GENETIC VALIDATION AND ITS ROLE IN CROP IMPROVEMENT

EDITED BY: Ahmed Sallam, Ahmad M. Alqudah, Peter Stephen Baenziger and Awais Rasheed

PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-431-4
DOI 10.3389/978-2-83250-431-4

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

GENETIC VALIDATION AND ITS ROLE IN CROP IMPROVEMENT

Topic Editors:
Ahmed Sallam, Assiut University, Egypt
Ahmad M. Alqudah, Qatar University, Qatar
Peter Stephen Baenziger, University of Nebraska System, United States
Awais Rasheed, Quaid-i-Azam University, Pakistan

Citation: Sallam, A., Alqudah, A. M., Baenziger, P. S., Rasheed, A., eds. (2022). Genetic Validation and its Role in Crop Improvement. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-431-4

Table of Contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1078246

OPEN ACCESS

EDITED AND REVIEWED BY
Humira Sonah,
Laval University, Canada

*CORRESPONDENCE
Ahmed Sallam,
amsallam@aun.edu.eg

SPECIALTY SECTION
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

RECEIVED 24 October 2022
ACCEPTED 16 December 2022
PUBLISHED 04 January 2023

CITATION
Sallam A, Alqudah AM, Baenziger PS and Rasheed A (2023), Editorial: Genetic validation and its role in crop improvement.
Front. Genet. 13:1078246.
doi: 10.3389/fgene.2022.1078246

Editorial: Genetic validation and its role in crop improvement

Ahmed Sallam[1,2]*, Ahmad M. Alqudah[3], P. Stephen Baenziger[4] and Awais Rasheed[5]

[1]Department of Genetics, Faculty of Agriculture, Assiut University, Assiut, Egypt, [2]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany, [3]Biological Science Program, Department of Biological and Environmental Sciences, College of Art and Science, Qatar University, Doha, Qatar, [4]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, United States, [5]Department of Plant Sciences, Quaid-i-Azam University, Islamabad, Pakistan

KEYWORDS

genetic validation, QTLs, genetic improvment, molecular breeding approaches, crops

Editorial on the Research Topic
Genetic validation and its role in crop improvement

Gene discovery for economically important traits has remained a challenging Frontier in crop genomics and breeding. The recent advances in DNA sequencing technologies and genetic analysis approaches paved the way for discovering many genes and hotspot genomic regions controlling target traits. The detection of novel genomic regions or candidate genes is very useful for plant breeders and geneticists to improve crops, dissect the genetics of complex traits, and understand the biological mechanisms of genes underpinning traits of interest. Quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS) dominated recent crop gene discovery research. These studies are becoming routine activities to discover the genetic basis of important phenotypes and result in underlying allelic variations, marker-traits associations, and frequency of favorable alleles in the target germplasm to help in understanding crop functional genomics (Rasheed and Xia 2019). However, the discovered loci need further validation before consideration to be used in breeding. In most GWAS cases, the outputs can be ambiguous due to problems of confounding population structure with low-frequency causal alleles leading to false-negative results and other unaccounted factors including low-accuracy genotype calls at some loci (Browning and Yu, 2009) and small population size (Finnoet al., 2014; Alqudah et al., 2020). Therefore, further validation is necessary, using cross-population approaches where candidate loci are either validated in bi-parental populations or independent germplasm collections (Finnoet al., 2014).

Genetic validation (QTL, genomic regions, candidate genes, gene expression, marker development, *etc.*) is one of the basic steps for marker-assisted and genomic selection for any breeding or genetic program to achieve its goals. Genetic validation examines whether the same QTL or gene tends to be significantly detected when the material is grown in other locations or years and whether its effect can still be significantly detected when tested in different genetic backgrounds (Sallam et al., 2016). Furthermore, validation of the polymorphic DNA markers in different populations is useful for further genetic

diversity studies. The cost of sequencing and genotyping for genetic diversity studies could be an obstacle for some researchers. Providing validated and highly polymorphic markers could save time and effort in breeding programs.

The current Research Topic was planned to seek articles that aim to validate putative genetic results controlling target traits for the genetic improvement of crops. In total, 18 manuscripts were published in this Research Topic. The Research Topic covered many studies presenting proof of genetic validation which can be used for supporting future research studies.

Quan et al. conducted GWAS for salinity tolerance in wheat and identified five candidate genes including kinase family protein, E3 ubiquitin-protein ligase-like protein, and transmembrane protein. Tian et al. identified 41 loci in wheat underpinning such wheat flour properties and validated loci on chr3B and chr7B as important haplotypes. Likewise, Irshad et al., discovered nucleotide variations in gene encoding starch branching enzymes *TaSBEIII* and developed a KASP marker to identify the causal mutation. They also identified the allele frequencies of *TaSBEIII* in wheat collections from different countries. For gene validation, Ur Rehman et al. developed and validated KASP markers for drought tolerance related genes in wheat and validated their effect in different germplasm resources. Xu et al. demonstrated the effect of asymmetric somatic hybridization on synonymous codon usage bias in wheat. The validation approach was also used for disease resistant genetic region. Eltaher et al., identified and validated a high linkage disequilibrium region on chr1B, chr2A, and chr7B in wheat harboring stem rust resistance.

Maize lethal necrosis (MLN) is a viral disease with a devastating effect on maize production. Combined use of QTL and GWAS identified a major effect QTL, qMLN06_157, on chr6 and this QTL was proposed to be used in both marker-assisted forward breeding and marker-assisted backcrossing schemes to improve MLN resistance of breeding populations and key lines for eastern Africa (Murithi et al.). The maize (*Zea mays* L.) *ZmCNR13* gene, encoding a protein of *fw2.2-like* (*FWL*) family, has been demonstrated to be involved in cell division, expansion, and differentiation. In the present study, the genomic sequences of the *ZmCNR13* locus were re-sequenced in 224 inbred lines, 56 landraces and 30 teosintes, and the nucleotide polymorphism and selection signature were estimated (Zuo et al.). It was validated that natural variations of *ZmCNR13* might be involved in ear development and can be used in the genetic improvement of maize ear-related traits. Ahmed et al. evaluated maize inbred lines for pollen viability related traits under heat stress and identified five loci associated with these traits that had been validated and can be incorporated in maize breeding programs for improving heat stress tolerance.

Elattar et al. identified and validated seven major QTLs for soybean seed size and shape. Based on gene annotation analyses and RNA-Seq, they identified candidate genes which are potentially involved in seed-related traits.

The gelatinization of wheat flour is an important characteristic and useful in evaluating the eating and cooking quality of wheat. Mesocotyl is a crucial organ for pushing plants out of the soil, which plays a vital role in seedling emergence and establishment in direct-seeded rice. Eighteen QTLs for mesocotyl length were identified in rice, out of which 6 QTLs were validated in two mapping populations (Wang et al.). Further association analysis and gene expression studies confirmed the cross-population effect of two loci on chr1 and chr7 on mesocotyl length. Tomato is among the most valuable fruit crop and sugar and organic acids contribute to the overall flavor intensity of tomato. Metabolome and transcriptome studies in two tomato cultivars TM-1 and TM-38 identified that citric acid may play a more dominant role in the sugar/organic acid ratio of the tomato fruit (Li et al.). The contribution of both L-malic acid and citric acid to the fruit Brix was much greater than that of D-glucose and D-fructose. Genes involved in CHO and TCA metabolism, which have a significant correlation with the sugar/organic acid ratio were considered to be the contributing factors of fruit Brix.

Choudhury et al. used 36 SNP markers to identify the core set of 247 rice accessions from India's east coast and validated the utility of using these SNP markers for the development of core collection and diversity studies. In another study, the genetic architecture of grain yield in wheat was revealed using a 50K SNP array, and 38 loci were identified (Li et al.). Khan et al. characterized the DECT1 gene in rice and the GUS and qRT-PCR analysis indicated that *DCET1* is specifically expressed in the anther till the developmental stage 9, consistent with the observed phenotype. The characterization of *DCET1* in callose regulation, pollen wall patterning, and tapetal programmed cell death (PCD) strengthens our knowledge for knowing the regulatory pathways involved in rice male reproductive development and has prospects in hybrid rice breeding. The phenomenal increase in the use of nitrogenous fertilizers coupled with poor nitrogen use efficiency is among the most important threats to the environment, economic, and social health. Sandhu et al. performed a meta-QTL analysis on 1,330 QTL from 29 studies published in the past 2 decades. A hot spot region associated with correlated traits on Chr 1, 4, and 8 and candidate genes associated with nitrate transporters, nitrogen content, and ammonium uptake on chromosomes 2, 4, 6, and 8 have been identified.

Iqbal et al. performed GWAS to identify the loci associated with single fiber quality in Asian cotton (*Gossypium arboreum*). They further identified 56 differentially expressed genes among the trait-associated SNPs to pinpoint the candidate genes.

Because, the functional gene validation is very important step to understand the mechanisms, Riaz et al. reviewed the recent breakthroughs in Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated protein (Cas) mediated gene editing in cereal grasses and overviewed 25 different studies about

it. The importance of CRISPR technology application in cultivated grasses improvement had been extensively explained.

In conclusion, a proof of genetic validation provides a very promising improvement in the production and productivity of crops for future research. Unfortunately, few research studies validated their genetic findings. The results of genetic validation in different crops (e.g. wheat, rice, maize, and cotton, *etc.*) presented in this Research Topic can be used for future and necessary genetic validation studies.

Author contributions

AS suggested, the Research Topic, and edited the articles. AA, PB, AR, edited the articles.

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

Alqudah, A. M., Sallam, A., Stephen Baenziger, P., and Börner, A. (2020). Gwas: Fast-forwarding gene identification and characterization in temperate cereals: Lessons from barley – a review. *J. Adv. Res.* 22, 119–135. doi:10.1016/J.JARE.2019.10.013

Browning, B. L., and Yu, Z. (2009). Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am. J. Hum. Genet.* 85 (6), 847–861.

Finno, C. J., Aleman, M., Higgins, R. J., Madigan, J. E., and Bannasch, D. L. (2014). Risk of false positive genetic associations in complex traits with underlying population structure: a case study. *Vet. J.* 202, 543–549.

Rasheed, A., and Xia, X. (2019). From markers to genome-based breeding in wheat. *Theor. Appl. Genet.* 132, 767–784. doi:10.1007/s00122-019-03286-4

Sallam, A., Arbaoui, M., El-Esawi, M., Abshire, N., and Martsch, R. (2016). Identification and verification of QTL associated with frost tolerance using linkage mapping and GWAS in winter faba bean. *Front. Plant Sci.* 7, 1098. doi:10.3389/fpls.2016.01098

frontiers in Genetics

ORIGINAL RESEARCH
published: 20 May 2021
doi: 10.3389/fgene.2021.663941

Genome-Wide Association Study Uncover the Genetic Architecture of Salt Tolerance-Related Traits in Common Wheat (*Triticum aestivum* L.)

Xiaoyan Quan[1†], Jindong Liu[2,3†], Ning Zhang[1], Chunjuan Xie[1], Hongmei Li[1], Xianchun Xia[3], Wenxing He[1] and Yuxiang Qin[1*]*

[1] Department of Biological Science, School of Biological Science and Technology, University of Jinan, Jinan, China, [2] Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, [3] Institute of Crop Sciences, National Wheat Improvement Center, Chinese Academy of Agricultural Sciences, Beijing, China

Soil salinity is a serious threat to wheat yield affecting sustainable agriculture. Although salt tolerance is important for plant establishment at seedling stage, its genetic architecture remains unclear. In the present study, we have evaluated eight salt tolerance–related traits at seedling stage and identified the loci for salt tolerance by genome-wide association study (GWAS). This GWAS panel comprised 317 accessions and was genotyped with the wheat 90 K single-nucleotide polymorphism (SNP) chip. In total, 37 SNPs located at 16 unique loci were identified, and each explained 6.3 to 18.6% of the phenotypic variations. Among these, six loci were overlapped with previously reported genes or quantitative trait loci, whereas the other 10 were novel. Besides, nine loci were detected for two or more traits, indicating that the salt-tolerance genetic architecture is complex. Furthermore, five candidate genes were identified for salt tolerance–related traits, including kinase family protein, E3 ubiquitin-protein ligase-like protein, and transmembrane protein. SNPs identified in this study and the accessions with more favorable alleles could further enhance salt tolerance in wheat breeding. Our results are useful for uncovering the genetic mechanism of salt tolerance in wheat at seeding stage.

Keywords: candidate gene, GWAS, Na/K, QTL, salt tolerance, wheat

OPEN ACCESS

Edited by:
Awais Rasheed, Quaid-i-Azam University, Pakistan

Reviewed by:
Satinder Kaur, Punjab Agricultural University, India
Prashant Vikram, International Center for Biosaline Agriculture (ICBA), United Arab Emirates
Mian Abdur Arif, Nuclear Institute for Agriculture and Biology, Pakistan

***Correspondence:**
Wenxing He
163.hwx@163.com
Yuxiang Qin
yuxiangqin@126.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *04 February 2021*
Accepted: *24 March 2021*
Published: *20 May 2021*

Citation:
Quan X, Liu J, Zhang N, Xie C, Li H, Xia X, He W and Qin Y (2021) Genome-Wide Association Study Uncover the Genetic Architecture of Salt Tolerance-Related Traits in Common Wheat (Triticum aestivum L.). Front. Genet. 12:663941. doi: 10.3389/fgene.2021.663941

INTRODUCTION

Soil salinity is a serious abiotic stress affecting more than 800 million hectares of total agricultural land worldwide (Song and Wang, 2015; Ding et al., 2018; Li W. H. et al., 2020). Plant growth and development, including seed germination, shoot height (SH), root length (RL), and biomass, are significantly inhibited under salt stress (Hasan et al., 2017; Liu et al., 2018; Füzy et al., 2019). It continues to be a serious threat to agricultural sustainability as greater than 50% of agricultural land

will be salinized by 2050 because of unscientific irrigation management, such as flood irrigation (Flowers, 2004; Demiral and Türkan, 2006). Hence, developing salt-tolerant cultivars is one of the most effective and sustainable ways to utilize the saline alkali land.

Salt stress damages plant tissues through osmotic stress and ionic toxicity to cells (Zhu, 2003; Munns and Tester, 2008). Osmotic stress is the first stress to confront when plants are subjected to salt stress, because salt stress first decreases root water-uptake ability by reducing soil water potential (Liang et al., 2018). Subsequently, Na^+ at high concentrations is absorbed by the roots and transported to the shoots. Excessive cellular Na^+ has a negative effect on metabolic processes and photosynthetic efficiency (Munns, 2005; Ahmadi et al., 2011), and it also makes it difficult to maintain intracellular ion homeostasis due to impeding K^+ absorption (Munns and Tester, 2008; Shabala and Cuin, 2008). Both osmotic stress and ion toxicity cause oxidative damage of membrane lipids and proteins and further disturb plant growth rate and development (Mittler, 2002; Imlay, 2003).

Plant salt tolerance is controlled by a series of quantitative trait loci (QTLs) or genes and is a typically genetic and physiological trait (Flowers, 2004; Munns and Tester, 2008). The complexity of mechanisms involved in salt stress tolerance for plants limits the progress toward salt-tolerance breeding. Mapping the QTLs for salt tolerance–related traits and breeding cultivars with high salt tolerance have become an effective way to reduce the losses caused by salt threat (Akram et al., 2020). QTL mapping based on biparental populations is a traditional approach to dissect the genetic mechanisms of complex quantitative inheritance traits (Kumar et al., 2015). However, only two allelic effects in any single locus can be evaluated in a biparental mapping, which limits its power to uncover the nature of genetic variation in wheat (Shi et al., 2017). Thus, the genome-wide association analysis (GWAS) method, which is based on linkage disequilibrium (LD), has become an alternative way to identify markers significantly associated with complex agronomic traits (Mackay and Powell, 2007; Cockram et al., 2010; Wang et al., 2014; Winfield et al., 2015; Liu et al., 2017; Akram et al., 2020), biotic stress (Singh et al., 2020; Dababat et al., 2021), and abiotic stress (Rehman-Arif and Börner, 2020). Compared with traditional biparental mapping, GWAS is more efficient and less expensive with no need to develop biparental population and a more representative gene pool. More recently, GWAS has become an effective tool for dissecting the genetic architecture of salt tolerance–related traits in crops such as wheat (Liu et al., 2017), rice (Kumar et al., 2015), barley (Long et al., 2013), and soybean (Kan et al., 2015).

Land salinization has become a serious threat in China because of the increased irrigation management, climate change, and fertilizer use. Breeding for salt-tolerant cultivars could be greatly improved by the identification and application of molecular markers. Furthermore, plants are the most sensitive to salinity at seeding stage, other than flowering, and the grain filling stage (Gerona et al., 2019). Thus, it is of great significance to identify loci related with salt tolerance at seedling stage of wheat. We investigated a diverse panel of 317 elite wheat cultivars employing GWAS to (1) uncover the genetic mechanism of salt tolerance, (2) detect the markers associated with salt tolerance, and (3) search the candidate genes and accessions with more favorable alleles for salt tolerance.

MATERIALS AND METHODS

Plant Materials and Treatments

To evaluate the salt tolerance of the modern cultivars, 317 various wheat accessions, mainly including modern cultivars and improved accessions, comprising 260 accessions from China, and 57 accessions from other countries (**Supplementary Table 1**) were used for GWAS. The experiment was carried out in a growth chamber at 20 ± 2°C with a photoperiod of 16-h light/8-h dark and a light intensity of 300 μmol m^{-2} s^{-1}. Seeds of all the accessions were germinated and grown to 7-day-old seedlings. The uniform seedlings were transplanted into 20-L containers covered with polystyrol plates with 60 evenly spaced holes. Half-strength modified Hoagland's solution was used for the cultivation of the seedlings with continuous aeration (Wu et al., 2015; Shan et al., 2018). The cultivation solution was renewed every 5 days. Completely randomized experiment was designed with four replicates, involving 16 plants grown separately. Each experiment comprised eight randomized units allocated to control (0 mM NaCl) and salt (200 mM NaCl) treatment groups. Two-leaf seedlings were treated with 200 mM NaCl; NaCl was gradually added with a 100-mM increment per day.

Phenotypic Measurement and Data Analysis

The plants were harvested for measuring SH, RL, shoot fresh weight (SFW), and root fresh weight (RFW) after 1 week of salt treatment. Shoots and roots were separated after washing with distilled water and oven-dried at 105°C for 30 min and then dried further at 80°C for 72 h. Record shoot dry weight (SDW) and root dry weight (RDW). And then 100 mg fine powder of dried shoot was incubated with 5 mL extraction buffer in a 90°C water bath for 30 min. The extraction buffer was a mixture of 60% trichloroacetic acid, nitric acid, and sulfuric acid (2:10:1). The supernatant was taken after centrifugation. Na^+ and K^+ content was estimated using atomic absorption spectrophotometer (TAS-990).

Analyses of variance (ANOVA) of all the tested traits among genotypes and replicates were determined with SAS v9.3 (SAS, Institute, http://www.sas.com), and the differences at $P < 0.01$ were considered highly significant. The salt-tolerance index was employed the relative changes of SDW, RDW, SFW, RFW, SH, RL, and shoot K content (labeled as RSDW, RRDW, RSFW, RRFW, RSH, RRL, and RK, respectively), as well as the values of Na content and Na/K ratio under salt stress (Munns and James, 2003; Long et al., 2013). The relative change of each trait was calculated by the value of salt stress/control, for example, RSDW = SDW (salt)/SDW (control). Basic statistical analysis of the data of all the phenotypic traits under both treatments and all the salt-tolerance indices was performed using Microsoft Excel

2019. Pearson correlations of the salt-tolerance indices of 317 wheat accessions were conducted in SPSS version 16.0.

Genotyping and Population Structure Analysis

All the 317 accessions were genotyped with the wheat 90 K single-nucleotide polymorphism (SNP; Illumina, 81,587 SNPs) chip by Capital Bio[1]. SNPs used for subsequent GWAS analysis were obtained after quality control (minor allele frequency > 0.05 and missing data < 20%; Liu and Muse, 2005). The physical positions of SNPs were obtained from the International Wheat Genome Sequencing Consortium website (IWGSC, http://www.wheatgenome.org/; IWGSC v1.1).

Population structure was analyzed using 1,000 filtered SNPs with Admixture 1.3.0 program[2]. ADMIXTURE ran from $K = 2$ to $K = 12$ clusters to identify the optimal K value. Principal components analysis (PCA) with a number of five components and phylogenetic trees [neighbor-joining (NJ)] were also used to uncover the population information with the software GAPIT (Lipka et al., 2012) based on R 3.5.3 and Tassel v5.1 (Bradbury et al., 2007), respectively. Besides, the LD was estimated using GAPIT software (Lipka et al., 2012) according to Liu et al., 2017.

Genome-Wide Association Analysis

Genome-wide association analysis was conducted on the salt-tolerance indices of phenotypic traits by employing the kinship matrix (K matrix) in a mixed linear model (MLM) to avoid the spurious marker trait associations (MTAs) caused by genetic background. In the present study, the P value indicated whether an SNP was associated with corresponding trait, and the R^2 indicated the phenotypic variation explained by the markers. As Bonferroni–Holm correction (Holm, 1979) for multiple testing ($\alpha = 0.05$) was too conserved and no significant MTA was detected with this criterion, markers with an adjusted $-\log_{10}$ (P value) ≥ 3.0 were selected as significantly associated markers (Gurung et al., 2014; Houston et al., 2014; Bellucci et al., 2015; Liu et al., 2017, 2019). Besides, Manhattan and Q–Q plot were drawn using CMplot package (Yin et al., 2020) implemented in R 3.5.3.

Allelic Effects and Candidate Genes

Candidate genes were identified as all the genes located in LD block region around the significant SNP (±3 Mb based on LD decay analysis) of each important locus from the physical position of IWGSC[3] (IWGSC v1.1). In the present study, alleles with positive effects on higher salt tolerance at seeding stage are referred to as "favorable alleles," and those contributing to lower tolerance are "unfavorable alleles." The peak SNPs for each locus were used to count the alleles frequencies and allelic effects. Regression analysis between favorable or unfavorable alleles and corresponding traits were conducted using Microsoft Excel 2019.

RESULTS

Marker Coverage, Population Structure, and LD

In total, 54,121 SNPs (14,063.9 Mb, 0.26 Mb per marker) was applied to GWAS for salt tolerance–related traits in 317 wheat cultivars (**Supplementary Tables 1**, **2**). In total, 18,623 (34.41%), 21,091 (38.97%), and 14,407 (26.62%) markers were from the A, B, and D genomes, with 4,934.5, 5,179.0, and 3,950.4 Mb, respectively. The average genetic diversity for the whole genome was 0.356 (0.009–0.500; **Supplementary Table 2**). All the 317 accessions could be divided into three subgroups based on the population structure, PCA, and NJ tree analysis (**Figure 1**). Of these, the subgroup I carried 93 accessions and was dominated by Anhui, Henan, Shandong, and foreign cultivars; subgroup II included 89 accessions and mainly comprising varieties from Hebei, Shanxi, and Shandong provinces; most accessions (135) belonged to subgroup III and mainly from Henan, Shandong, and Sichuan province (**Supplementary Table 1**). The LD decay analysis indicated that the LD decay distance was about 3 Mb for the whole genome (**Supplementary Figure 1**).

Phenotypic Variations for Salt Tolerance–Related Traits

The salt-tolerance indices for SDW, RDW, SFW, RFW, SH, RL, and K content (labeled as RSDW, RRDW, RSFW, RRFW, RSH, RRL, and RK, respectively) were estimated in the current study. The contents of Na and Na/K in shoot under salt stress were also assessed as salt-tolerance index. Continuous variation was observed across all traits with approximately normal distributions (**Supplementary Figure 2** and **Supplementary Tables 2–4**). The result showed that RRL, RSH, RRFW, RSFW, RRDW, RSDW, and RK contents ranged from 0.31 to 0.9, 0.62 to 1.03, 0.18 to 1.05, 0.26 to 0.78, 0.17 to 0.94, 0.41 to 0.94, and 0.58 to 1.32, respectively (**Table 1**). Meanwhile, the contents of Na and Na/K under salt stress ranged from 24.67 to 88.10 mg/g DW and 0.33 to 1.58, respectively. It indicated a wide range of variation in each salt tolerance–related trait. ANOVA for salt tolerance–related traits revealed significant differences ($P \leq 0.01$) among genotypes (G; **Supplementary Table 4**), suggesting that much of the phenotypic variation was derived from genetic factors. However, the broad sense heritability (h^2) could not be calculated because of the phenotype evaluated only in laboratory.

Pearson correlations among the salt tolerance–related traits (salt-tolerance indices) of seedlings under salt stress were calculated, and the correlation coefficients for nine traits are shown in **Table 2**. RSDW showed significantly negative correlation with Na and Na/K and positive correlation with the other traits. Pairwise positive correlation among six traits (RRDW, RSFW, RRFW, RSH, RRL, and RK) was observed. Notably, RSHs were uncorrelated with Na, Na/K, and RK in

[1]http://www.capitalbiotech.com/

[2]http://software.genetics.ucla.edu/admixture/

[3]http://www.wheatgenome.org/

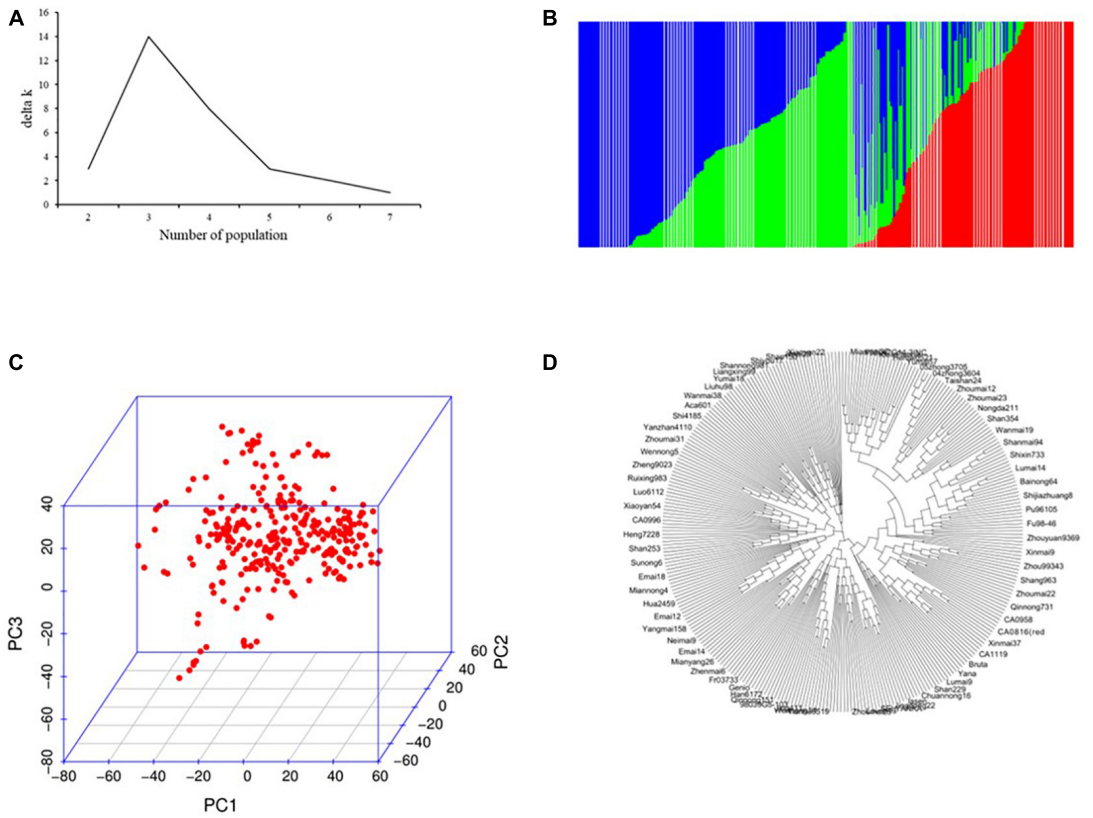


FIGURE 1 | Population structure analysis of 317 wheat accessions. **(A)** Estimated ΔK over five repeats of structure analysis; **(B)** three subgroups inferred by structure analysis; **(C)** principal components analysis (PCA) plots; and **(D)** neighbor-joining (NJ) tree.

TABLE 1 | Phenotypic variation of salt-tolerance index traits in the 317 wheat accessions.

Trait	Minimum	Maximum	Mean	SD	CV (%)	Skewness	Kurtosis
RSDW	0.41	0.94	0.62	0.09	12.96	0.43	0.76
RRDW	0.17	0.94	0.52	0.11	19.65	0.31	0.97
RSFW	0.26	0.78	0.43	0.07	16.34	0.63	2.10
RRFW	0.18	1.05	0.47	0.12	24.55	0.66	1.71
RSH	0.62	1.03	0.79	0.06	7.25	–0.05	0.33
RRL	0.31	0.90	0.62	0.08	12.82	–0.13	0.80
RK	0.58	1.32	0.98	0.12	12.67	–0.24	0.26
Na (mg/g DW)	24.67	88.10	50.40	11.60	23.35	0.58	0.34
Na/K	0.33	1.58	0.74	0.23	30.93	0.84	0.61

RSDW, ratio of shoot dry weight under salt stress and control; RRDW, ratio of root dry weight under salt stress and control; RSFW, ratio of shoot fresh weight under salt stress and control; RRFW, ratio of root fresh weigh under salt stress and control; RSH, ratio of shoot height under salt stress and control; RRL, ratio of root length under salt stress and control; RK, ratio of shoot K content under salt stress and control; and Na, shoot Na content under salt stress.
Values of minimum, maximum, and mean for each trait except Na represent ratio of each trait value between salt stress and control.
SD, standard deviation; CV (%), coefficient of variation.

shoot, whereas Na and Na/K negatively correlated with all the other traits except RSH.

MTA Analysis

At *P* value of 0.001 ($-\log_{10}$ value of 3), a total of 37 significantly associated SNPs (MTAs) for eight salt tolerance–related traits including RSDW, RRDW, RRFW, RSH, RRL, RK, Na content, and Na/K were identified (**Figure 2**, **Supplementary Figure 3**, and **Table 3**). The 37 associated SNPs were distributed on 16 unique loci and located on chromosomes 1B, 1D, 2A (2), 4A, 5A, 5B (4), 6B, 7A (4), and 7B, respectively (**Table 3**), explaining phenotypic variation ranging from 6.3 to 18.6%. The number of loci found most in the A genome (8) and B genome (7), whereas only one locus was identified in the D genome (**Table 3**). Among these loci, four were detected for RSDW, four for RRDW, six for RRFW, five for RSH, one

TABLE 2 | Correlation coefficients among nine traits of seedlings under salt stress.

Pearson correlation	RSDW	RRDW	RSFW	RRFW	RSH	RRL	RK	Na	Na/K
RSDW	1.0	0.687**	0.777**	0.611**	0.393**	0.369**	0.163**	−0.269**	−0.272**
RRDW	0.687**	1.0	0.619**	0.832**	0.207**	0.469**	0.142*	−0.406**	−0.390**
RSFW	0.777**	0.619**	1.0	0.618**	0.395**	0.379**	0.211**	−0.485**	−0.467**
RRFW	0.611**	0.832**	0.618**	1.0	0.200**	0.490**	0.253**	−0.404**	−0.413**
RSH	0.393**	0.207**	0.395**	0.200**	1.0	0.304**	0.108	0.009	0.006
RRL	0.369**	0.469**	0.379**	0.490**	0.304**	1.0	0.107	−0.125*	−0.171**
RK	0.163**	0.142*	0.211**	0.253**	0.108	0.107	1.0	−0.408**	−0.606**
Na	−.269**	−0.406**	−0.485**	−0.404**	0.009	−0.125*	−0.408**	1.0	0.927**
Na/K	−0.272**	−0.390**	−0.467**	−0.413**	0.006	−0.171**	−0.606**	0.927**	1.0

RSDW, ratio of shoot dry weight under salt stress and control; RRDW, ratio of root dry weight under salt stress and control; RSFW, ratio of shoot fresh weight under salt stress and control; RRFW, ratio of root fresh weigh under salt stress and control; RSH, ratio of shoot height under salt stress and control; RRL, ratio of root length under salt stress and control; RK, ratio of shoot K content under salt stress and control; and Na, shoot Na content under salt stress.
***Correlation is significant at the 0.01 level. *Correlation is significant at the 0.05 level.*

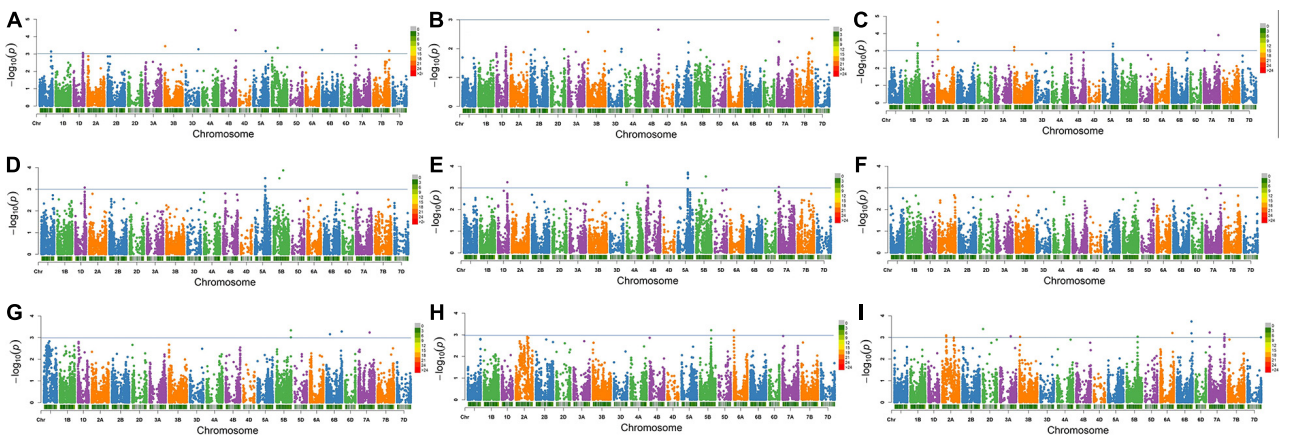


FIGURE 2 | Manhattan plots for salt tolerance–related traits in 317 wheat accessions by the mixed linear model (MLM) in Tassel v5.0. **(A)** Relative shoot dry weight; **(B)** relative shoot fresh weight; **(C)** relative shoot height; **(D)** relative root dry weight; **(E)** relative root fresh weight; **(F)** relative root length; **(G)** relative shoot K content; **(H)** shoot Na content under salt stress; and **(I)** Na/K in shoot under salt stress. The $-\log_{10}(P)$ values from a genome-wide scan are plotted against positions on each of the 21 chromosomes. Horizontal lines indicate genome-wide significance thresholds.

for RRL, two for RK, one for Na content, and five for Na/K (**Table 3**). Notably, nine loci (*IAAV8839*, *Kukri_rep_c68263_453*, *RAC875_c25567_1204*, *GENE-3440_199*, *Ku_c15213_388*, *Ku_c5191_340*, *BS00022442_51*, *BobWhite_c149_3064*, *and BS00071025_51*) were identified for two or more traits, proving the existence of pleiotropic regions. To add it, the quantile and quantile (Q-Q) plots of all the traits (**Supplementary Figure 3**) show that the expected −log10(*p*) value was close to the observed distribution (**Supplementary Figure 2**), indicating that the GWAS analysis through MLM (PC + K) was appropriate to locate MTAs related to salt tolerance in the germplasm under investigation.

Relationship Between Salt Tolerance and the Number of Tolerance Alleles

Here, the RSDW of plants was used to stand for plant tolerance to salt stress (Szira et al., 2008). According to the relation between RSDW and other salt tolerance–related traits, favorable alleles for salt tolerance were counted as favorable alleles for RSH, RRDW, RRFW, RRL, RK, and unfavorable alleles for Na content and Na/K. To further understand the combined effects of alleles on salt tolerance, we examined the number of favorable alleles for salt tolerance in each accession (**Supplementary Table 1**). Interestingly, the number of favorable and unfavorable alleles in single accession both ranged from 1 to 15 (**Supplementary Table 1**). Linear regression analysis showed that RSDW displayed significantly positive correlation with total number of favorable alleles for salt tolerance (R^2 = 0.616; **Figure 3A**) and notably negative correlation with number of unfavorable alleles (R^2 = 0.581; **Figure 3B**). Thus, accessions with more favorable alleles and less unfavorable alleles were more tolerant to salt stress.

DISCUSSION

To date, many studies involved in physiological and molecular mechanisms on salt tolerance have been carried out in wheat, but the studies on identification of salt tolerance–related QTLs

TABLE 3 | Loci for salt tolerance–related traits in 317 wheat accessions identified by Tassel v5.0.

Trait[a]	Marker	Chr[b]	Pos[c]	*P* value[e]	Marker R^{2f}	QTL/gene[g]	Favorable allele[d]
SH	*Excalibur_c112015_118*	1B	566.8	4.94E-04	8.5	Li L. et al., 2020	T
SH	*Ku_c956_1797*	1B	571.4	3.63E-04	8.4		C
RFW	*IAAV8839*	1D	430.4	5.51E-04	7.9	Li L. et al., 2020	T
RDW	*IAAV8839*	1D	430.4	8.46E-04	7.2		T
SDW	*IAAV8839*	1D	430.4	9.53E-04	7.1		T
SDW	*Ex_c18035_602*	1D	430.6	8.81E-04	7.1		G
SDW	*IACX1549*	1D	431.4	9.00E-04	7.3		A
SDW	*Kukri_c59519_352*	1D	431.7	9.61E-04	7.0		A
SH	*Tdurum_contig11803_306*	2A	36.0	1.22E-04	9.7		A
SH	*wsnp_Ex_c19556_28530243*	2A	36.1	2.19E-05	12.2		C
SH	*BobWhite_c26374_339*	2A	36.1	9.00E-04	7.6		T
Na/K	*Ku_c1217_312*	2A	182.3	8.09E-04	7.0		G
Na/K	*Ra_c22724_1137*	2A	182.3	9.14E-04	7.0		A
RFW	*JD_c21248_511*	4A	2.8	7.29E-04	7.5	Li L. et al., 2020	A
RFW	*BS00074614_51*	4A	3.0	5.61E-04	9.5		A
SH	*Kukri_rep_c68263_453*	5A	464.5	5.60E-04	7.9		T
SH	*Kukri_c17430_972*	5A	468.5	3.87E-04	8.6		C
RFW	*BS00027465_51*	5A	471.7	1.99E-04	9.0		G
RDW	*BS00027465_51*	5A	471.7	3.10E-04	8.4		T
RFW	*BS00027466_51*	5A	471.7	2.28E-04	8.9		C
RDW	*BS00027466_51*	5A	471.7	7.22E-04	7.5		C
RFW	*BS00022509_51*	5A	472.3	3.61E-04	8.5		A
RFW	*Kukri_c2326_1037*	5A	476.6	3.40E-04	8.4		T
RDW	*Kukri_c2326_1037*	5A	476.6	7.62E-04	7.3		G
RFW	*Kukri_c2326_995*	5A	476.6	2.39E-04	8.8		C
RDW	*RAC875_c25567_1204*	5B	273.1	3.13E-04	8.7		G
SDW	*RAC875_c25567_1204*	5B	273.1	4.46E-04	8.2		G
RDW	*GENE-3440_199*	5B	445.5	1.36E-04	9.5		A
RFW	*GENE-3440_199*	5B	445.5	2.99E-04	8.7		C
Na	*Ku_c15213_388*	5B	515.2	6.18E-04	18.6		T
Na/K	*Ku_c15213_388*	5B	515.2	9.25E-04	16.5		A
K	*Kukri_c16864_398*	5B	684.6	9.81E-04	7.3	Li L. et al., 2020	C
K	*Tdurum_contig65330_190*	5B	684.6	4.61E-04	8.4	Oyiga et al., 2018	G
Na/K	*Ku_c5191_340*	6B	668.8	1.84E-04	8.8		T
Na/K	*RAC875_rep_c69963_514*	6B	668.8	6.49E-04	7.4		C
K	*Kukri_rep_c101126_469*	6B	705.8	5.25E-04	8.4		A
Na/K	*BS00022442_51*	7A	30.1	5.93E-04	7.4		C
SH	*BS00062724_51*	7A	36.5	9.61E-04	7.2		A
SDW	*BS00040929_51*	7A	83.9	3.16E-04	8.6		G
SDW	*IAAV1971*	7A	84.7	4.65E-04	7.9		T
SH	*BobWhite_c149_3064*	7A	670.8	1.23E-04	11.4		C
RL	*wsnp_Ku_c19943_29512612*	7A	675.4	7.66E-04	7.2		C
Na/K	*IAAV6119*	7A	709.4	7.10E-04	7.2	Li L. et al., 2020	T
Na/K	*BobWhite_c46250_98*	7A	709.4	9.89E-04	6.7		C
SFW	*BS00071025_51*	7B	730.2	2.81E-04	6.3	Li L. et al., 2020	T
SDW	*BS00071025_51*	7B	730.2	6.81E-04	8.3		C

[a] *RSDW, ratio of shoot dry weight under salt stress and control; RRDW, ratio of root dry weight under salt stress and control; RSFW, ratio of shoot fresh weight under salt stress and control; RRFW, ratio of root fresh weigh under salt stress and control; RSH, ratio of shoot height under salt stress and control; RRL, ratio of root length under salt stress and control; RK, ratio of shoot K content under salt stress and control; and Na, shoot Na content under salt stress.*

[b] *Chr: chromosome.*

[c] *The physical positions of SNP markers based on wheat genome sequences from the International Wheat Genome Sequencing Consortium (IWGSC, http://www.wheatgenome.org/).*

[d] *Favorable allele (SNP) for traits is underlined.*

[e] *The P values were calculated by Tassel v5.0.*

[f] *Percentage of phenotypic variance explained by the MTA from the results of Tassel v5.0.*

[g] *The previously reported QTL or genes within the same chromosomal regions.*

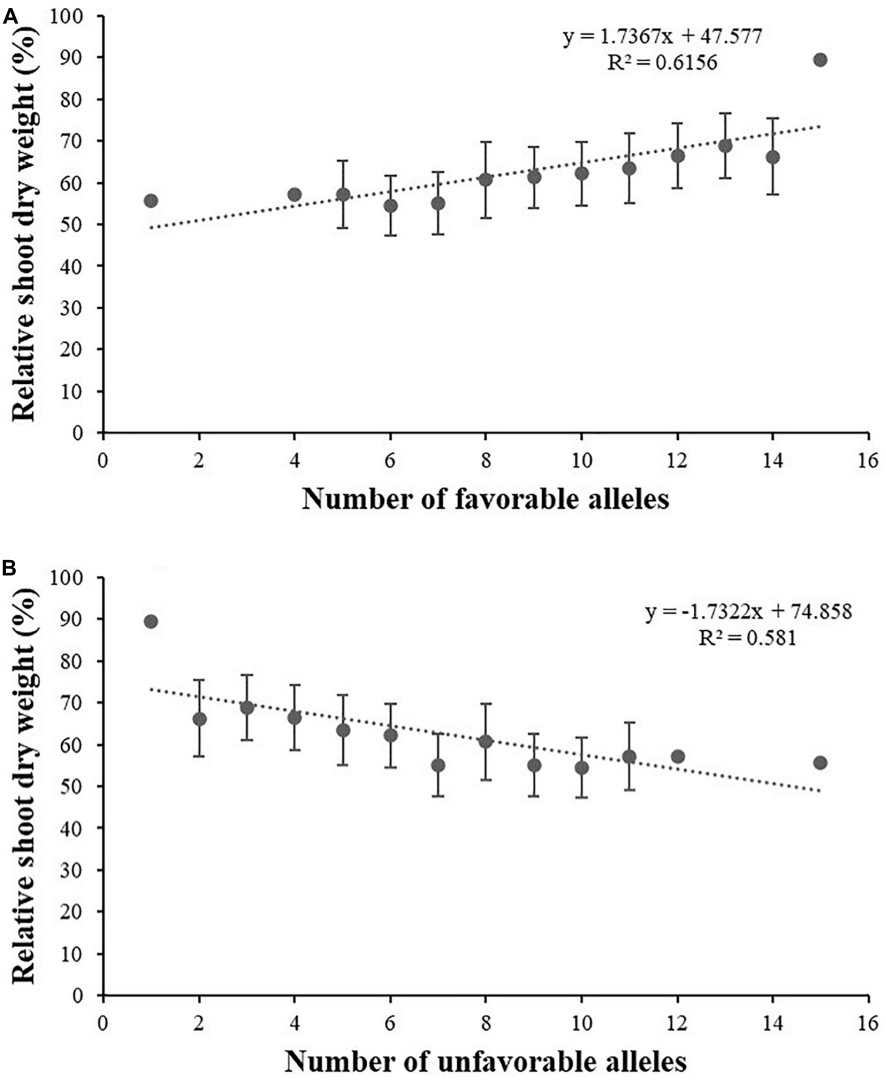


FIGURE 3 | Linear regression between relative shoot dry weight and the number of favorable alleles **(A)** and unfavorable alleles **(B)** for salt tolerance.

TABLE 4 | Candidate gene for salt tolerance–related traits.

Marker	Candidate gene	Gene position	mRNA position	Annotation
Tdurum_contig11803_306	*TraesCS2A01G079300*	chr2A:36037402-36042137	chr2A:36037725-36041668	Kinase family protein
GENE-3440_199	*TraesCS5B01G261000*	chr5B:445453118-445459346	chr5B:445453337-445459168	E3 ubiquitin-protein ligase-like protein
Ku_c5191_340	*TraesCS6B01G393800*	chr6B:668776400-668778663	chr6B:668776660-668778663	Transmembrane 9 superfamily member
Kukri_rep_c101126_469	*TraesCS6B01G441700*	chr6B:705756764-705758664	chr6B:705756764-705758664	Leucine-rich repeat Receptor-like protein Kinase family protein
BS00022442_51	*TraesCS7A01G060900*	chr7A:30062542-30064125	chr7A:30062602-30063750	Transmembrane protein

are limited (Ma et al., 2007; Genc et al., 2010; Xu et al., 2013; Asif et al., 2018; Oyiga et al., 2018). The information achieved is far from enough in salt tolerance improvement. Therefore, identification of much more QTLs/genes and further study were needed to improve salt tolerance in wheat. Association analysis is a highly efficient tool for the identification of

QTLs for complicated quantitative traits. It makes possible to identify markers significantly associated with salt tolerance, thereby facilitating salt-tolerance breeding in wheat by MAS. The panel used in the present study showed a high genetic diversity of salt tolerance, consisting of landraces, advanced lines, and released cultivars from different ecological regions and countries (**Table 1**).

It has been reported that the biomass parameters of RSFW, RRFW, RSDW, and RRDW have been used together to determine salt tolerance under abiotic stress in previous studies (Zhao et al., 2010; Liu et al., 2018). RRL and RSH were also taken as important indicators to show plant salt tolerance in numerous studies (Hasan et al., 2017; Liu et al., 2018; Füzy et al., 2019). Moreover, the agronomic characters are always the primary targets in plant breeding (Zeng et al., 2003). In addition, shoot Na content and Na/K ratio were also considered as two of the key determinants for plant salt tolerance (Liang et al., 2018). Thus, these characters were used to identify QTLs for salt tolerance in the present study. Salt tolerance at seedling stage is important because the initial plant growth will affect the final production. Our study demonstrated large variation in various parameters among different wheat varieties at seedling stage under salinity stress (**Table 1**). The salt-tolerance index was employed for GWAS, as it is considered a reliable measure for assessing salt tolerance, which permits the control for background effects among different genotypes (Karan and Subudhi, 2012; Long et al., 2013; Kan et al., 2015). Significant genetic variations for the agronomic traits measured and shoot Na contents and Na/K ratio in this panel indicated the possibility of genetic improvement of salt tolerance (Yong et al., 2015; Oyiga et al., 2018).

The diverse panel consisted of 317 wheat accessions originating from 10 countries and could belong to three unique subgroups (**Figure 1**). The subgroups were generally consistent with geographic origins and pedigrees. For example, the CA1005, CA1062, and the CA1133 shared the common parent of Jing 411 in subgroup I; the released wheat cultivars Zhongmai 871, Zhongmai 875, and Zhongmai 895 were derived from Zhoumai 16, clustered with Zhoumai 16 in subgroup II; the Xiaoyan 54 and Xiaoyan 81 were clustered with Xiaoyan 6 in the subgroup III. To avoid spurious MTAs caused by population structure and kinship, MLM model (Q + K) based on Tassel v5.1 was adopted for association analysis (Zhao et al., 2007; Stich et al., 2008; Long et al., 2013). The LD decay is influenced by population structure, allele frequency, recombination rate, etc., and affects the precision of GWAS. Previous studies have been reported that LD decay in common wheat ranged between 1.5 and 15 cM using SSR, DArT, or SNP markers and various in landraces and cultivars (Liu et al., 2017), which showed a longer LD decay distance (Breseghello and Sorrells, 2006; Liu et al., 2017; Akram et al., 2020). Most of the accessions used in our study originated from the released cultivars and the improved cultivars. In this panel, the LD decay distance was approximately 3 Mb for the whole genome, in accordance with the previous studies (Breseghello and Sorrells, 2006; Liu et al., 2017; Akram et al., 2020). The marker densities for the whole genome are higher than LD decay distance and thus highly reliable for detecting MTAs in the diverse panel according to Breseghello and Sorrells (2006).

Among 16 QTLs discovered, nine co-associated QTLs (*IAAV8839*, *Kukri_rep_c68263_453*, *RAC875_c25567_1204*, *GENE-3440_199*, *Ku_c15213_388*, *Ku_c5191_340*, *BS00022442_51*, *BobWhite_c149_3064*, *and BS00071025_51*) were detected for two or more traits, suggesting that the genetic mechanisms of salt tolerance are complex. During the past two decades, more than 50 QTLs/genes for salt tolerance–related traits in wheat were identified using linkage or association mapping. Munns et al. (2003) have identified two major genes for Na^+ exclusion, named *Nax1* and *Nax2*. Of these, *Nax1* was located on chromosome 2A and was identified by fine mapping as an Na^+ transporter of the HKT gene family HKT7 (HKT1;4; Huang et al., 2006), whereas *Nax2* was located on chromosome 5A and identified as HKT8 (HKT1;5; Byrt et al., 2007). A locus for Na/K was identified in chromosome 2AS (182.3 Mb) in this study and is different with the *Nax1*. Besides, one locus for K content was identified in the 5B chromosome. The flanking sequence of the SNP *Kukri_c16864_398* and *Tdurum_contig65330_190* was compared with the homologous gene sequence in 5B, and the result demonstrated that the locus identified in this study is different with *Nax2*. Six common loci were identified in the present study in comparison with the QTLs detected previously based on the physical map of IWGSC V1.0 and integrated map of Maccaferri et al. (2015). Li L. et al. (2020) have found 269 associated loci on all 21 chromosomes in wheat for salt-responsive traits based on GWAS in a diverse panel of 323 wheat accessions and 150 doubled haploid lines; these overlapped with *Excalibur_c112015_118* (RSH, 1B), *Ku_c956_1797* (RSH, 1B), *IAAV8839* (RRFW, RRDW, and RSDW, 1D), *Kukri_c16864_398* (RK, 5B), *BobWhite_c46250_98* (Na/K, 7A), and *BS00071025_51* (RSDW and RSDW, 7B) in the present study. Oyiga et al. (2018) have identified a hub locus on chromosome 5B was simultaneously associated with germination vigor and index of dry root weight under different salt stress conditions by GWAS corresponding to *Kukri_c16864_398* for RK content in our study. Also, this locus was identified for germination rate under salt stress by Li L. et al. (2020). The results confirmed that GWAS is a powerful and reliable tool for identification of complex quantitative genes. The alignment of several loci for salt tolerance–related traits with previous studies also serves as a validation of the accuracy and powerfulness of the GWAS. The remaining 10 loci located on chromosomes 2A (*Tdurum_contig11803_306* and *Ku_c1217_312*), 5A (*BS00027465_51*), 5B (*RAC875_c25567_1204*, GENE-3440_199, and *Ku_c15213_388*), 6B (*Ku_c5191_340*), and 7A (*BS00022442_51*, *BS00040929_51*, and BobWhite_c149_3064) are likely novel, which may contribute to uncover the genetic mechanism of salt tolerance, and provide more opportunities for MAS breeding.

To identify candidate genes for the loci of salt tolerance, the sequences of SNPs associated with salt tolerance–related traits were used to BLAST against the National Center for Biotechnology Information. In total, five candidate genes were identified (**Table 4**). The SNP marker *Tdurum_contig11803_306* on chromosome 2AL corresponded to kinase family protein with 2.89 Mb, which may play crucial role in plant responses to salt stress by regulating the hypersensitivity to Na^+ and superfluous

accumulation of Na^+ (Liang et al., 2018). Marker *GENE-3440_199* on chromosome 5B corresponded to E3 ubiquitin-protein ligase-like protein with 1.96 Mb, which plays an important role in plant growth and development (Cyr et al., 2002; Craig et al., 2009). It has been reported that the E3 ubiquitin protein ligase was involved in the regulation of the development of shoot and roots under abiotic stress (Guerra et al., 2012; Wang et al., 2020). Marker *Ku_c5191_340* on chromosome 6B and *BS00022442_51* on chromosome 7A corresponded to transmembrane 9 superfamily member with 2.31 Mb and transmembrane protein with 2.98 Mb, respectively. Various transmembrane transporters, such as H^+/glycerol symporters, Na^+/H^+ antiporters, and the *P*-type ATPases HwENA1/2, either directly or through the electrochemical driving force of the proton gradient to respond to the salt stress in plants (Montpetit et al., 2012). In addition, many *Nax1* and *Nax2* (*HKT1;4* and *HKT1;5*) genes, well known as Na^+ transporter genes, have been cloned and demonstrated to contribute to leaf Na^+ exclusion and salt tolerance in wheat (Byrt et al., 2007, 2014; Tounsi et al., 2016). *Kukri_rep_c101126_469* on 6B encoded leucine-rich repeat receptor-like protein kinase family protein with 1.32 Mb, which may trigger multiple physiological pathways (Montenegro et al., 2017).

For crops, we are concerned with yield much more than other characters. Biomass yield is often taken as a salt tolerance–related indicator because of its permission of the direct estimation of economic return under salt stress (Munns and James, 2003). Thus, in this study, biomass was considered the most important parameter related to growth under salt stress at seedling stage. The number of favorable alleles showed a significant positive effect on RSDW by the linear regression analysis, suggesting that pyramiding of favorable alleles may favor plant salt tolerance. Nine of the 16 loci were identified for two or more traits and should be applicable for MAS breeding. Some accessions with higher salt tolerance and relatively more favorable alleles, such as Xinong 291, Lumai 14, Wengnong 5, and Bima 4, should be excellent germplasms for breeding.

CONCLUSION

In this study, a GWAS for salt tolerance in a diversity panel was conducted with the 90 K SNP chip. In total, 16 loci explained 6.3 to 18.6% of the phenotypic variations, demonstrating that GWAS can be used as a powerful and reliable tool for dissecting complex traits in wheat. The markers and accessions identified in this study can be used as valuable markers and excellent parent materials for salt tolerance breeding. This study improves our understanding of the genetic architecture of salt tolerance in common wheat.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the (**Supplementary Material**) repository, accession number (317).

AUTHOR CONTRIBUTIONS

XQ designed the research, analyzed the physiology data, and drafted the manuscript. XQ, NZ, CX, and HL performed the experiment. JL did the GWAS analysis. JL, YQ, and WH revised the manuscript. All authors have read, edited, and approved the current version of the manuscript.

FUNDING

This work was supported by the Key Research and Development Program of Shandong Province (2019GNC106070) and National Natural Science Foundation of China (31901437, 31871622, and 31872415).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.663941/full#supplementary-material

Supplementary Figure 1 | The LD decay for the 317 wheat accessions.

Supplementary Figure 2 | The frequency distribution of salt tolerance–related traits in 317 wheat accessions. **(a)** Relative shoot dry weight; **(b)** relative shoot fresh weight; **(c)** relative shoot height; **(d)** relative root dry weight; **(e)** relative root fresh weight; **(f)** relative root length; **(g)** relative shoot K content; **(h)** shoot Na content under salt stress; and **(i)** Na/K in shoot under salt stress.

Supplementary Figure 3 | Quantile–quantile (Q-Q) plot for salt tolerance–related traits in 317 wheat accessions by the mixed linear model (MLM) in Tassel v5.1. **(a)** relative shoot dry weight; **(b)** relative shoot fresh weight; **(c)** relative shoot height; **(d)** relative root dry weight; **(e)** relative root fresh weight; **(f)** relative root length; **(g)** relative shoot K content; **(h)** shoot Na content under salt stress; and **(i)** Na/K content in shoot under salt stress.

Supplementary Table 1 | The 317 wheat accessions used in the genome-wide association study (GWAS) for salt tolerance–related traits and their origins.

Supplementary Table 2 | Basic statistical analysis of SNP markers used for genome-wide association study (GWAS) of 317 wheat accessions.

Supplementary Table 3 | The salt tolerance–related traits in 317 wheat accessions.

Supplementary Table 4 | The analysis of variance (ANOVA) for salt tolerance–related traits.

REFERENCES

Ahmadi, N., Negrao, S., Katsantonis, D., Frouin, J., Ploux, J., Letourmy, P., et al. (2011). Targeted association analysis identified japonica rice varieties achieving Na^+/K^+ homeostasis without the allelic make-up of the salt tolerant indica variety Nona Bokra. *Theor. Appl. Genet.* 123, 881–895. doi: 10.1007/s00122-011-1634-4

Akram, S., Rehman-Arif, M. A., and Hameed, A. (2020). A GBS-based GWAS analysis of adaptability and yield traits in bread wheat (*Triticum aestivum* L.). *J. Appl. Genet.* 62, 27–41. doi: 10.1007/s13353-020-00593-1

Asif, M. A., Schilling, R. K., Tilbrook, J., Brien, C., Dowling, K., Rabie, H., et al. (2018). Mapping of novel salt tolerance QTL in an Excalibur× Kukri doubled haploid wheat population. *Theor. Appl. Genet.* 131, 2179–2196. doi: 10.1007/s00122-018-3146-y

Bellucci, A., Torp, A. M., Bruun, S., Magid, J., Andersen, S. B., and Rasmussen, S. K. (2015). Association mapping in scandinavian winter wheat for yield, plant height, and traits important for second-generation bioethanol production. *Front. Plant Sci.* 6:1046. doi: 10.3389/fpls.2015.01046

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

Breseghello, F., and Sorrells, M. E. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172, 1165–1177. doi: 10.1534/genetics.105.044586

Byrt, C. S., Platten, J. D., Spielmeyer, W., James, R. A., Lagudah, E. S., Dennis, E. S., et al. (2007). HKT1;5-like cation transporters linked to Na^+ exclusion loci in wheat, *Nax2* and *Kna1*. *Plant Physiol.* 143, 1918–1928. doi: 10.1104/pp.106.093476

Byrt, C. S., Xu, B., Krishnan, M., Lightfoot, D. J., Athman, A., Jacobs, A. K., et al. (2014). The Na^+ transporter, TaHKT1;5-D, limits shoot Na^+ accumulation in bread wheat. *Plant J.* 80, 516–526. doi: 10.1111/tpj.12651

Cockram, J., White, J., Zuluaga, D. L., Smith, D., Comadran, J., Macaulay, M., et al. (2010). Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21611–21616. doi: 10.1073/pnas.1010179107

Craig, A., Ewan, R., Mesmar, J., Gudipati, V., and Sadanandom, A. (2009). E3 ubiquitin ligases and plant innate immunity. *J. Exp. Bot.* 60, 1123–1132. doi: 10.1093/jxb/erp059

Cyr, D. M., Hohfeld, J., and Patterson, C. (2002). Protein quality control: U-box-containing E3 ubiquitin ligases join the fold. *Trends Biochem. Sci.* 27, 368–375. doi: 10.1016/S0968-0004(02)02125-4

Dababat, A., Rehman-Arif, M. A., Toktay, H., Atiya, O., Shokat, S., E-Orakci, G., et al. (2021). A GWAS to identify the cereal cyst nematode (Heterodera filipjevi) resistance loci in diverse wheat prebreeding lines. *J. Appl. Genet.* 62, 93–98. doi: 10.1007/s13353-020-00607-y

Demiral, T., and Türkan, I. (2006). Exogenous glycinebetaine affects growth and proline accumulation and retards senescence in two rice cultivars under NaCl stress. *Environ. Exp. Bot.* 56, 72–79. doi: 10.1016/j.envexpbot.2005.01.005

Ding, T. L., Yang, Z., Wei, X. C., Yuan, F., Yin, S. S., and Wang, B. S. (2018). Evaluation of salt-tolerant germplasm and screening of the salt-tolerance traits of sweet sorghum in the germination stage. *Funct. Plant. Biol.* 45, 1073–1081. doi: 10.1071/FP18009

Flowers, T. J. (2004). Improving crop salt tolerance. *J. Exp. Bot.* 55, 307–319. doi: 10.1093/jxb/erh003

Füzy, A., Kovács, R., Cseresnyés, I., Parádi, I., Szili−Kovács, T., Kelemen, B., et al. (2019). Selection of plant physiological parameters to detect stress effects in pot experiments using principal component analysis. *Acta Physiol. Plant* 41:56. doi: 10.1007/s11738-019-2842-9

Genc, Y., Oldach, K., Verbyla, A. P., Lott, G., Hassan, M., Tester, M., et al. (2010). Sodium exclusion QTL associated with improved seedling growth in bread wheat under salinity stress. *Theor. Appl. Genet.* 121, 877–894. doi: 10.1007/s00122-010-1357-y

Gerona, M. E. B., Deocampo, M. P., Egdane, J. A., Ismail, A. M., and Dionisio-Sese, M. L. (2019). Physiological responses of contrasting rice genotypes to salt stress at reproductive stage. *Rice Sci.* 26, 207–219. doi: 10.1016/j.rsci.2019.05.001

Guerra, D., Mastrangelo, A. M., Lopez-Torrejon, G., Marzin, S., Schweizer, P., Stanca, A. M., et al. (2012). Identification of a protein network interacting with TdRF1, a wheat RING ubiquitin ligase with a protective role against cellular dehydration. *Plant Physiol.* 158, 777–789. doi: 10.1104/pp.111.183988

Gurung, S., Mamidi, S., Bonman, J. M., Xiong, M., Brown-Guedira, G., and Adhikari, T. B. (2014). Genome-wide association study reveals novel quantitative trait loci associated with resistance to multiple leaf spot diseases of spring wheat. *PLoS One* 9:e108179. doi: 10.1371/journal.pone.0108179

Hasan, M. M., Baque, M. A., Habib, M. A., Yeasmin, M., and Hakim, M. A. (2017). Screening of salt tolerance capability of wheat genotypes under salt stress condition. *Univers. J. Agric. Res.* 5, 235–249. doi: 10.13189/ujar.2017.050405

Holm, S. A. (1979). Simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.

Houston, K., Russell, J., Schreiber, M., Halpin, C., Oakey, H., Washington, J. M., et al. (2014). A Genome wide association scan for (1,3;1,4)-β-glucan content in the grain of contemporary 2-row spring and winter barleys. *BMC Genomics* 15:907. doi: 10.1186/1471-2164-15-907

Huang, S., Spielmeyer, W., Lagudah, E. S., James, R. A., Platten, J. D., Dennis, E. S., et al. (2006). A sodium transporter (HKT7) is a candidate for Nax1, a gene for salt tolerance in durum wheat. *Plant Physiol.* 142, 1718–1727. doi: 10.1104/pp.106.088864

Imlay, J. A. (2003). Pathways of oxidative damage. *Annu. Rev. Microbiol.* 57, 395–418. doi: 10.1146/annurev.micro.57.030502.090938

Kan, G. Z., Zhang, W., Yang, W. M., Ma, D. Y., Zhang, D., Hao, D. R., et al. (2015). Association mapping of soybean seed germination under salt stress. *Mol. Genet. Genomics* 290, 2147–2162. doi: 10.1007/s00438-015-1066-y

Karan, R., and Subudhi, P. K. (2012). "Approaches to increasing salt tolerance in crop plants," in *Abiotic Stress Responses in Plants*, eds P. Ahmad and M. Prasad (New York, NY: Springer), 63–88. doi: 10.1007/978-1-4614-0634-1_4

Kumar, V., Singh, A., Mithra, S. V. A., Krishnamurthy, S. L., Parida, S. K., Jain, S., et al. (2015). Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* 22, 133–145. doi: 10.1093/dnares/dsu046

Li, L., Peng, Z., Mao, X., Wang, J., Li, C., Chang, X., et al. (2020). Genetic insights into natural variation underlying salt tolerance in wheat. *J. Exp. Bot.* 72, 1135–1150. doi: 10.1093/jxb/eraa500

Li, W. H., Zhang, H. Z., Zeng, Y. L., Xiang, L. J., Lei, Z. H., Huang, Q. X., et al. (2020). A salt tolerance evaluation method for sunflower (Helianthus annuus L.) at the seed germination stage. *Sci. Rep.* 10:106261. doi: 10.1038/s41598-020-67210-3

Liang, W., Ma, X., Wan, P., and Liu, L. (2018). Plant salt-tolerance mechanism: a review. *Biochem. Biophys. Res. Commun.* 495, 286–291. doi: 10.1016/j.bbrc.2017.11.043

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Liu, J. D., He, Z. H., Rasheed, A., Wen, W. E., Yan, J., Zhang, P. Z., et al. (2017). A Genome-wide association mapping of black point reaction in common wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 17:220. doi: 10.1186/s12870-017-1167-3

Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282

Liu, P., Jin, Y. R., Liu, J. D., Liu, C. Y., Yao, H. P., Luo, F. Y., et al. (2019). Genome-wide association mapping of root system architecture traits in common wheat (*Triticum aestivum* L.). *Euphytica* 215:121. doi: 10.1007/s10681-019-2452-z

Liu, Y., Liu, Y., Zhang, Q., Fu, B., Cai, J., Wu, J., et al. (2018). Genome-wide association analysis of quantitative trait loci for salinity-tolerance related morphological indices in bread wheat. *Euphytica* 214:176. doi: 10.1007/s10681-018-2265-5

Long, N. V., Dolstra, O., Malosetti, M., Kilian, B., Graner, A., Visser, R. G. F., et al. (2013). Association mapping of salt tolerance in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 126, 2335–2351. doi: 10.1007/s00122-013-2139-0

Ma, L. Q., Zhou, E. F., Huo, N. X., Zhou, R. H., Wang, G. Y., and Jia, J. Z. (2007). Genetic analysis of salt tolerance in a recombinant inbred population of wheat (*Triticum aestivum* L.). *Euphytica* 153, 109–117. doi: 10.1007/s10681-006-9247-8

Maccaferri, M., Ricci, A., Salvi, S., Milner, S. G., Noli, E., Martelli, P. L., et al. (2015). A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. *Plant Biotechnol. J.* 13, 648–663. doi: 10.1111/pbi.12288

Mackay, I., and Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.* 12, 57–63. doi: 10.1016/j.tplants.2006.12.001

Mittler, R. (2002). Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci.* 7, 405–410. doi: 10.1016/S1360-1385(02)02312-9

Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C. K. K., et al. (2017). The pangenome of hexaploid bread wheat. *Plant J.* 90, 1007–1013. doi: 10.1111/tpj.13515

Montpetit, J., Vivancos, J., Mitani-Ueno, N., Yamaji, N., Rémus-Borel, W., Belzile, F., et al. (2012). Cloning, functional characterization and heterologous

expression of *TaLsi1*, a wheat silicon transporter gene. *Plant Mol. Biol.* 79, 35–46. doi: 10.1007/s11103-012-9892-3

Munns, R. (2005). Genes and salt tolerance: bringing them together. *New Phytol.* 167, 645–663. doi: 10.1111/j.1469-8137.2005.01487.x

Munns, R., and James, R. A. (2003). Screening methods for salinity tolerance: a case study with tetraploid wheat. *Plant Soil* 253, 201–218. doi: 10.1023/A:1024553303144

Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Ann. Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911

Munns, R., Rebetzke, G. J., Husain, S., James, R. A., and Hare, R. A. (2003). Genetic control of sodium exclusion in durum wheat. *Aust. J. Agr. Res.* 54, 627–635.

Oyiga, B. C., Sharma, R. C., Baum, M., Ogbonnaya, F. C., Leon, J., and Ballvora, A. (2018). Allelic variations and differential expressions detected at quantitative trait loci for salt stress tolerance in wheat. *Plant Cell Environ.* 41, 919–935. doi: 10.1111/pce.12898

Rehman-Arif, M. A., and Börner, A. (2020). An SNP Based GWAS analysis of seed longevity in wheat. *Cereal Res. Commun.* 48, 149–156. doi: 10.1007/s42976-020-00028-x

Shabala, S., and Cuin, T. A. (2008). Potassium transport and plant salt tolerance. *Physiol. Plant* 133, 651–669. doi: 10.1111/j.1399-3054.2007.01008.x

Shan, C. J., Zhang, S. L., and Ou, X. Q. (2018). The roles of H_2S and H_2O_2 in regulating AsA-GSH cycle in the leaves of wheat seedlings under drought stress. *Protoplasma* 255, 1257–1262. doi: 10.1007/s00709-018-1213-5

Shi, Y. Y., Gao, L. L., Wu, Z. C., Zhang, X. J., Wang, M., Zhang, C. S., et al. (2017). Genome-wide association study of salt tolerance at the seed germination stage in rice. *BMC Plant Biol.* 17:92. doi: 10.1186/s12870-017-1044-0

Singh, S., Sehgal, D., Kumar, S., Rehman-Arif, M. A., Vikram, P., Sansaloni, C. P., et al. (2020). GWAS revealed a novel resistance locus on chromosome 4D for the quarantine disease Karnal bunt in diverse wheat pre-breeding germplasm. *Sci. Rep.* 10:5999. doi: 10.1038/s41598-020-62711-7

Song, J., and Wang, B. S. (2015). Using euhalophytes to understand salt tolerance and to develop saline agriculture: *Suaeda salsa* as a promising model. *Ann. Bot.* 115, 541–553. doi: 10.1093/aob/mcu194

Stich, B., Mohring, J., Piepho, H. P., Heckenberger, M., Buckler, E. S., and Melchinger, A. E. (2008). Comparison of mixed-model approaches for association mapping. *Genetics* 178, 1745–1754. doi: 10.1534/genetics.107.079707

Szira, F., Bálint, A. F., Börner, A., and Galiba, G. (2008). Evaluation of drought-related traits and screening methods at different developmental stages in spring barley. *J. Agron. Crop Sci.* 194, 334–342. doi: 10.1111/j.1439-037X.2008.00330.x

Tounsi, S., Ben Amar, S., Masmoudi, K., Sentenac, H., Brini, F., and Very, A. A. (2016). Characterization of two HKT1;4 transporters from *Triticum monococcum* to elucidate the determinants of the wheat salt tolerance *Nax1* QTL. *Plant Cell Physiol.* 57, 2047–2057. doi: 10.1093/pcp/pcw123

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotech. J.* 12, 787–796. doi: 10.1111/pbi.12183

Wang, W. L., Wang, W. Q., Wu, Y. Z., Li, Q. X., Zhang, G. Q., Shi, R. R., et al. (2020). The involvement of wheat U-box E3 ubiquitin ligase TaPUB1 in salt stress tolerance. *J. Integr. Plant Biol.* 62, 631–651. doi: 10.1111/jipb.12842

Winfield, M. O., Allen, A. M., Burridge, A. J., Barker, G. L., Benbow, H. R., Wilkinson, P. A., et al. (2015). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* 14, 1195–1206. doi: 10.1111/pbi.12485

Wu, D. Z., Sato, K., and Ma, J. F. (2015). Genome-wide association mapping of cadmium accumulation in different organs of barley. *New Phytol.* 208, 817–829. doi: 10.1111/nph.13512

Xu, Y., Li, S., Li, L., Zhang, X., Xu, H., and An, D. (2013). Mapping QTLs for salt tolerance with additive, epistatic and QTL treatment interaction effects at seedling stage in wheat. *Plant Breed.* 132, 276–283. doi: 10.1111/pbr.12048

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2020). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *bioRxiv* [Preprint] doi: 10.1101/2020.08.20.258491

Yong, H. Y., Wang, C. L., Bancroft, I., Li, F., Wu, X. M., Kitashiba, H., et al. (2015). Identification of a gene controlling variation in the salt tolerance of rapeseed (*Brassica napus* L.). *Planta* 242, 313–326. doi: 10.1007/s00425-015-2310-8

Zeng, L., Poss, J. A., Wilson, C., Draz, A. E., Gregorio, G. B., and Grieve, C. M. (2003). Evaluation of salt tolerance in rice genotypes by physiological characters. *Euphytica* 129, 281–292. doi: 10.1023/A:1022248522536

Zhao, J., Paulo, M. J., Jamar, D., Lou, P., van Eeuwijk, F., Bonnema, G., et al. (2007). Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*. *Genome* 50, 963–973. doi: 10.1139/G07-078

Zhao, J., Sun, H., Dai, H., Zhang, G., and Wu, F. (2010). Difference in response to drought stress among Tibet wild barley genotypes. *Euphytica* 172, 395–403. doi: 10.1007/s10681-009-0064-8

Zhu, J. K. (2003). Regulation of ion homeostasis under salt stress. *Curr. Opin. Plant Biol.* 6, 441–445. doi: 10.1016/S1369-5266(03)00085-2

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fgene.2021.666440

Identification and Validation of Major QTLs, Epistatic Interactions, and Candidate Genes for Soybean Seed Shape and Weight Using Two Related RIL Populations

Mahmoud A. Elattar[1,2], Benjamin Karikari[1†], Shuguang Li[1], Shiyu Song[1], Yongce Cao[1], Muhammed Aslam[1], Aiman Hina[1], Salah Fatouh Abou-Elwafa[3] and Tuanjie Zhao[1]**

[1] National Center for Soybean Improvement, Key Laboratory of Biology and Genetic Improvement of Soybean (Ministry of Agriculture), State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China, [2] Agronomy Department, Faculty of Agriculture, Minia University, Minia, Egypt, [3] Agronomy Department, Faculty of Agriculture, Assiut University, Assiut, Egypt

Understanding the genetic mechanism underlying seed size, shape, and weight is essential for enhancing soybean cultivars. High-density genetic maps of two recombinant inbred line (RIL) populations, LM6 and ZM6, were evaluated across multiple environments to identify and validate M-QTLs as well as identify candidate genes behind major and stable quantitative trait loci (QTLs). A total of 239 and 43 M-QTLs were mapped by composite interval mapping (CIM) and mixed-model-based composite interval mapping (MCIM) approaches, from which 180 and 18, respectively, are novel QTLs. Twenty-two QTLs including four novel major QTLs were validated in the two RIL populations across multiple environments. Moreover, 18 QTLs showed significant AE effects, and 40 pairwise of the identified QTLs exhibited digenic epistatic effects. Thirty-four QTLs associated with seed flatness index (FI) were identified and reported here for the first time. Seven QTL clusters comprising several QTLs for seed size, shape, and weight on genomic regions of chromosomes 3, 4, 5, 7, 9, 17, and 19 were identified. Gene annotations, gene ontology (GO) enrichment, and RNA-seq analyses of the genomic regions of those seven QTL clusters identified 47 candidate genes for seed-related traits. These genes are highly expressed in seed-related tissues and nodules, which might be deemed as potential candidate genes regulating the seed size, weight, and shape traits in soybean. This study provides detailed information on the genetic basis of the studied traits and candidate genes that could be efficiently implemented by soybean breeders for fine mapping and gene cloning, and for marker-assisted selection (MAS) targeted at improving these traits individually or concurrently.

Keywords: *Glycine max*, QTL mapping, QTL clusters, marker assisted breeding, epistatic interactions, candidate genes

OPEN ACCESS

Edited by:
Ahmad M. Alqudah, Martin Luther University of Halle-Wittenberg, Germany

Reviewed by:
Shouvik Das, Indian Agricultural Research Institute (ICAR), India
Kaixin Zhang, Yancheng Institute of Technology, China

***Correspondence:**
Salah Fatouh Abou-Elwafa
elwafa75@aun.edu.eg
Tuanjie Zhao
tjzhao@njau.edu.cn

†Present address:
Benjamin Karikari, Department of Crop Science, Faculty of Agriculture, Food and Consumer Sciences, University for Development Studies, Tamale, Ghana

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *10 February 2021*
Accepted: *29 March 2021*
Published: *28 May 2021*

Citation:
Elattar MA, Karikari B, Li S, Song S, Cao Y, Aslam M, Hina A, Abou-Elwafa SF and Zhao T (2021) Identification and Validation of Major QTLs, Epistatic Interactions, and Candidate Genes for Soybean Seed Shape and Weight Using Two Related RIL Populations. Front. Genet. 12:666440. doi: 10.3389/fgene.2021.666440

INTRODUCTION

Soybean (*Glycine max* L. Merr.) is one of the most important food crops, being a rich source of dietary protein (69%) and providing over 50% edible oil globally, and has a significant role in health and biofuel (Hoeck et al., 2003). Besides improving soil fertility by integrating atmospheric nitrogen in the soil through a synergistic interaction with microorganisms, because of its high nutritional value, soybean is used in human food and animal feed (Wang et al., 2019). Throughout the last five decades, soybean production in China slightly increased. To meet domestic demands, China imports almost 80% of its requirements of soybean; therefore, improving soybean production is a major aim of soybean breeders to make the country self-sufficient (Liu et al., 2018). Most plant breeders are targeting yield-related traits to improve soybean production.

Seed size is an essential trait in flowering plants and plays a critical role in adaptation to the environment (Tao et al., 2017). However, these traits are complex quantitative traits regulated by polygenes and strongly influenced by environment and genotype × environment (G × E) interaction, and hence it is more difficult to select for based on phenotype (Yao et al., 2014). All soybean varieties developed in tropical and subtropical countries have small seed size compared to the temperate-region varieties. Besides, seed size, shape, and weight are important seed quality traits with significant influence on seed use (Basra, 1995; Teng et al., 2017; Wu et al., 2018). A positive correlation between seed size/weight and seed yield has been reported in several studies. Seed size/weight revealed a positive association with seed germination capability and vigor, thereby significantly affecting the competitive capability of the seedling for nutrient and water resources and light, hence enhancing stress tolerance (Edwards and Hartwig, 1971; Haig, 2013). Dissecting the genetic factors underlying seed size, shape, and weight and their relationship to the ambient environment is essential for improving soybean yield and quality-related traits. In addition, understanding the additive and additive × environment (AE) effects of quantitative trait loci (QTLs) and their contribution to the phenotypic variations would facilitate the application marker-assisted selection (MAS) because it will prominently lead the breeders in the QTL selection and expectation of the outcomes of MAS (Jannink et al., 2009).

A major aim of utilizing linkage mapping in plant breeding is to deepen our understanding of the inheritance and genetic architecture of quantitative traits and detect markers that can be employed as indirect selection tools in plant breeding (Bernardo, 2008; Abou-Elwafa, 2016a). In this regard, QTL mapping has been regularly used for detecting the QTL/gene underlying the quantitative traits such as seed size, shape, and weight in crop plants. As known, parental diversity and marker density greatly influence the accuracy and precision of QTL mapping. Besides, the population size used in most of the previously published reports for genetic mapping studies usually varied from 50 to 250 individuals; however, larger populations are needed for high-resolution mapping. A high-density genetic map facilitates the detection of closely linked markers associated with QTLs and provides an effective base for investigating quantitative traits (Mohan et al., 1997; Galal et al., 2014; Tewodros and Zelalem, 2016). The statistical difference between phenotypic data obtained from various environments could enhance the accuracy to detect QTL position (Zhao and Xu, 2012). Previous studies identified important seed size and shape QTLs, which were also associated with hundred seed weight (HSW); however, most of the studies used low-density genetic maps based on restriction fragment length polymorphism, simple-sequence repeat markers, and biochemical and morphological markers which have a large confidence interval with a low resolution of QTLs that are not suitable for candidate gene identification (Bernardo, 2008; Han et al., 2012; Abou-Elwafa, 2016a,b). Therefore, it is crucial to employ high-density genetic maps to detect more new recombination in a population, which will increase the accuracy of QTL mapping, candidate gene identification, and MAS (Mahmoud et al., 2018; Cao et al., 2019; Hina et al., 2020). Recent advances in genetic and genomic tools and approaches have facilitated the identification of QTLs associated with various agronomic traits in different crop species, including soybean, and the identification of candidate genes underlying these genomic regions (Peterson et al., 2012; Sun et al., 2012; Xie et al., 2018; Zhang F. et al., 2019; Hina et al., 2020; Kajiya-Kanegae et al., 2020). Although epistatic interaction has a stronger effect on inbreeding depression, heterosis, adaptation, speciation, and reproductive isolation (Ma et al., 2015), previous studies focused mostly on identifying main-effect QTLs associated with seed sizes, shapes, and 100-seed weight in soybean. To date, at least 441, 52, and 297 QTLs for seed size, shape, and HSW have been reported[1] based on various genetic contexts, advances in marker technology, statistical methods, and multiple environments. However, most of these QTLs are minor (R^2 < 10%), not stable, and with larger genomic regions/confidence intervals (Han et al., 2012; Hu et al., 2013; Kato et al., 2014). Recently, there have been limited studies on detecting QTLs with epistatic effects and their interactions with the environment (QEs) (Panthee et al., 2005; Zhang et al., 2011, 2018; Liang et al., 2016). Knowledge about the molecular mechanisms underlying soybean seed size, shape, and weight is still limited. So far, only two seed sizes/weight-related genes have been cloned and characterized from the soybean, i.e., the *Glyma20g25000 (ln)* gene that has a significant impact on seed size and number of seeds per pod (Jeong et al., 2012) and the *PP2C-1* gene that enhances seed size/weight (Lu et al., 2017). Therefore, it is essential to identify major and stable QTLs and candidate genes related to seed size, shape, and weight to improve our understanding of genetic mechanisms controlling these important traits in soybean (Kato et al., 2014; Zhang et al., 2018). The present study aimed to (i) map main-effect QTLs (M-QTLs), additive × additive (AA) QTLs, and QE for seed size, shape, and weight traits; (ii) employ two QTL mapping approaches to validate the identified QTLs in two mapping populations across multiple environments; (iii) analyze the epistatic QTL pairs and their interactions with the environment for further utilization of

[1] www.soybase.org

these QTLs in soybean genetic improvement; and (iv) mine potential candidate genes for the major ($R^2 > 10\%$) and stable (identified across multiple environments or populations) QTLs. We hypothesize that the results of this study would provide comprehensive knowledge on the genetic bases for these traits and mined candidate genes would serve as a foundation for functional validation and verification of some genes for seed size, shape, and weight in soybean. Besides, the results would be useful for the application of marker-assisted breeding (MAB) in soybean.

MATERIALS AND METHODS

Plant Materials and Experiments

Two recombinant inbred line (RIL) populations, i.e., ZM6 and LM6, comprising 126 and 104 lines, respectively, were used in the present study. The two populations were developed by single seed descent (SSD) from crosses between the genotypes Zhengyang (Z) and Linhefenqingdou (L) as female parents and the M8206 (M6) genotype as the male parent. The two female parents, Z and L, have an average 100-seed weight of 17.1 and 35 g, respectively, whereas the male parent has an average 100-seed weight of 13.7 g.

The two RIL populations along with their parents were evaluated for seed size, shape and HSW across multiple environments. Experiments were conducted in the Jiangpu Experimental Station (33° 030' N and 63° 118' E), Nanjing, Jiangsu Province, in the 2012, 2013, 2014, and 2017 growing seasons (designated as 12JP, 13JP, 14JP, and 17JP, respectively), the Fengyang Experimental Station, Chuzhou, Anhui Province (32° 870' N and 117° 560' E), in the 2012 growing season (designated as 12FY), and the Yancheng Experimental Station, Yancheng, Jiangsu Province (33° 410' N and 120° 200' E), in 2014 (designated as 14YC). Plants were sown in June and harvest was done in October of the same year. Experiments were designed in a randomized complete blocks design (RCBD) with three replications. The experimental plot was one row of 2-m-long at 5-cm plant-to-plant distance and 50-cm row-to-row distance. Planting and post-planting operations were carried out following the recommended agronomic practices.

Phenotypic Evaluation and Statistical Analysis

Eight seed-related traits including seed length (SL), seed width (SW), seed thickness (ST), seed length/width (SLW), seed length/thickness (SLT), seed width/thickness (SWT), flatness index (FI), and 100-seed weight (HSW) were evaluated in LM6 and ZM6 populations under all environments. Phenotypic data were measured and recorded according to standard procedures (Tomooka et al., 2002; Cheng et al., 2006). In brief, seeds harvested from 10 guarded plants in the middle of each row were used for estimating SL, SW, ST, and HSW. The SL was measured as the longest dimension over the seed equivalent to the hilum. SW was measured as the longest dimension across the seed vertical to the hilum. ST was measured as the longest dimension from top to bottom of the seed. The SL, SW, and ST were estimated in millimeters (mm) using the Vernier caliper instrument, according to Kaushik et al. (2007) (**Figure 1**). Seed shape was identified by calculating three different ratios, i.e., SL/SW (SLW), SL/ST (SLT), and SW/ST (SWT), and FI. The ratios between the SL, SW, and ST were estimated from the individual values of the length, width, and thickness of the seeds according to Omokhafe and Alika (2004), while FI was calculated following the formula elaborated by Cailleux (1945) and Cerdà and Garcıa-Fayos (2002) to describe seed shape:

$$FI = \frac{(L+W)}{2T}$$

where L is the SL, W is the SW, and T is the ST.

It extended from a value of 1 for the round seeds to more than 2 for skinny seeds. The HSW was expressed as an average of five measurements of 100 randomly selected seeds.

Descriptive statistics of the seed size, seed shape, and HSW traits were calculated using the SPSS software, version 24[2]. Analysis of variance (ANOVA) for each environment and the combined overall environments (CE) was performed using the PROC GLM procedure in SAS software based on the random model (SAS Institute Inc. v. 9.02, 2010, Cary, NC, United States). Broad-sense heritability (h^2) in individual environments was estimated as:

$$h^2 = \sigma_g^2/(\sigma_g^2+\sigma_e^2)$$

[2]http://www.spss.com

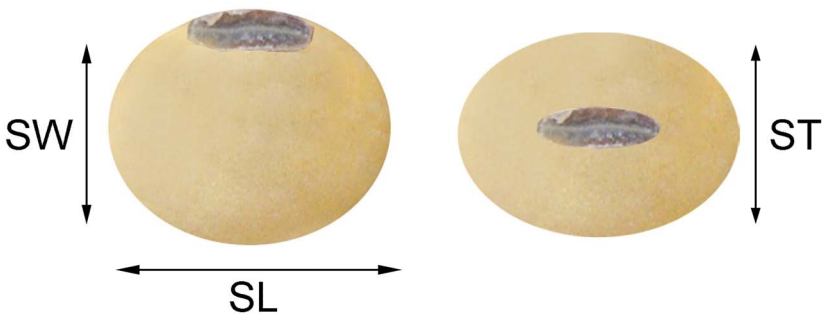


FIGURE 1 | Measuring seed width (SW), length (SL), and thickness (ST).

whereas in the CE h^2 was estimated as follows:

$$h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/nr)$$

where $\sigma^2{}_g$, $\sigma^2{}_e$, and $\sigma^2{}_{ge}$ are the variance components estimated from the ANOVA for the genotypic, error, and genotype × experiment variances, respectively, with r as the number of replicates and n as the number of environments. All the parameters were assessed from the expected mean squares in ANOVA. The Pearson correlation coefficient (r) between seed size, seed shape, and HSW traits was calculated from the mean data using the SAS PROC CORR with data obtained for CE (average across environments) for each population.

Construction of Genetic Maps and QTL Analysis

Genetic map information was obtained from the National Center for Soybean Improvement, Nanjing Agricultural University. High-density genetic maps of the ZM6 and LM6 populations comprise 2601 and 2267 bin markers by using the RAD-seq technique, respectively (**Supplementary Table 1**), which were constructed as previously reported (Karikari et al., 2019; Zhang X. et al., 2019). The total lengths of the ZM6 and LM6 genetic maps were 2630.22 and 2453.79 cM, with an average distance between the markers 1.01 and 1.08 cM, respectively (**Supplementary Table 1**). The average marker per chromosome was 130 and 113 for the ZM6 and LM6 maps, respectively, with average genetic distances per chromosome of 131.51 and 122.69 cM, respectively (**Figure 2** and **Supplementary Table 1**).

Mapping of Main- and Epistatic-Effect QTLs

The WinQTLCart 2.5 software (Wang et al., 2006) was employed to identify the M-QTLs using the average values of seed size, seed shape, and 100-seed weight from the individual environments and overall environments with the composite interval mapping model (CIM) (Zeng, 1994). The software running features were 10 cM window size, 1 cM running speed, the logarithm of odds (LOD) (Morton, 1955) threshold which was computed using 1000 permutations because of an experiment-wide error proportion of $P < 0.05$ (Churchill and Doerge, 1994), and the confidence interval which was determined using a 1-LOD support interval, which was controlled by finding the local on the two sides of a QTL top that is compatible with a reduction of the 1 LOD score. QTLs detected within overlapping intervals in different environments were considered the same (Qi et al., 2017). To identify the genetic effects of the QTLs, i.e., additive QTLs, additive × additive (AA), AE, and AA × environment (AAE), the mixed-model-based composite interval mapping (MCIM) procedure was employed in the QTLNetwork V2.1 software (Yang et al., 2008). The critical F-value was calculated by a permutation test with 1000 permutations for MCIM. The effects of QTLs were assessed using the Markov Chain Monte Carlo (MCMC) approach. Epistatic effects, candidate interval selection, and putative QTL detection were estimated with an experiment-wide error proportion of $P < 0.05$ (Yang et al., 2007; Xing et al., 2012).

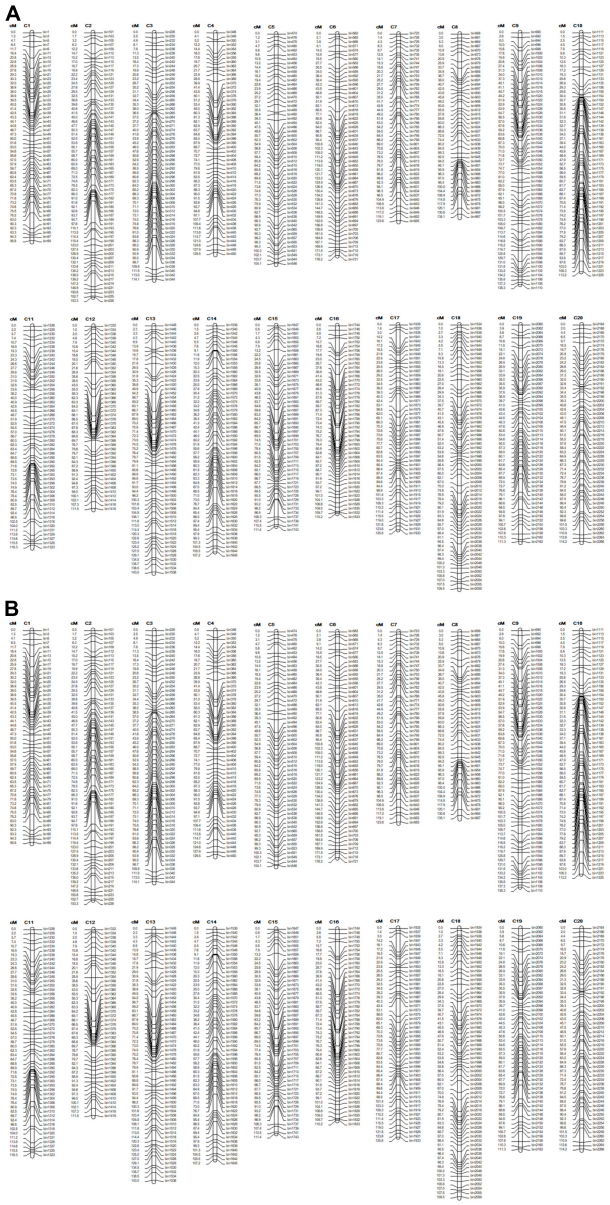


FIGURE 2 | The genetic linkage map of two soybean RIL populations. **(A)** The genetic linkage map for the LM6 population, and **(B)** the genetic linkage map for the ZM6 population. Map distances are shown in centimorgans (cM).

In silico Identification of Candidate Genes

QTLs identified in two or more environments with an $R^2 > 10\%$ were considered as stable and major QTLs (Qi et al., 2017). Genomic regions with several M-QTLs related to different studied traits were identified as a QTL cluster.

The Phytozome[3] and SoyBase (see text footnote 1) online platform repositories were employed to retrieve all model genes within the physical interval position of the QTL clusters. Potential candidate genes were identified based on gene annotations (see text footnotes 1 and 3) and the reported putative function of genes implicated in these traits. Gene ontology (GO) enrichment analysis was performed for the identified candidate genes within each QTL cluster region using AgriGO V2.0[4] (Tian et al., 2017). Gene classification was then carried out using the Web Gene Ontology (WeGO) Annotation Plotting tool, Version 2.0 (Ye et al., 2006). The publicly available RNA-Seq database on the SoyBase website was used to analyze the expression of the candidate genes in various soybean tissues and developmental stages. A heat map to visualize the fold-change patterns of these candidate genes was constructed using the TBtools_JRE 1.068 software (Chen C. et al., 2020).

RESULTS

Phenotypic Variations in RIL Populations in Multiple Environments

All measured (SL, ST, SW, and HSW) and calculated (SLW, SLT, SWT, and FI) phenotypic traits exhibited significant differences among the three parental lines across all environments as indicated by the ANOVA (**Supplementary Tables 2** and **3**). ANOVA revealed that all studied traits were significantly ($P < 0.001$ or <0.05) influenced by the environment, genotypes, and the genotype × environment interaction (**Supplementary Tables 4**, **5**), indicating the differential response of the genotypes to the changes in environmental cues. The two populations showed continuous phenotypic variations in all studied traits, implying a polygenic inheritance of these traits (**Figure 3**). The differences in mean phenotypic values among the three parental lines for seed size, seed shape, and HSW traits were constantly high across all studied environments, and their multi-environment means for both populations (**Figure 4**). Compared to the male parent, the female parent of the LM6 population, Linhefenqingdou, exhibited an average increase of 27.80, 28.19, 31.10, and 41.37% in SL, ST, SW, and HSW, respectively. Meanwhile, in the ZM6 population the female parent Zhengyang surpassed the male parent M8206 by an average of 11.00, 9.66, 7.65, and 17.53% in SL, ST, SW, and HSW across all environments, respectively (**Figure 4**, **Supplementary Tables 2**, **3**). In both populations, several lines overstep their parents in both directions in all studied traits across all environments, suggesting the occurrence of transgressive segregations within the two populations (**Figures 3**, **4**). The broad-sense heritability (h^2) under individual environments ranged from 66.75 to 98.08%, 64.39 to 95.72%, and 81.45 to 99.36% for seed size, HSW, and seed shape, respectively (**Supplementary Tables 2**, **3**). Meanwhile, h^2 under combined environments (CE) ranged from 78.25 to 87.31%, 65.70 to 90.80%, and 92.95 to 95.72% for seed size, shape, and HSW, respectively. The correlation coefficient (r^2) among SL, ST, and SW exhibited significant positive correlations with each other and with two of the seed shape traits (SLT and SLW) in both populations with r^2 values ranging from 0.79 to 0.91. Meanwhile, SL, ST, and SW exhibited significant negative correlations with the other two seed shape traits (SWT and FI) (**Supplementary Table 6**). Except for the correlation between SLW and SWT, all the seed shape traits showed significant positive correlations with each other in both populations with r^2 values ranged from 0.33 to 0.95. All seed size traits, i.e., SL, SW, and ST, showed significant positive correlations with HSW with r^2 values ranging from 0.29 to 0.70 in both populations.

Validation of Identified QTLs Employing Two Mapping Approaches

A total of 92, 99, and 48 M-QTLs associated with seed size, seed shape, and HSW, respectively, were mapped by the CIM approach (**Supplementary Tables 7–9**). Meanwhile, forty-three QTLs were identified for seed size, shape, and HSW by using MCIM approach (**Tables 1**, **2**). Among these, 22 QTLs were identified and validated by both approaches within the same physical chromosomal position, indicating the dependability and stability of these QTLs. A comparison of the physical chromosomal regions of the QTLs detected by both approaches revealed the identification and validation of four QTLs, i.e., $qSL\text{-}7\text{-}1_{LM6}$, $qSW\text{-}19\text{-}2_{LM6}$, $qFI\text{-}3\text{-}1_{LM6}$, and $qHSW\text{-}3\text{-}2_{LM6}$, for the first time in the two populations (LM6 and ZM6) with an $R^2 > 10\%$. Therefore, we considered these QTLs as novel stable and major QTLs that could be used for map-based cloning, candidate gene identification, and QTL stacking into elite cultivars targeted at improving seed size, shape, and HSW in soybean.

Identification of the Main Effects of the Stable Additive Seed Size QTLs

A total of 92 M-QTLs were mapped for seed size-related traits, i.e., SL, SW, and ST, on all soybean chromosomes, except chromosomes 1 and 12, with LOD scores and R^2 values ranging from 2.5 to 10.3 and 5.0 to 19.7%, respectively, in the two populations (**Supplementary Table 7** and **Supplementary Figures 1a–c**). Out of these, 30 M-QTLs for SL, 35 for SW, and 27 for ST with alleles underlying QTLs inherited from either of the parents. Seventy-two M-QTLs were mapped in one environment while the remaining 20 were mapped within overlapping regions in at least one environment with or without CE. Forty-seven QTLs exhibiting $R^2 > 10\%$ were considered as major QTLs. The most prominent QTL was the $qSW\text{-}17\text{-}2_{LM6}$ (LOD = 6.70–10.29, and R^2 = 16.60–18.30%), which was detected within the physical position 6,844,412–9,645,325 bp in 14JP and CE (**Figure 5b**). Likewise, the $qSL\text{-}10\text{-}2_{ZM6,LM6}$ (LOD = 6.08–6.89, and R^2 = 15.4–17.1% in ZM6 (17JP) and LM6

[3]http:/phytozome.jgi.doe.gov

[4]http:/systemsbiology.cau.edu.cn

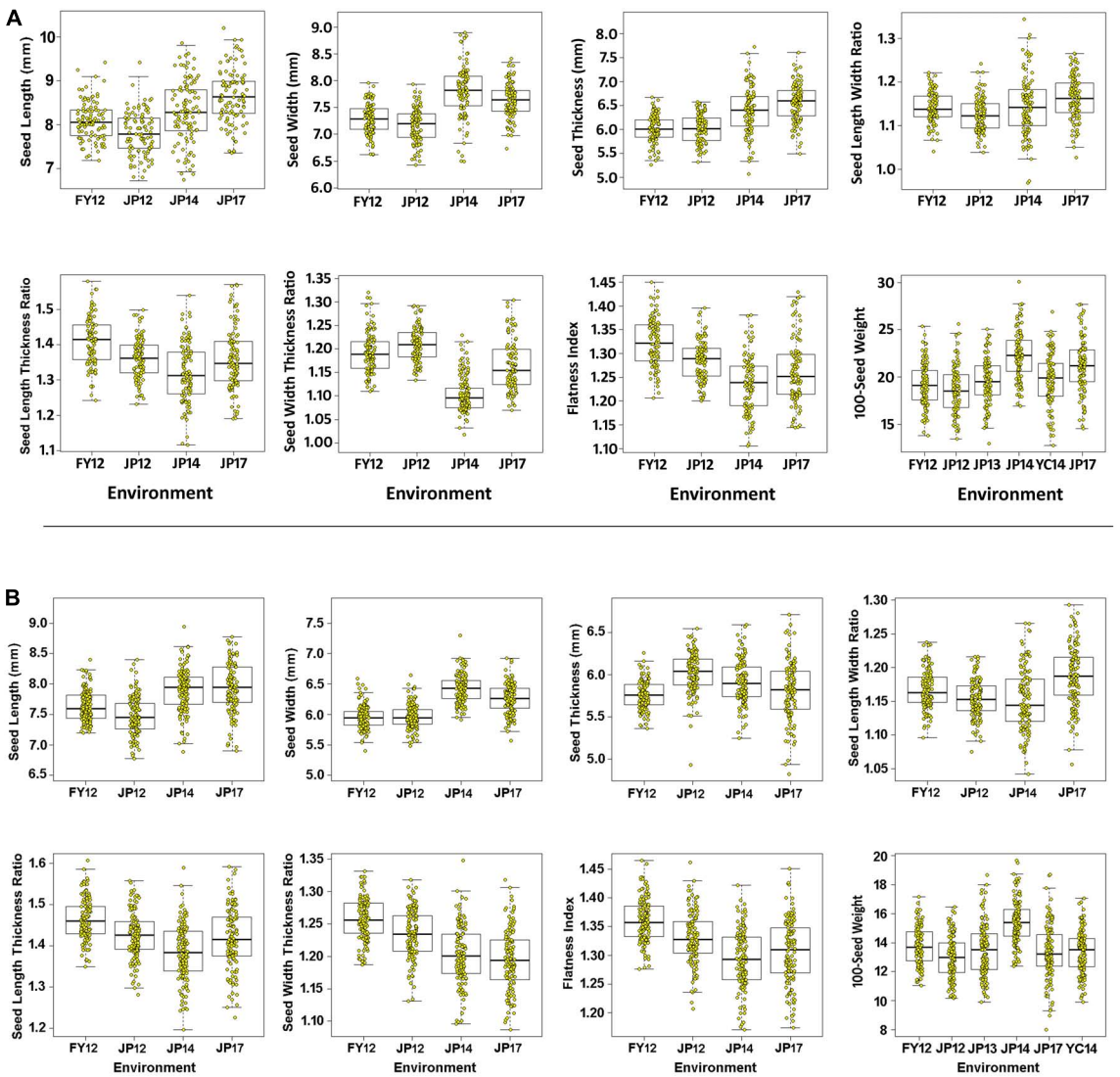


FIGURE 3 | Boxplot for seed size, seed shape, and 100-seed weight traits. The black line in the middle of the box shows the median, the white box indicates the range from the lower quartile to the upper quartile, and the dashed black line and yellow dots represent the dispersion and frequency distribution of the phenotypic data in each of the six environments, i.e., 12FY, 12JP, 13JP, 14JP, 14YC, and 17JP, while **(A,B)** represent LM6 and ZM6 populations.

(14JP) populations) was located to the physical position between 41,454,163 and 43,944,243 bp.

Main Effects of the Stable Additive Seed Shape QTLs

In total, 99 M-QTLs related to seed shape traits (SLT, SLW, SWT, and FI) were mapped to 19 soybean chromosomes excluding chromosome 2 in both populations across four environments and the CE with LOD scores of 2.50–10.44 and R^2 of 5.12–31.56% by the CIM approach (**Supplementary Table 8** and **Supplementary Figures 1d–g**). From the 99 M-QTLs, 22, 33, 11, and 22 were detected for SLT, SLW, SWT, and FI, respectively (**Supplementary Table 8**). Among them, 71 M-QTLs were detected in multiple environments, while 28 were mapped in at least one environment either with or without the CE. Eight M-QTLs for SLW (*qSLW-3-2*$_{LM6}$, *qSLW-5-3*$_{ZM6,LM6}$, *qSLW-9-3*$_{LM6,ZM6}$, *qSLW-13-4*$_{ZM6,\ LM6}$, *qSLW-15-1*$_{LM6}$, *qSLW-15-3*$_{LM6}$, *qSLW-16-2*$_{ZM6}$, and *qSLW-16-3*$_{ZM6}$) were mapped in at least one environment with or without the CE. Similarly, seven M-QTLs for SLT (*qSLT-1-3*$_{LM6}$, *qSLT-5-3*$_{ZM6}$, *qSLT-11-1*$_{LM6}$, *qSLT-13-1*$_{LM6}$, *qSLT-14-1*$_{LM6}$, *qSLT-16-1*$_{LM6}$, and *qSLT-20-1*$_{ZM6}$) were mapped in at least one environment with or without the CE (**Supplementary Table 8** and **Supplementary Figure 1e**). Likewise, four M-QTLs (*qSWT-8-1*$_{LM6}$, *qSWT-11-2*$_{ZM6,LM6}$, *qSWT-13-1*$_{ZM6}$, and *qSWT-17-1*$_{ZM6}$) were mapped for SWT in at least one environment with or without the CE (**Supplementary Table 8** and **Supplementary Figure 1f**). Also, a total of 10 M-QTLs (*qFI-1-1*$_{ZM6}$, *qFI-1-2*$_{ZM6,LM6}$, *qFI-1-3*$_{LM6}$, *qFI-1-4*$_{ZM6,LM6}$, *qFI-3-1*$_{LM6}$, *qFI-3-3*$_{ZM6}$, *qFI-5-2*$_{ZM6}$, *qFI-11-1*$_{LM6}$,

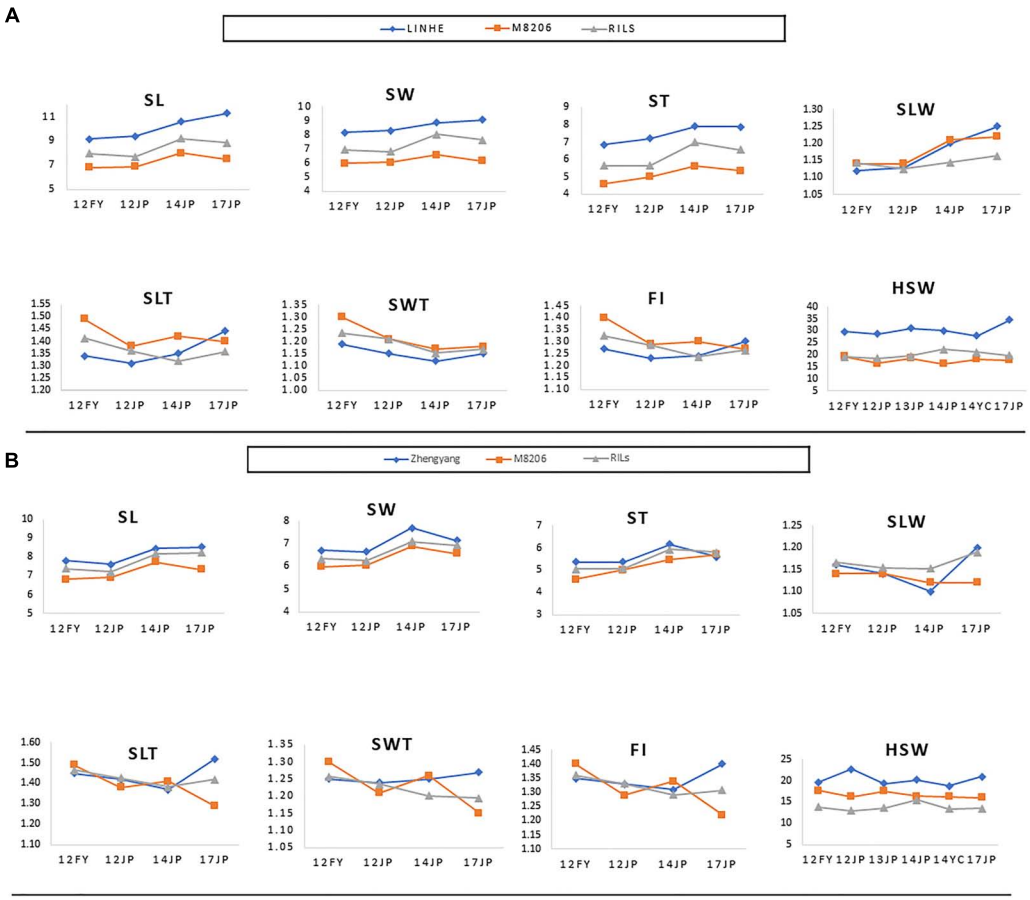


FIGURE 4 | Performance of the parents of the two RIL populations, i.e., LINHE and M8206 **(A)** and Zhengyang and M8206 **(B)** along with the two derived RIL populations, LM6 and ZM6, respectively for seed size and shape traits as well as 100-seed weight among multiple environments. 12JP, 13JP, 14JP, and 17JP indicate phenotyping at the Jiangpu Experimental Station in the 2012, 2013, 2014, and 2017 growing seasons, respectively. 12FY indicates the Fengyang Experimental Station, Chuzhou in the 2012 growing season. 14YC indicates the Yancheng Experimental Station in 2014. SL, seed length (mm); SW, seed width (mm); ST, seed thickness (mm); SLW, seed length to width ratio; SLT, seed length to thickness ratio; SWT, seed width to thickness ratio; FI, flatness index; HSW, 100-seed weight (g).

qFI-14-1$_{LM6}$, *qFI-17-1*$_{ZM6}$, and *qFI-20-1*$_{ZM6}$) were considered as stable QTLs as they were detected in multiple environments. Several physical regions which harbored at least two seed shape-related traits were identified, e.g., the chromosomal region between 1,730,667 and 3,014,518 bp on chr01 harbors *qSLT-1-2*$_{ZM6}$, *qSWT-1-1*$_{ZM6}$, and *qFI-1-2*$_{ZM6,LM6}$, and the chromosomal region between 4,946,300 and 35,955,471 bp on chr01 which comprises *qSLT-1-3*$_{LM6}$, *qSWT-1-2*$_{LM6}$, and *qFI-1-3*$_{LM6}$ (**Supplementary Figure 2** and **Supplementary Table 8**). Two M-QTLs (*qSLW-3-2*$_{LM6}$ and *qFI-3-1*$_{LM6}$) located to the physical region between 1,509,548 and 3,515,594 bp on chr03. The QTLs *qSLT-5-3*$_{ZM6}$, *qFI-5-2*$_{ZM6}$, and *qSLW-5-3*$_{ZM6,LM6}$ reside the physical region between 38,035,798 and 41,186,985 bp on chr05 (**Supplementary Figure 2** and **Supplementary Table 8**). The *qSLT-11-1*$_{LM6}$ and *qFI-11-1*$_{LM6}$, *qSLT-13-1*$_{LM6}$, *qSLW-13-4*$_{ZM6,LM6}$ and *qSWT-13-1*$_{ZM6}$, and *qFI-14-1*$_{LM6}$, *qSLT-14-1*$_{LM6}$ and *qFI-14-1*$_{LM6}$ were located to the physical regions of 17,145,381–23,469,672, 33,303,067–39,562,563, and 3,468,251–8,668,367 bp on chr11, chr13, and chr14, respectively (**Supplementary Figure 2** and **Supplementary Table 8**). The *qSLW-16-2*$_{ZM6}$, *qSLT-16-1*$_{LM6}$ and *qSLW-16-3*$_{ZM6}$, *qFI-17-1*$_{ZM6}$ and *qSWT-17-1*$_{ZM6}$, and *qSLT-20-1*$_{ZM6}$ and *qFI-20-*1$_{ZM6}$ were located to the chromosomal regions of 26,903,205–31,959,397 on Chr16, 40,207,655–41,672,092 bp on Chr17, and 1–1,115,156 bp on Chr20, respectively (**Supplementary Figure 2** and **Supplementary Table 8**).

Main Effects of the Stable Additive Seed Weight QTLs

A total of 48 M-QTLs for HSW were detected, from which 27 were detected in a specific environment and 21 were mapped in at least one environment with or without the CE (**Supplementary Table 9** and **Supplementary Figure 1h**). The LOD scores and R^2 values of these M-QTLs ranged from 2.51 to 10.61 and 4.8 to 24.5%, respectively. The highest number of M-QTLs (six QTLs) was mapped on Chr 04 followed by Chr10 with five M-QTLs, and the lowest number of one M-QTL was mapped

TABLE 1 | 1 QTL naming followed nomenclature q for quantitative trait locus, trait name seed size [seed length (SL), thickness (ST), and width (SW)] and seed shape [length to-thickness (SLT), length-to-width (SLW), width-to-thickness (SWT) ratios, and flatness index (FI)] followed by chromosome number, number of QTLs detected on each chromosome for each trait and subscripted by name of recombinant inbred line population (LM6 and ZM6). 2 Genetic position of QTL (cM). 3 Additive effect (A) and phenotypic variation explained by QTL(PVE). 4 Additive by environment effect in FY2012 (AE1); JP2012 (AE2); JP2014 (AE3); and JP2017 (AE4).

QTL[1]	Marker interval	Position (cM)[2]	Physical position (bp)	Additive – effect (A)[3]		Additive × environment effect (AE)[4]					References
				A	H^2%	AE1	AE2	AE3	AE4	H^2%	
qSL-7-1$_{LM6}$	bin744–bin745	15.25	3,324,836–3,459,470	0.16**	17.45	NS	NS	NS	NS	0.07	Hu et al., 2013
qSL-13-6$_{LM6}$	bin1535–bin1536	140.15	43,244,220–44,026,619	0.13**	6.08	NS	NS	NS	NS	0.28	Salas et al., 2006
qSW-13-5$_{LM6}$	bin1536–bin1537	143.04	43,953,331–44,408,971	0.51**	10.52	0.18*	−0.18*	0.16*	0.21**	0.15	Salas et al., 2006
qSW-19-2$_{LM6}$	bin2100–bin2101	42.57	34,493,194–34,882,495	0.18**	19.47	−0.13**	−0.11*	0.06*	0.09*	0.18	New
qST-9-5$_{LM6}$	bin1022–bin1023	51.88	7,308,659–7,459,924	0.12**	8.6	NS	NS	NS	NS	0.03	Salas et al., 2006
qST-18-4$_{LM6}$	bin1979–bin1980	43.69	9,222,099–10,402,370	0.06**	11.34	NS	NS	NS	NS	0.07	Fang et al., 2017
qSLT-3-1$_{LM6}$	bin247–bin248	19.49	3,119,582–3,515,594	0.12**	2.67	NS	−.22**	NS	NS	4.93	New
qSLT-14-1$_{LM6}$	bin1586–bin1587	42.38	7,850,227–8,143,522	−0.09**	1.96	NS	0.27**	NS	NS	6.73	Li et al., 2010
qSLT-17-5$_{LM6}$	bin1883–bin1884	70.26	13,441,932–13,696,232	−0.13**	2.41	NS	NS	NS	NS	1.05	New
qSWT-7-5$_{LM6}$	bin816–bin817	79.69	29,822,346–35,034,728	−0.67**	12.58	NS	NS	NS	NS	1.27	Fang et al., 2017
qFl-3-3$_{LM6}$	bin244–bin245	17.30	2,790,829–2,980,527	0.16**	11.48	NS	NS	NS	NS	0.13	New
qFl-5-1$_{LM6}$	bin476–bin477	0.21	1–5,29,217	0.17**	21.22	NS	NS	NS	NS	0.15	New
qFl-8-6$_{LM6}$	bin954–bin955	95.88	35,158,414–37,964,850	0.17**	15.74	0.21*	−0.19*	0.14**	0.21*	0.11	New
qFl-9-5$_{LM6}$	bin1030–bin1031	56.45	20,192,294–27,035,074	−0.13**	4.27	NS	NS	NS	NS	0.21	New
qFl-11-3$_{LM6}$	bin1290–bin1291	69.93	18,546,688–18,767,705	0.1**	2.27	NS	NS	NS	NS	1.47	New
qFl-16-1$_{LM6}$	bin1745–bin1746	1.66	6,97,999–9,08,917	−0.08**	14.74	NS	NS	NS	NS	0.11	New
qSL-1-4$_{ZM6}$	bin4–bin5	3.23	7,54,691–1,375,000	0.05**	3.33	NS	NS	NS	NS	0.08	New
qSL-9-2$_{ZM6}$	bin1174–bin1175	90.55	3,850,7474–38,736,001	0.04*	2.06	NS	NS	0.07*	NS	2.29	New
qSL-10-1$_{ZM6}$	bin1236–bin1237	24.62	3,150,454–3,297,961	0.05**	13.67	NS	NS	NS	NS	1.4	New
qSL-10-2$_{ZM6}$	bin1334–bin1335	106.35	44,226,599–44,378,813	0.05**	5.57	−0.1**	0.06*	0.09**	−0.07*	3.58	Li et al., 2010
qSL-12-4$_{ZM6}$	bin1553–bin1554	97.99	38,615,116–38,812,896	−0.05**	3.17	NS	NS	NS	NS	0.83	New
qSL-13-2$_{ZM6}$	bin1612–bin1613	71.65	25,830,321–26,065,585	−0.08**	6.17	NS	NS	NS	NS	1.33	Fang et al., 2017
qSL-15-5$_{ZM6}$	bin1918–bin1919	85.59	17,503,517–17,963,129	0.14**	3.56	NS	NS	NS	NS	0.08	Salas et al., 2006
qSW-8-5$_{ZM6}$	bin959–bin960	73.74	11,970,511–12,228,336	0.04**	3	NS	NS	NS	NS	0.65	New
qST-10-5$_{ZM6}$	bin1334–bin1335	106.35	44,226,599–44,378,813	0.42**	3.99	NS	NS	NS	NS	3.07	Hu et al., 2013
qST-10-6$_{ZM6}$	bin1336–bin1337	107.17	44,378,814–44,741,960	−0.41**	2.25	NS	NS	0.15**	NS	3.17	New
qST-13-5$_{ZM6}$	bin1609–bin1610	67.36	24,985,496–25,641,179	−0.41**	4.1	0.8**	−0.6*	−0.7**	0.61*	3.41	New
qST-14-3$_{ZM6}$	bin1809–bin1810	104.68	47,489,495–47,717,306	0.04**	2.87	NS	NS	0.05**	NS	2.42	New
qST-20-1$_{ZM6}$	bin2463–bin2464	4.37	6,62,753–1,045,131	−0.06**	3.78	NS	NS	NS	NS	0.71	New
qSLW-9-4$_{ZM6}$	bin1172–bin1173	89.06	38,139,739–38,507,473	0.99**	2.51	NS	NS	NS	NS	1.25	Li et al., 2010
qSLW-10-2$_{ZM6}$	bin1275–bin1279	60.40	14,218,565–17,808,941	0.82**	12.74	NS	−0.92*	0.95*	NS	2.3	New
qSLW-13-5$_{ZM6}$	bin1653–bin1654	102.63	32,704,220–33,303,066	−0.1**	1.15	NS	0.3**	NS	NS	3.95	Salas et al., 2006
qSLT-5-3$_{ZM6}$	bin600–bin599	93.89	40,328,493–40,882,874	0.013**	11.18	NS	NS	NS	NS	0.32	Salas et al., 2006
qFl-17-6$_{ZM6}$	bin2177–bin2178	130.75	41,009,636–41,399,912	0.06**	14.81	NS	NS	NS	NS	0.6	New
qFl-20-1$_{ZM6}$	bin2461–bin2462	2.80	1–6,62,752	0.08**	5.45	NS	NS	0.12*	0.09*	0.28	New

*$p < 0.05$; **$p < 0.01$; NS, non-significant. A indicates additive effects, those with positive values show beneficial alleles from parents Zhengyang and Linhefenqingdou, while those with negative values show beneficial alleles from parent Meng 8206.*

TABLE 2 | 1 QTL naming followed nomenclature q for quantitative trait locus, trait name 100-seed weight (HSW) followed by chromosome number, number of QTLs detected on each chromosome for each trait and subscripted by name of recombinant inbred line population (LM6 and ZM6). 2 Genetic position of QTL (cM). 3 Additive effect (A) and phenotypic variation explained by QTL (PVE). 4 Additive by environment effect in FY2012 (AE1); JP2012 (AE2); JP2013 (AE3); JP2014 (AE4); YC2014 (AE5); and JP2017 (AE6).

QTL[1]	Marker interval	Position (cM)[2]	Physical position (bp)	Additive – effect (A)[3]		Additive × environment effect (AE)[4]						References
				A	H^2%	AE1	AE2	AE3	AE4	AE5	AE6	H^2%	
qHSW-3-2$_{LM6}$	bin255–bin256	30.64	5,83,3775–6,78,0840	0.61**	10.75	NS	NS	NS	NS	NS	NS	0.16	Li et al. (2008)
qHSW-14-3$_{LM6}$	bin1640–bin1641	101.33	48,267,526–48,523,627	0.48**	0.28	0.52*	NS	NS	NS	−0.16*	NS	1.99	New
qHSW-8-3$_{ZM6}$	bin963–bin964	80.64	12,871,276–13,803,222	0.29**	3.21	NS	NS	NS	0.34*	−0.41*	NS	0.42	Han et al., 2012
qHSW-9-1$_{ZM6}$	bin1162–bin1163	77.51	35,758,796–36,561,550	0.40**	2.47	NS	NS	NS	NS	NS	NS	0.88	New
qHSW-13-3$_{ZM6}$	bin1611–bin1612	69.99	25,641,180–26,012,595	−0.33**	5.7	NS	NS	−0.38*	NS	NS	NS	1.13	Funatsuki et al. (2005)
qHSW-14-2$_{ZM6}$	bin1746–bin1747	28.67	4,176,245–4,861,311	−0.15**	1.66	NS	NS	NS	NS	NS	NS	0.17	New
qHSW-14-4$_{ZM6}$	bin1809–bin1810	104.68	47,489,495–47,717,306	0.46**	2.78	0.54**	NS	NS	0.43*	NS	−0.69**	4.75	New
qHSW-16-3$_{ZM6}$	bin2043–bin2044	103.81	35,441,262–35,607,069	0.21**	1.67	NS	NS	NS	NS	NS	NS	0.29	New

*$p < 0.05$; **$p < 0.01$; *NS, non-significant. A indicates additive effects, those with positive values show beneficial alleles from parents Zhengyang and Linhefenqingdou, while those with negative values show beneficial alleles from parent Meng 8206.*

to Chr 02, 09, 17, and 18. The most prominent M-QTLs were *qHSW-14-2*$_{ZM6}$, *qHSW-10-3*$_{LM6}$, and *qHSW-10-4*$_{LM6}$ with LOD scores and R^2 values of 10.61 and 24.50% (**Figure 5**), 7.57 and 17.60%, and 7.20 and 16.90%, respectively. Among those 21 M-QTLs, *qHSW-4-3*$_{LM6}$, $_{ZM6}$, *qHSW-6-2*$_{LM6}$, *qHSW-10-1*$_{LM6}$, *qHSW-13-1*$_{ZM6}$, *qHSW-15-2*$_{LM6}$, and *qHSW-15-4*$_{LM6}$ were mapped in at least three environments with an average R^2 of 13.01%.

Analysis of Additive-Effect QTLs and Additive × Environment QTL Interactions

The mixed-MCIM approach implemented in the QTL Network V2.1 software for both RIL populations across multiple environments identified 35 *AA* QTLs on 17 chromosomes related to seven seed size and seed shape traits. These comprise 9, 3, 7, 3, 4, 1, and 8 *A* QTLs associated with SL, SW, ST, SLW, SLT, SWT, and FI, respectively, in the LM6 and ZM6 populations across all environments (**Table 1**). The contributed allele of 11 QTLs of them which reduces seed size and seed shape values through significant additive effects is inherited from the M8206 parent. Meanwhile, the contributed allele of the remaining 24 QTLs, which enhances seed size and shape values through significant additive effects, is inherited from either Zhengyang or Linhefenqingdou parent of the ZM6 or LM6 population, respectively (**Table 1**). Thirteen out of 35 QTLs revealed significant *AE* effects in at least one environment. However, five QTLs, i.e., *qSW-13-5*$_{LM6}$, *qSW-19-2*$_{LM6}$, *qFI-8-6*$_{LM6}$, *qSL-10-2*$_{ZM6}$, and *qST-13-5*$_{ZM6}$, showed significant or highly significant *AE* among all studied environments (**Table 1**). The influence of *AE* effects on seed size and seed shape values was environmentally dependent (**Table 1**). Eight *AA* QTLs associated with HSW were identified on six chromosomes, i.e., Chr 03, 08, 09, 13, 14, and 16, in LM6 and ZM6 populations across six environments (**Table 2**). Six of those eight QTLs displayed a positive additive effect with the beneficial allele that could increase HSW which is inherited from the female parents (Linhefenqingdou or Zhengyang). Meanwhile, the remaining two QTLs, i.e., *qHSW-13-3*$_{ZM6}$ and *qHSW-14-2*$_{ZM6}$, revealed negative additive effects with the allele that reduces HSW which is inherited from the common male parent (Meng8206) (**Table 2**). Two QTLs associated with HSW, i.e., *qHSW-14-3*$_{LM6}$ and *qHSW-8-3*$_{ZM6}$, displayed significant *AE* effects in two environments, whereas *qHSW-13-3*$_{ZM6}$ showed a significant *AE* only in one environment (the 13JP environment). The *qHSW-14-4*$_{ZM6}$ QTL revealed significant *AE* effects across three different environments, i.e., 12FY, 12JP, and 17JP (**Table 2**).

Validation of QTLs and Identification of QTL Clusters

A chromosomal region comprising several identified M-QTLs for different studied seed phenotypic traits was designated as a QTL cluster. Accordingly, 24 QTL clusters on 17 chromosomes with the exception to Chr 02, 12, and 18 were identified (**Supplementary Table 10** and **Supplementary Figure 2**). Among the identified 24 clusters, seven clusters harbored QTLs related

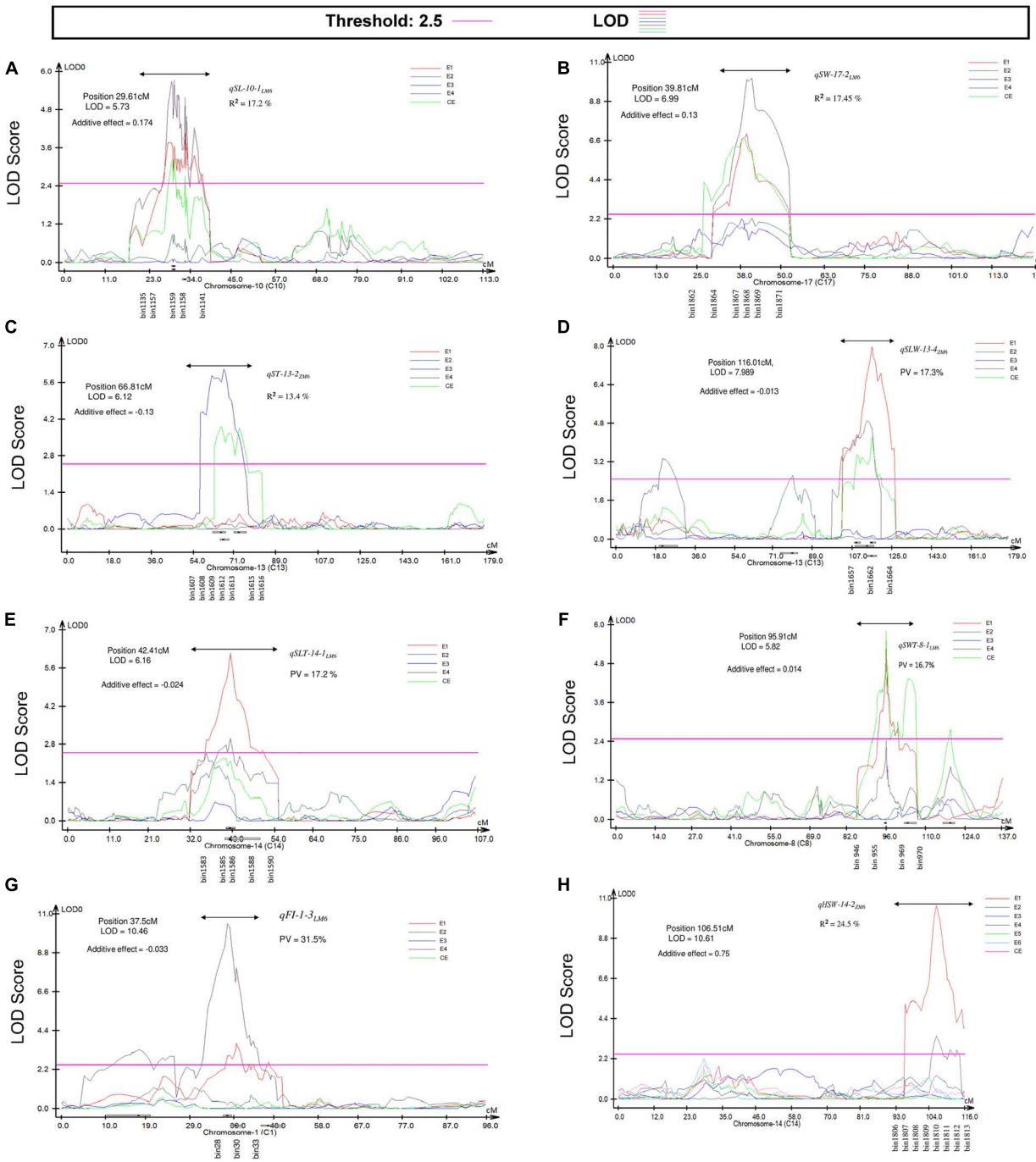


FIGURE 5 | Position of the most prominent QTL detected by the CIM approach associated with seed size and seed shape traits in the LM6 and ZM6 RIL populations grown in multiple environments indicated as with E1, FY2012; E2, JP2012; E3, JP2013; E4, JP2014; E5, YC2014; E6, JP2017, respectively, in addition to the combined environment (CE). **(A)** LOD curve for *qSL-10-1*$_{LM6}$, **(B)** LOD curve for *qSW-17-2*$_{LM6}$, **(C)** LOD curve for *qST-13-2*$_{ZM6}$, **(D)** LOD curve for *qSLW-13-4*$_{ZM6}$, **(E)** LOD curve for *qSLT-14-1*$_{LM6}$, **(F)** LOD curve for *qSWT-8-1*$_{LM6}$, **(G)** LOD curve for *qFl-1-3*$_{LM6}$, and **(H)** LOD curve for *qHSW-14-2*$_{ZM6}$. The LOD threshold (2.5) is indicated by a pink line. The double-headed arrow denotes the location of prominent QTL. The *X* and *Y*-axes represent chromosome and LOD score, respectively.

to seed size, seed shape, and HSW, five clusters harbored QTLs related only to seed size and seed shape traits, nine clusters comprised QTLs related to seed size and HSW traits, and three clusters harbored QTLs for only seed shape traits (**Supplementary Table 10**). QTLs within 15 clusters revealed positive additive effects with the beneficial alleles which are inherited from the big seed size and heavy seed weight parents (Zhengyang or Linhefenqingdou). Seven out of 24

clusters contain QTLs that have been detected and validated in the low RIL populations (**Supplementary Table 10**). The most prominent M-QTL (*qFI-1-3*$_{LM6}$) with a LOD score of 3.71–10.44 and R^2 (10.45–31.50%) was located to Cluster-01. Each cluster comprised a different number of QTLs, with the highest number of QTLs, i.e., seven, associated with seed size, shape, and HSW traits which were in cluster-03 at the physical position of 1,509,548–6,780,840 bp allocated as two QTLs related to seed size (*qSL-3-1*$_{LM6}$ and *qSL-3-2*$_{LM6}$), four QTLs for seed shape (*qSLW-3-2* $_{LM6}$, *qFI-3-1*$_{LM6}$, *qSLT-3-1*$_{LM6}$, and *qFI-3-2*$_{LM6}$), and one QTL HSW (*qHSW-3-1*$_{LM6}$). Except for *qHSW-3-1*$_{LM6}$, all QTLs in this cluster were major QTLs (with an $R^2 > 10\%$). Each of the clusters-13, 16.2, and 17.1 comprises five to six M-QTLs related to seed size and HSW traits identified in one of the two RIL populations and displayed R^2 values of 8.85–13.43, 5.96–11.26, and 6.8–18.30%, respectively (**Supplementary Figure 2** and **Supplementary Table 10**). Another rich region of QTLs was cluster-20 on chr20 that harbors five seed size and shape M-QTLs, i.e., *qFI-20-1*$_{ZM6}$, *qSLT-20-1*$_{ZM6}$, *qSLW-20-1*$_{ZM6}$, *qST-20-1*$_{ZM6}$, and *qSW-20-1*$_{ZM6}$, from which three are major QTLs with R^2 of 11.2–19.2% within a physical region of 1.2 Mb (**Supplementary Table 10**). Cluster-09 comprises five stable (identified in the two populations) seed size, shape, and HSW QTLs with R^2 values ranging from 12.5 to 16.3%. Cluster-14.1 comprises four major M-QTLs in both populations with R^2 values ranging from 10.4 to 18.4% within the chromosomal region between 5,834,015 and 9,844,637 bp, one from which *qSW-14-2*$_{ZM6}$ is associated with seed size traits, whereas the other three (*qSLW-14-1*$_{LM6}$, *qFI-14-1*$_{LM6}$, and *qSLT-14-1*$_{LM6}$) were associated with seed shape traits (**Supplementary Table 10**). Six clusters harboring four M-QTLs each were identified, from which four clusters, i.e., cluster-07, cluster-19.1, cluster-08, and cluster-14.2, comprise QTLs associated with HSW, seed size, and shape trait cluster-10.2 that comprises M-QTLs for seed size and shape traits, and cluster-16.1 that comprises only M-QTLs related to seed shape traits (**Supplementary Table 10**). The remaining nine clusters have three QTLs each; out of them, cluster-01 and cluster-17.2 comprise major QTLs related only to seed shape traits. Meanwhile, cluster-04.1 and cluster-19.2 contain minor M-QTLs associated with SW, SL, and HSW. Another two clusters comprise M-QTLs for both seed size and shape traits (**Supplementary Table 10**). The other three M-QTL clusters, i.e., cluster-10.1, cluster-11, and cluster-15, comprise both major and minor QTLs for seed size traits and HSW. Cluster-04.2 comprises the two QTLs *qHSW-4-3*$_{LM6,ZM6}$ and *qSL-4-1*$_{ZM6}$ with R^2 values of 13.1–17.7%.

Analysis of Epistatic-Effect QTLs and Epistatic × Environment QTL Interactions

Analysis of the seed size and shape trait data under all environments identified 38 pairwise epistatic effects (AA) QTLs, from which 2, 13, 6, 2, 3, 5, and 7 pairs were related to SL, SW, ST, SLW, SLT, SWT, and FI traits, respectively, with R^2 values ranging 0.51–11.35% (**Table 3**). All QTL pairs displayed a high significant *AA* effect. Further analyses revealed that 20 AA QTLs showed significant or highly significant pairwise additive–additive–environment (*AAE*) interaction effects in at least one environment with R^2 values ranging from 0.13 to 5.31% (**Table 3**). Ten pairs showed significant *AAE* in two environments, i.e., 12FY (AAE1) and 12JP (AAE2), while three pairs displayed significant *AAE* in 12JP (AAE2) and 14JP (AAE3) environments (**Table 3**). This shows the effect of the environment on gene expression on phenotype development through epistatic effects. Out of the 38 QTLs, 16 pairwise interactions exhibited negative epistatic effects (*AA*) that decrease the values of seed size and shape traits, whereas 22 pairwise interactions exhibited positive epistatic effects (*AA*) that increase the values of seed size and shape traits (**Table 3**). The pairwise interaction between *qFI-1-1*$_{ZM6}$ and *qFI-7-3*$_{ZM6}$ revealed the strongest positive epistatic effect (0.65), whereas the pairwise *qSLT-6-1*$_{LM6}$ and *qSLW-9-1*$_{LM6}$ revealed the weakest positive epistatic effect (0.02). Conversely, *qSWT-3-1*$_{LM6}$ and *qSWT-13-1*$_{LM6}$ resulted in the strongest negative epistatic effect (−0.71), whereas the *qSLW-2-6*$_{ZM6}$ and *qSLW-18-3*$_{ZM6}$ pairwise resulted in the weakest negative epistatic effect (-0.02) (**Table 3**). Two digenic positive pairwise epistatic QTLs for HSW with highly significant additive × additive (*AA*) effects were identified on four chromosomes (**Table 4**). The first pairwise is composed of two QTLs, *qHSW-11-1*$_{LM6}$ and *qHSW-20-1*$_{LM6}$, with an R^2 of 3.46%, whereas the second pairwise comprises the two QTLs *qHSW-9-1*$_{ZM6}$ and *qHSW-16-3*$_{ZM6}$ with an R^2 of 1.38%. However, the two pairs did not show any significant *AAE* interaction effects across all six environments (**Table 4**).

Candidate Gene Mining of the Main-Effect QTLs

The 24 M-QTL clusters were filtered based on the richness in QTLs associated with all or some of the seed size, shape, and HSW traits. As a result, seven QTL clusters, i.e., cluster-03, 04.1, 05.1, 07, 09, 17.1, and 19.1, were used to identify candidate genes based on publicly available databases such as SoyBase and Phytozome and published papers. According to the physical intervals of the seven QTL clusters, 242, 190, 444, 367, 437, 523, and 116 genes were identified within cluster-03, 04.1, 05.1, 07, 09, 17.1, and 19, respectively, which were retrieved from the SoyBase database (see text footnote 1; **Supplementary Table 11**). GO enrichment analyses via AgriGO V2.0 (see text footnote 4) (Tian et al., 2017) were used to classify the model genes in each cluster. The classification was based on molecular function, biological process, and cellular components visualized on the Web-based GO (WeGO) V2.0 https://wego.genomics.cn (Ye et al., 2006). In all seven QTL clusters, high percentages of genes were related to catalytic activity, cell part, cell, cellular process, binding, and metabolic process terms, besides the response to stimulus in cluster-03 (**Figure 6**). These indicate the essential roles of these terms in the seed size, shape, and seed weight development in soybean. Probable candidate genes underlying these QTL clusters responsible for seed size, shape, and HSW in soybean were further predicted based on gene annotations, GO enrichment

TABLE 3 | Estimated epistatic effects (AA) and environmental (AAE) interaction of QTLs for soybean seed size traits (SL, SW, and ST) and seed shape (SLW, SLT, SWT, and FI) traits across all environments.

RIL	Trait	QTL_i	Chr_i	Interval_i	Pos_i	QTL_j	Chr_j	Interval_j	Pos_j	Epistasis – effect (AA)		Epistasis × environment effect (AAE)				
										AA	H^2%	AAE1	AAE2	AAE3	AAE4	H^2%
LM6	SW	*qSW-2-2*$_{LM6}$	2	bin124–bin125	23.92	*qSW-16-2*$_{LM6}$	16	bin1757–bin1758	18.19	0.08**	7.24	NS	NS	NS	NS	1.11
		qSW-16-2$_{LM6}$	3	bin229–bin230	1.96	*qSW-13-1*$_{LM6}$	13	bin1509–bin1510	107.72	−0.1**	5.40	NS	NS	NS	NS	0.24
		qSW-4-2$_{LM6}$	4	bin353–bin354	13.86	*qSW-15-3*$_{LM6}$	15	bin1738–bin1739	106.86	−0.53**	2.85	NS	NS	NS	NS	2.34
		qSW-4-2$_{LM6}$	4	bin353–bin354	13.86	*qSW-15-4*$_{LM6}$	15	bin1740–bin1741	110.00	0.33**	7.30	0.11**	−0.11**	NS	NS	3.95
		qSW-5-1$_{LM6}$	5	bin525–bin526	74.76	*qSW-12-1*$_{LM6}$	12	bin1352–bin1353	21.59	−0.11**	3.96	NS	NS	NS	NS	0.96
		qSW-7-1$_{LM6}$	7	bin784–bin785	53.84	*qSW-15-4*$_{LM6}$	15	bin1740–bin1741	110.00	−0.10**	5.83	NS	NS	NS	NS	0.05
		qSW-8-1$_{LM6}$	8	bin984–bin985	130.62	*qSW-10-3*$_{LM6}$	10	bin1219–bin1220	97.61	−0.21**	3.20	−0.13*	0.12*	NS	NS	1.39
		qSW-10-1$_{LM6}$	10	bin1184–bin1185	62.20	*qSW-20-1*$_{LM6}$	20	bin2231–bin2232	69.64	−0.09**	2.20	NS	NS	NS	NS	1.19
		qSW-10-3$_{LM6}$	10	bin1219–bin1220	97.61	*qSW-16-4*$_{LM6}$	16	bin1810–bin1811	87.22	0.14**	1.47	NS	NS	NS	NS	0.18
		qSW-11-1$_{LM6}$	11	bin1291–bin1292	70.56	*qSW-15-1*$_{LM6}$	15	bin1715–bin1716	85.15	−0.17**	4.79	NS	NS	NS	NS	1.08
		qSW-11-2$_{LM6}$	11	bin1292–bin1296	71.58	*qSW-15-2*$_{LM6}$	15	bin1717–bin1718	85.58	0.23**	9.22	NS	NS	NS	0.10**	1.48
	ST	*qST-3-1*$_{LM6}$	3	bin237–bin238	8.12	*qST-3-3*$_{LM6}$	3	bin344–bin345	114.06	0.05**	0.51	NS	NS	NS	NS	0.56
		qST-6-3$_{LM6}$	6	bin647–bin648	109.3	*qST-11-1*$_{LM6}$	11	bin1274–bin1275	54.45	0.1**	5.55	NS	NS	NS	0.12**	1.40
		qST-7-2$_{LM6}$	7	bin749–bin750	19.81	*qST-16-3*$_{LM6}$	16	bin1755–bin1756	17.12	0.09**	6.31	NS	NS	NS	NS	0.57
		qST-7-3$_{LM6}$	7	bin783–bin784	53.32	*qST-15-2*$_{LM6}$	15	bin1741–bin1742	110.5	−0.12**	5.74	NS	NS	NS	NS	0.69
		qST-16-1$_{LM6}$	16	bin1744–bin1745	1.66	*qST-17-2*$_{LM6}$	17	bin1886–bin1887	72.87	−0.14**	9.36	0.06*	−0.06*	NS	NS	2.04
	SLW	*qSLW-8-2*$_{LM6}$	8	bin941–bin942	86.24	*qST-14-3*$_{LM6}$	14	bin1625–bin1626	84.91	0.03**	11.35	NS	NS	NS	NS	1.41
	SLT	*qSLT-5-3*$_{LM6}$	5	bin560–bin543	92.74	*qSLW-6-2*$_{LM6}$	6	bin625–bin626	82.38	−0.04**	5.40	0.03*	NS	NS	NS	2.85
		qSLT-6-1$_{LM6}$	6	bin584–bin585	28.72	*qSLW-9-1*$_{LM6}$	9	bin1095–bin1096	128.66	0.02**	1.61	−0.03*	0.03*	NS	NS	3.00
	SWT	*qSWT-3-1*$_{LM6}$	3	bin236–bin237	7.79	*qSWT-13-1*$_{LM6}$	13	bin1435–bin1434	15.23	−0.71**	5.19	0.81**	−0.8**	NS	NS	4.57
		qSWT-6-1$_{LM6}$	6	bin588–bin589	33.52	*qSWT-18-1*$_{LM6}$	18	bin2036–bin2037	95.49	−0.62**	3.88	−0.64*	0.63*	NS	NS	0.67
		qSWT-11-1$_{LM6}$	11	bin1262–bin1263	40.91	*qSWT-20-1*$_{LM6}$	20	bin2177–bin2178	19.84	0.14**	8.03	−0.11**	0.15**	NS	NS	4.19
		qSWT-16-1$_{LM6}$	16	bin1744–bin1745	1.66	*qSWT-17-3*$_{LM6}$	17	bin1886–bin1887	72.87	0.11**	11.08	−0.15**	0.17**	NS	NS	5.31
	FI	*qFI-1-5*$_{LM6}$	1	bin59–bin60	55.61	*qFI-14-4*$_{LM6}$	14	bin1627–bin1628	85.96	0.17**	2.89	NS	NS	0.09*	NS	2.24
		qFI-5-2LM6	5	bin516–bin517	65.79	*qFI-10-2*$_{LM6}$	10	bin1223–bin1224	108.26	0.04**	10.02	NS	NS	NS	NS	0.02
		qFI-16-2$_{LM6}$	16	bin1745–bin1746	1.66	*qSWT-17-2*$_{LM6}$	17	bin1886–bin1887	38.76	0.02**	6.76	−0.18*	0.02*	NS	0.03*	2.75
	SL	*qSL-12-4*$_{ZM6}$	12	bin1553–bin1554	97.99	*qSL-15-5*$_{ZM6}$	15	bin1919–bin1920	85.99	−0.06**	1.88	NS	NS	NS	NS	0.13
		qSL-2-3$_{ZM6}$	2	bin214–bin211	96.57	*qSL-8-6*$_{ZM6}$	8	bin1084–bin1085	186.90	0.2**	8.61	NS	NS	NS	NS	0.10
ZM6	SW	*qSW-4-4*$_{ZM6}$	4	bin434–bin435	62.42	*qSW-20-5*$_{ZM6}$	20	bin2590–bin2591	97.79	−0.12**	6.68	NS	NS	NS	0.07*	0.33
		qSW-6-3$_{ZM6}$	6	bin684–bin685	86.36	*qSW-6-5*$_{ZM6}$	6	bin703–bin704	109.84	0.05**	0.66	−0.08**	0.08**	NS	NS	3.35
	ST	*qST-10-5*$_{ZM6}$	10	bin1334–bin1335	106.35	*qST-10-6*$_{ZM6}$	10	bin1336–bin1337	107.17	−0.7**	2.36	NS	NS	NS	NS	0.48
	SLW	*qSLW-2-6*$_{ZM6}$	2	bin260–bin261	158.24	*qSLW-18-3*$_{ZM6}$	18	bin2336–bin2337	123.90	−0.02**	1.14	NS	0.04**	−0.03**	NS	3.74
	SLT	*qSLT-1-4*$_{ZM6}$	1	bin53–bin54	43.95	*qSLT-7-2*$_{ZM6}$	7	bin884–bin885	103.56	0.03**	4.22	NS	−0.03*	0.03**	NS	2.58
	SWT	*qSWT-1-3*$_{ZM6}$	1	bin62–bin63	47.73	*qSWT-8-1*$_{ZM6}$	8	bin914–bin915	15.98	0.13**	5.30	NS	NS	NS	NS	0.29
	FI	*qFI-17-6*$_{ZM6}$	17	bin2177–bin2178	130.75	*qFI-20-1*$_{ZM6}$	20	bin2461–bin2462	2.79	0.51**	1.75	NS	NS	NS	NS	1.20
		qFI-1-1$_{ZM6}$	1	bin58–bin59	46.05	*qFI-7-3*$_{ZM6}$	7	bin884–bin885	103.56	0.65**	5.14	NS	NS	1.01**	NS	1.26
		qFI-1-3$_{ZM6}$	1	bin72–bin73	66.36	*qFI-7-1*$_{ZM6}$	7	bin872–bin873	96.62	0.3**	6.01	NS	NS	NS	NS	0.65
		qFI-3-1$_{ZM6}$	3	bin289–bin290	27.39	*qFI-18-2*$_{ZM6}$	18	bin2313–bin2314	112.03	0.12**	7.29	NS	0.9*	−0.86*	NS	1.88

*Chr_i and Chr_j indicate the two sites involved in epistatic interactions; Pos indicates genetic position for each of the sites. *$p < 0.05$; **$p < 0.01$; NS, non-significant. AA indicates epistatic effects between two QTLs; those with positive values showing two-locus genotypes being the same as those in parent Linhefenqingdou and Zhengyang (or Meng 8206) have the beneficial effects, while the two-locus recombinants take the negative effects. The case of negative values is the opposite. H^2 indicates phenotypic variation explained by epistatic effects. AE1, FY2012; AE2, JP2012; AE3, JP2014; and AE4, JP2017.*

TABLE 4 | Estimated epistatic effects (AA) and environmental (AAE) interaction of QTLs for soybean 100-seed weight across all environments.

QTL_i	Chr_i	Interval_i	Pos_i	Physical position (bp)_i	QTL_j	Chr_j	Interval_j	Pos_j	Physical position (bp)_j	Epistasis – effect (AA)		Epistasis × environment effect (AAE)						
										AA	H^2%	AAE1	AAE2	AAE3	AAE4	AAE5	AAE6	H^2%
qHSW-11-1$_{LM6}$	*11*	bin1245–bin1246	27.25	6,13,558,464,942,24	*qHSW–20–1*$_{LM6}$	*20*	bin2175–bin2176	17.48	12725901470471	0.51**	3.46	NS	NS	NS	NS	NS	NS	0.32
qHSW-9-1$_{ZM6}$	*9*	bin1162–bin1163	77.51	3,575,879,636,561,550	*qHSW–16–3*$_{ZM6}$	*16*	bin2043–bin2044	103.81	35441262235607069	0.34**	1.38	NS	NS	NS	NS	NS	NS	0.06

*Chr_i and Chr_j indicate the two sites involved in epistatic interactions; Pos indicates genetic position for each of the sites. *p < 0.05; **p < 0.01; NS, non-significant. AA indicates epistatic effects between two QTLs; those with positive values showing two-locus genotypes being the same as those in parent Linhefenqingdou and Zhengyang (or Meng 8206) have the beneficial effects, while the two-locus recombinants take the negative effects. The case of negative values is the opposite. H^2 indicates phenotypic variation explained by epistatic effects. AE1, FY2012; AE2, JP2012; AE3, JP2013; AE4, JP2014; AE5, YC2014; and AE6, JP2017.*

analysis, and the previously known putative biological function of the gene. Based on these, 19, 12, 26, 18, 22, 30, and 16 candidate genes were identified within the QTL clusters-03, 04.1, 05.1, 07, 09, 17.1, and 19.1, respectively (**Supplementary Table 12**). These genes may function directly or indirectly in regulating seed development in soybean, which regulates seed size, shape, and HSW. These genes are involved in response to brassinosteroid stimulus, regulation of cell proliferation and differentiation, regulation of transcription, secondary metabolism and signaling, storage of proteins and lipids, hormone-mediated signaling pathway, regulation of the cell cycle process, transport, ubiquitin-dependent protein catabolic process, embryonic pattern specification, and response to auxin stimulus (**Table 5**). However, the RNS-seq data of genes in the soybean genome (Severin et al., 2010) that is publicly available on SoyBase was used to heatmap the expression of those candidate genes in the young leaf, flower, pod, seed, root, and nodule (**Figure 7** and **Supplementary Table 13**). From the heatmaps, 47 genes out of the identified 143 candidate genes are highly expressed during seed developmental stages and in seed-related tissues (**Figure 7** and **Supplementary Table 13**); hence, they could be potential seed size, shape, and HSW regulatory genes.

DISCUSSION

The present study has implemented high-density genetic maps constructed from two-related RIL populations LM6 and ZM6 comprising 2,267 and 2,601 bin markers, respectively (Li et al., 2017), to validate QTLs associated with seed size, shape, and weight. To minimize the environmental errors, the two RIL populations were evaluated in four environments. The transgressive segregation and continuous variations observed in the two populations in all studied phenotypic traits facilitate the identification of a high number of both major and minor effect QTLs including some novel QTLs associated with all studied traits (Teng et al., 2009; Xu et al., 2011; Zhang et al., 2018). All measured and calculated traits in both populations were significantly ($P < 0.01$) influenced by genotype (G), environment (E), and their interactions (G × E), suggesting that the seed size, shape, and weight traits are not only governed by both genetic and environment; however, there was an effect of the G × E interaction as well (Sun et al., 2012; Hu et al., 2013; Liang et al., 2016). This explains the observed high h^2 (99.04%) and accordingly deduces that these traits are amenable to manipulation by selection without the help of molecular markers. Except for SL, SW, and ST that exhibited a highly significant correlation between each other and with HSW, our data showed that seed size, shape, and weight traits are not correlated, which is favorable when breeding for a round type with smaller or bigger seed size (Cober et al., 1997; Salas et al., 2006).

For validation of identified QTLs, a comparative QTL analysis using the CIM QTL mapping approach with the SoyBase database identified 69, 82, and 29 novel QTLs for seed size, shape, and HSW, respectively, indicating the distinct genetic architecture

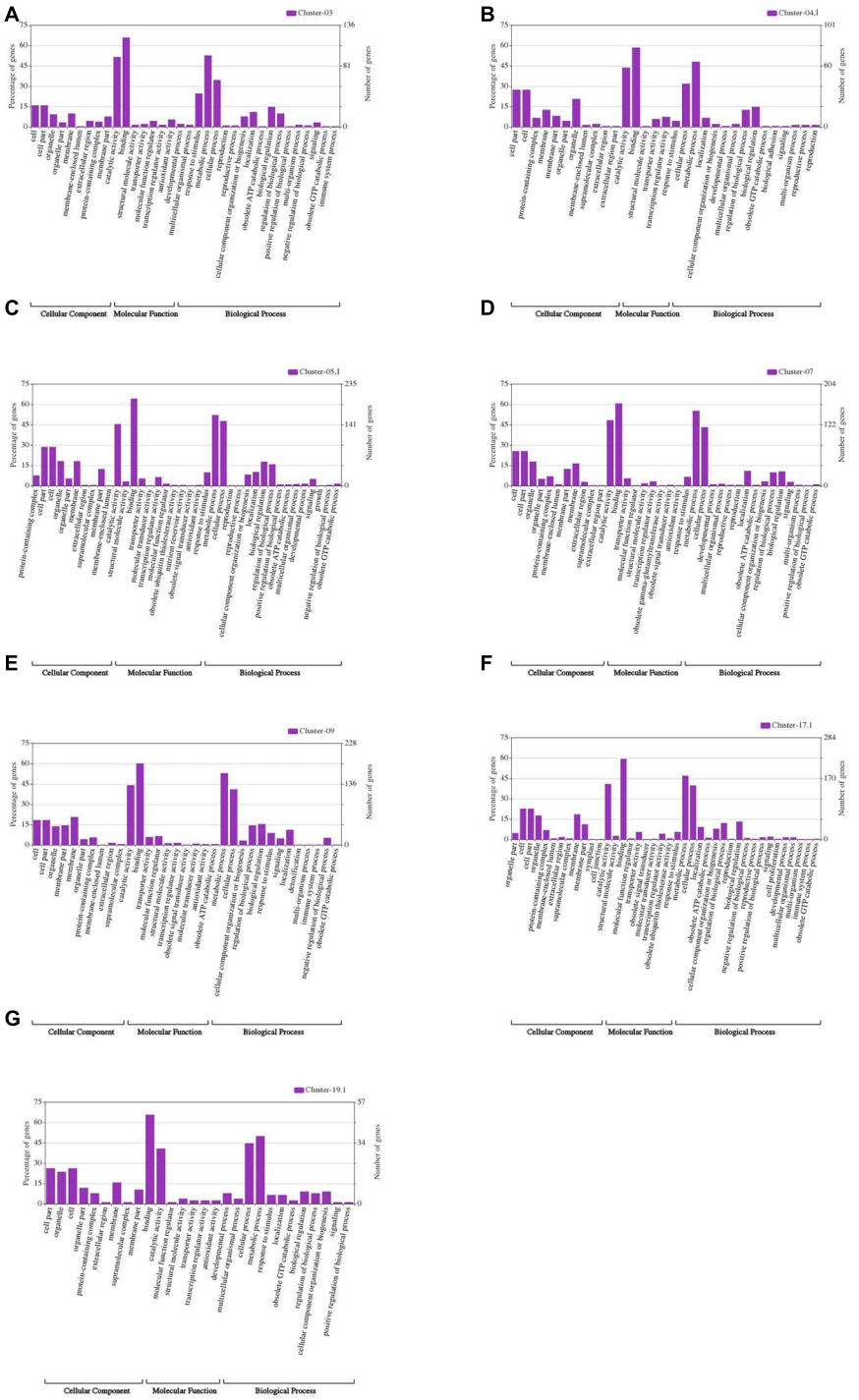


FIGURE 6 | WeGO analysis of the genes located within the seven major QTL clusters: **(A)** Cluster-03; **(B)** Cluster-4.1; **(C)** Cluster-5.1; **(D)** Cluster-07; **(E)** Cluster-09; **(F)** Cluster-17.1; and **(G)** Cluster-19.1.

of the LM6 and ZM6 populations. These novel QTLs together explain over 88.00% of phenotypic variance for seed size, shape, and weight, signifying their potential value for improving soybean cultivars. Besides, the identification of novel QTLs in the present study suggests that more germplasms are required for unraveling the complex genetic basis for seed size and shape traits in soybean. Among these novel QTLs, eight novel major QTLs associated with HSW where their physical intervals did not overlap with any of the previously reported HSW QTLs, suggesting them as potential loci for HSW and major QTLs for

TABLE 5 | Candidate genes identified within the seven QTL clusters that are highly expressed in soybean seed.

QTL clusters	Gene	Start	Stop	Gene functional annotation
Cluster-03	Glyma03g01880	1668601	1674475	Seed dormancy process; protein ubiquitination; lipid storage
	Glyma03g03210	3001933	3005606	Pollen development; embryo sac egg cell differentiation; DNA-dependent
	Glyma03g03760	3581308	3584468	Maintenance of shoot apical meristem identity; cell differentiation
	Glyma03g04330	3581308	3584468	Embryo development; regulation of seed maturation
	Glyma03g04620	4798039	4801122	Regulation of meristem growth; protein deubiquitination
Cluster-04.1	Glyma04g02970	2146489	2152500	Embryo sac egg cell differentiation
	Glyma04g03210	2347024	2349849	Fatty acid beta-oxidation; response to auxin stimulus; ovule development
	Glyma04g03610	2630227	2632308	Brassinosteroid mediated signaling pathway; seed development; ovule development
	Glyma04g04460	3305860	3308715	Response to cytokinin stimulus; response to brassinosteroid stimulus; seed development
	Glyma04g04540	3395831	3397238	Response to ethylene stimulus; seed dormancy process; floral organ morphogenesis
	Glyma04g04870	3628743	3634478	Embryo development ending in seed dormancy
Cluster-05.1	Glyma05g28950	34669156	34678593	Nucleotide biosynthetic process; embryo development ending in seed dormancy
	Glyma05g29700	35236284	35242029	Brassinosteroid biosynthetic process; starch biosynthetic process
	Glyma05g30380	35754306	35755603	Embryo development; protein ubiquitination; lipid storage; anther development
	Glyma05g31450	36578952	36583516	Post-embryonic development
	Glyma05g31490	36611301	36615160	Embryo development ending in seed dormancy
	Glyma05g31830	36870586	36873840	Ubiquitin-dependent protein catabolic process
	Glyma05g32030	37026301	37031440	Ubiquitin-dependent protein catabolic process; multicellular organismal development
	Glyma05g33790	38337126	38341410	Phosphatidylcholine biosynthetic process; metabolic process
	Glyma05g34070	38511154	38513219	Cellular response to abscisic acid stimulus
Cluster-07	Glyma07g13230	11764552	11784123	Embryo sac egg cell differentiation; protein ubiquitination; lipid storage
	Glyma07g13730	12749034	12753558	Embryo development; positive regulation of gene expression
	Glyma07g14460	13903037	13906228	Embryo development ending in seed dormancy
	Glyma07g15050	14900705	14909235	Seed dormancy process; regulation of cell cycle process
	Glyma07g15640	15378798	15384642	Response to hormone stimulus and auxin stimulus; response to brassinosteroid stimulus
	Glyma07g15840	15528948	15544150	Ubiquitin-dependent protein catabolic process; regulation of lipid catabolic process
Cluster-09	Glyma09g28640	35573357	35579018	Embryo development ending in seed dormancy; cellular response to abscisic acid stimulus
	Glyma09g29030	35989729	35993075	Ubiquitin-dependent protein catabolic process; fatty acid beta-oxidation
	Glyma09g29720	36540972	36548174	Response to auxin stimulus; auxin metabolic process
	Glyma09g30130	37014420	37023261	Protein import into nucleus; embryo sac egg cell differentiation
	Glyma09g30650	37426876	37433118	Phosphatidylcholine biosynthetic process; metabolic process; pollen development
	Glyma09g31620	38298193	38307446	Response to abscisic acid stimulus; embryo development
	Glyma09g32600	39100482	39107332	Translational elongation; embryo development ending in seed dormancy
	Glyma09g32680	39173955	39183935	Regulation of protein phosphorylation
	Glyma09g33630	40063507	40067999	Response to auxin stimulus; seed dormancy process
Cluster-17.1	Glyma17g09320	6889969	6894069	Seed maturation; histone deacetylation; response to abscisic acid stimulus
	Glyma17g09690	7171761	7186015	Seed maturation; protein ubiquitination; lipid storage
	Glyma17g10290	7707775	7711360	Pollen tube growth; seed dormancy process; ovule development
	Glyma17g10380	7768561	7778131	Ubiquitin-dependent protein catabolic process
	Glyma17g10990	8262700	8267178	Carbohydrate metabolic process
	Glyma17g11410	8557013	8563158	Regulation of embryo sac egg cell differentiation
	Glyma17g12950	9873806	9891306	Protein folding; embryo development response to starvation
	Glyma17g15490	12218497	12226562	Ubiquitin-dependent protein catabolic process
	Glyma17g15550	12302621	12306143	N-terminal protein myristoylation; pollen development; pollen tube growth
Cluster-19.1	Glyma19g32990	40666918	40669847	Glucose catabolic process; response to auxin stimulus
	Glyma19g33620	41194146	41196743	Maltose metabolic process; starch biosynthetic process; glucosinolate biosynthetic process
	Glyma19g33650	41237306	41242657	Abscisic acid biosynthetic process; plant-type cell wall modification; pollen tube growth

future fine mapping to delimit the physical interval. Numerous QTLs associated with SW, HSW, SLW, and SLT identified in this study are co-localized with previously reported corresponding QTLs (Salas et al., 2006; Li et al., 2010; Moongkanna et al., 2011; Hu et al., 2013; Jun et al., 2014; Fang et al., 2017; Hacisalihoglu et al., 2018; Hina et al., 2020). Our study identified for the first time 13 major QTLs ($R^2 > 10\%$) related to FI; thus, we considered them as novel QTLs. Besides, Chr01 and Chr03 harbored four and three FI QTLs, suggesting crucial roles of Chr 01 and 03 in controlling the inheritance of seed FI in soybean. The

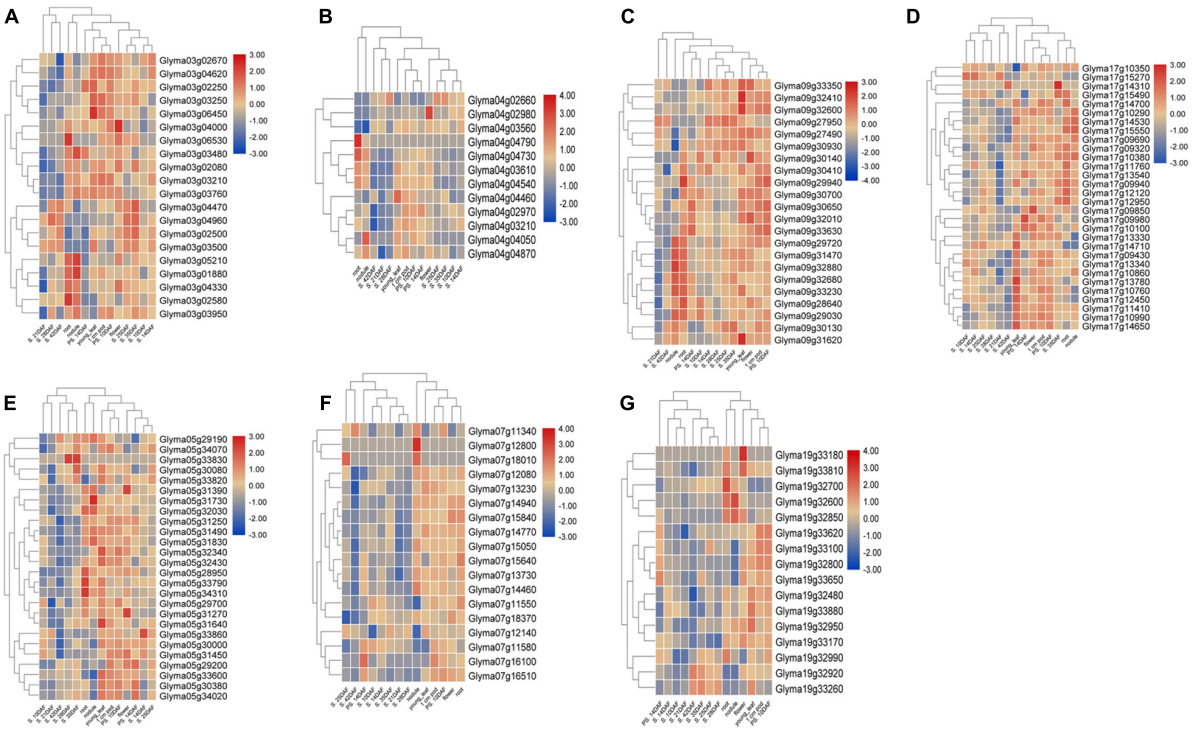


FIGURE 7 | Expression profiling of 47 candidate genes across different development stages in soybean and tissues from seven QTL clusters. **(A)** Cluster-03 candidate genes, **(B)** Cluster-4.1 candidate genes, **(C)** Cluster-5.1 candidate genes, **(D)** Cluster-07 candidate genes, **(E)** Cluster-09 candidate genes, **(F)** Cluster-17.1 candidate genes, and **(G)** Cluster-19.1 candidate genes. A heat map was generated using the RNA-sequencing data retrieved from publicly available database SoyBase. PS = pod shell, DAF = days after flowering and S = Seed.

positive alleles for seed size, shape, and HSW traits were inherited from both parents of the two RIL populations. Therefore, it is likely that not only the higher seed size and heavyweight parent (Linhefenqingdou or Zhengyang) contributed favorable alleles but also the lighter seed weight parent (M8206) might play a role (Cao et al., 2019; Hina et al., 2020).

Mapping of QTLs associated with seed size, shape, and weight-related traits using the MCIM approach was performed to (i) dissect the additive effect QTLs and Q × E interactions, which is essential for selecting the most compatible varieties adapted to particular environments, and (ii) further validate the QTLs identified by the CIM approach. The MCIM approach identified 18 QTLs for seed sizes, shapes, and weight traits that are co-localized in the same physical interval of the CIM-mapped QTLs. Therefore, these QTLs could also be stable QTLs for further fine mapping and map-based cloning to uncover the genetic control and mechanisms of seed size, shape, and weight traits in soybean, and molecular markers tightly linked to these QTLs could be used for MAS.

Dissecting the epistatic and QTL × environment effects are crucial for understanding the genetic mechanisms that contributed to the phenotypic variations of complex traits (Kaushik et al., 2007). Disregarding intergenic interactions will lead to the overestimation of individual QTL effects, and the underestimation of genetic variance resulting in a large drop in the genetic response to MAS especially in late generations (Nyquist and Baker, 1991; Zhang et al., 2004). The identified 40 pairwise digenic epistatic QTLs for seed size, shape, and weight-related traits in the present study could be considered as modifying genes that do not exhibit only additive effects but could affect the expression of seed size, shape, and weight-related genes through epistatic interactions. Similar results for the epistatic interaction of seed size, shape, and weight QTLs have been also previously reported by Xin et al. (2016) and Zhang et al. (2018). The appearance of epistatic interactions for a specific trait makes selection difficult. Noteworthily, all main-effect QTLs detected in our study had no epistatic effect, which raises the heritability of the trait guiding to easier selection.

Genomic regions were identified as QTL clusters based on the presence of several QTLs related to all or some of the seed size, shape, and HSW traits. Accordingly, 24 QTL clusters were identified on 17 chromosomes each containing three or more QTLs related to seed size, shape, and HSW traits. These QTL clusters have not been previously reported, which enhances the developing knowledge of the genetic control of these traits. The co-localization of QTLs for seed size, shape, and HSW and how they have exceptionally corresponded support the highly significant correlation with each other (Cai and Morishima, 2002) (**Supplementary Table 10**). Besides, the occurrence of the QTL clustering could signify a linkage of

QTLs/genes or outcome from the multiple effects of one QTL in the same genomic region (Wang et al., 2006; Cao et al., 2017; Liu et al., 2017). The QTL clusters reveal that the QTL linkage/gathering could make the enhancement of seed size and shape easier than single QTLs (Hina et al., 2020). Significant positive correlations of soybean seed protein and oil contents and seed yield with seed size and seed shape have been shown; therefore, these traits are directly associated with seed size and shape in soybean (Qi et al., 2011; Hacisalihoglu and Settles, 2017; Wu et al., 2018). This notion would explain the co-localization of QTLs associated with seed protein and oil contents in the genomic regions of several QTL (Panthee et al., 2005; Salas et al., 2006; Vieira et al., 2006; Moongkanna et al., 2011; Yang et al., 2011). The position of the first flower and the number of days to flowering have large effects on seed number per plant in soybean (Tasma et al., 2001; Yamanaka et al., 2001; Khan et al., 2008), which affects seed size and HSW indicating the existence of common genetic factors for these traits. QTLs associated with the position of the first flower identified previously (Tasma et al., 2001; Han et al., 2012) are located to the genomic region of clusters 16.1, 19.2, and 20 (Hyten et al., 2004). The extensive analysis of QTL clusters in our study suggests that breeding programs aiming to improve seed size, shape, and weight with enhanced quality should focus on QTL clustering and select QTLs within these regions. Besides, the existence of QTL clusters provides evidence that some traits-related genes are more densely concentrated in specific genomic regions of crop genomes than others (Fang et al., 2017).

Identification of candidate genes underlying QTL regions is of great interest for breeding programs (Abou-Elwafa, 2018; Abou-Elwafa and Shehzad, 2018). A bioinformatics pipeline implementing genomic sequences of identified QTL clusters was employed to identify candidate genes. The pipeline comprises three complementary steps, i.e., (1) retrieving candidate genes from the SoyBase database, (2) visualizing the molecular function of candidate genes by GO enrichment analyses and gene classification, and (3) implying candidate genes in seed size, shape, and weight based on their expression profiles. Accordingly, 47 genes were considered as potential candidates. Most of the identified candidate genes are related to the terms of catalytic activity, cell part, cell, cellular process, and binding and metabolic process as indicated by GO enrichment and gene classification analyses. These terms have functions related/involved in seed development, which influence the size, shape, and weight of seeds (Mao et al., 2010; Li and Li, 2014). For example, the *Glyma07g14460* gene underlying QTL cluster-7 belongs to the oxygenase (CYP51G1) protein class, which has been confirmed to regulate seed size in soybean (Zhao et al., 2016). Furthermore, 10 candidate genes were identified as a regulator of ubiquitin-dependent protein catabolic process, RING-type E3 ubiquitin ligases, and lipid catabolic process (**Table 5**). Several components of the ubiquitin pathway such as the ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin protein ligase (E3) have been reported to play important roles in regulation seed and organ size (Li and Li, 2014). Similarly, 16 candidate genes have functions in pollen tube development, embryo sac egg cell differentiation, post-embryonic development, regulation of seed maturation, positive regulation of gene expression, regulation of cell cycle process, ovule development, anther development, seed dormancy process, and seed maturation (**Table 5**), and hence they are likely to participate in regulating seed size, shape, and weight in plants, including soybean (Meng et al., 2016). Ten candidate genes are involved in response to auxin stimulus, response to ethylene stimulus, and abscisic acid biosynthetic process which are known to be implicated in promoting seed size and weight in Arabidopsis (**Table 5**) (Xie et al., 2014). Six genes play functions in the glucose catabolic process, phosphatidylcholine biosynthetic process, carbohydrate metabolic process, maltose metabolic process, and starch biosynthetic process and are implicated in the partitioning and translocation of photoassimilates and grain filling in rice (**Table 5**) (Chen J. et al., 2020; Zhang et al., 2020).

CONCLUSION

QTLs associated with seed size, shape, and weight in soybean were identified and validated using two mapping approaches in two populations across multiple environments. This is the first comprehensive investigation of the identification and validation of QTLs for the FI as a seed shape trait in soybean. Employing a bioinformatics pipeline identified candidate genes behind genomic regions harboring major and stable QTL clusters underlying the inheritance of seed size, shape, and weight. The implemented bioinformatics pipeline delimits the number of the identified candidate genes to 47-gene genomic regions involved directly or indirectly in seed size, shape, and weight. These genes are highly expressed in seed-related tissues and nodules, indicating that they may be involved in regulating these traits in soybean. Furthermore, some of the potential 47 candidate genes have been included in our ongoing projects for functional validation to confirm their effect on seed size, shape, and weight. Our study provides detailed information for genetic bases of the studied traits and candidate genes that could be efficiently implemented by soybean breeders for fine mapping and gene cloning and for MAS targeted at improving seed size, shape, and weight.

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

TZ designed the project. ME performed the experiments and drafted the manuscript. ME, BK, SS, SL, YC, MA, and AH analyzed the data. TZ and SA-E revised the manuscript.

All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Key R&D Program of China (2018YFD0100800), the National Natural Science Foundation of China (31571691 and 31871646), the MOE Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT_17R55), the Fundamental Research Funds for the Central Universities (KYT201801), and the Jiangsu Collaborative Innovation Center for Modern Crop Production (JCICMCP) Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.666440/full#supplementary-material

Supplementary Figure 1 | Seed size, shape, and seed weight traits related QTLs mapped using CIM approach in the two soybean RIL populations LM6 and ZM6 across multiple environments indicated with E1, FY2012; E2, JP2012; E3, JP2013; E4, JP2014; E5, YC2014; E6, JP2017 respectively, in addition to the combined environment (CE). **(a)** Seed length associated QTLs, **(b)** seed width associated QTLs, **(c)** seed thickness associated QTLs, **(d)** seed length/width associated QTLs, **(e)** seed length/thickness associated QTLs, **(f)** seed width/thickness associated QTLs, **(g)** FI associated QTLs, and **(h)** QTLs associated with HSW. The LOD threshold (2.5) is indicated by a green line. The *X* and *Y*-axis represent chromosomes and LOD score, respectively.

Supplementary Figure 2 | Diagram showing Chromosomal locations of the identified 24 QTL clusters on 17 different chromosomes, i.e., Chr 01, 03, 04, 05, 06, 07, 08, 09, 10, 11, 13, 14, 15, 16, 17, 19, and 20 in LM6 and ZM6 RIL populations for SL, SW, ST, SLW, SLT, SWT, FI, and HSW traits under multiple environments.

Supplementary Table 1 | Distribution of SNPs, recombination bins and markers mapped on soybean chromosomes/linkage groups. Bin-map (RAD-sequencing).

Supplementary Table 2 | Descriptive statistics, variance components, and broad-sense heritability (h^2) of seed shape and size traits evaluated in two recombinant inbred lines (RILs) LM6 and ZM6. The RILs and their parents were grown under different environments. * and ** represent significance at 5 and 1%, respectively.

Supplementary Table 3 | Descriptive statistics, variance components, and broad-sense heritability (h^2) for 100-Seed weight trait evaluated in two recombinant inbred lines (RILs) LM6 and ZM6. The RILs and their parents were grown under different environments. * and ** represent significance at 5 and 1%, respectively.

Supplementary Table 4 | Combined analysis of variance (ANOVA) for Seed shape and size trait (SL, SW, ST, SLW, SLT, SWT, FI) in ZM6 and LM6 RIL Populations across four different environments (2012FY, 2012JP, 2014JP and 2017JP).

Supplementary Table 5 | Combined analysis of variance (ANOVA) for 100-Seed weight trait in ZM6 and LM6 RIL Populations across six different environments (2012FY, 2012JP, 2013JP, 2014JP, 2014YC, and 2017JP).

Supplementary Table 6 | Correlation analysis among seven Seed size/Shape traits (SL, SW, ST, FI, SLT, SLW, SWT), and 100-seed weight (HSW). * and ** represent significance at 5 and 1%, respectively.

Supplementary Table 7 | Main-effect quantitative trait loci (M-QTLs) identified for three seed-size traits (SL, SW, and ST in ZM6 and LM6 recombinant inbred line (RIL) populations across multiple environments and combined environment.

Supplementary Table 8 | Main-effect quantitative trait loci (M-QTLs) identified for four seed-shape traits [Seed length/width ratio (SLW), Seed length thickness ratio (SLT), Seed width thickness ratio (SWT), and Flatness index (FI)] in ZM6 and LM6 recombinant inbred line (RIL) populations across multiple environments and combined environment.

Supplementary Table 9 | Main-effect quantitative trait loci (M-QTLs) identified for 100-Seed weight in ZM6 and LM6 recombinant inbred line (RIL) populations across multiple environments and combined environment.

Supplementary Table 10 | Twenty-Four QTL clusters detected in LM6 and ZM6 RIL populations across multiple environments.

Supplementary Table 11 | Model genes within Cluster-03, Cluster-04.1, Cluster-05.1, Cluster-07, Cluster-09, Cluster-17.1, and Cluster-19.1 regions.

Supplementary Table 12 | Predicted candidate genes within Cluster-03, Cluster-04.1, Cluster-05.1, Cluster-07, Cluster-09, Cluster-17.1, and Cluster-19.1 regions in both RIL Populations (LM6 and ZM6) based on known functional annotation (highlighted lines show candidate genes).

Supplementary Table 13 | Candidate genes within seven QTL clusters and their relative expression data retrieved from RNA-seq data available in SoyBase (DAF = days after flowering) (highlighted lines show candidate genes).

REFERENCES

Abou-Elwafa, S. F. (2016a). Association mapping for drought tolerance in barley at the reproductive stage. *Comptes Rendus Biol.* 339, 51–59. doi: 10.1016/j.crvi.2015.12.002

Abou-Elwafa, S. F. (2016b). Association mapping for yield and yield-contributing traits in barley under drought conditions with genome-based SSR markers. *C. R. Biol.* 339, 153–162. doi: 10.1016/j.crvi.2016.03.001

Abou-Elwafa, S. F. (2018). Identification of genes associated with drought tolerance in barley. *Biol. Plant.* 62, 299–306. doi: 10.1007/s10535-017-0765-0

Abou-Elwafa, S. F., and Shehzad, T. (2018). Genetic identification and expression profiling of drought responsive genes in sorghum. *Environ. Exp. Bot.* 155, 12–20. doi: 10.1016/j.envexpbot.2018.06.019

Basra, A. S. (1995). *Seed quality: Basic Mechanisms and Agricultural Implications*. New York, NY: Food Products Press.

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci.* 48, 1649–1664. doi: 10.2135/cropsci2008.03.0131

Cai, H., and Morishima, H. (2002). QTL clusters reflect character associations in wild and cultivated rice. *Theor. Appl. Genet.* 104, 1217–1228. doi: 10.1007/s00122-001-0819-7

Cailleux, A. (1945). Distinction des galets marins et fluviatiles. *Bull. de la Soc. Géologique de France* 5, 375–404. doi: 10.2113/gssgfbull.s5-xv.7-8.375

Cao, Y., Li, S., Chen, G., Wang, Y., Bhat, J. A., Karikari, B., et al. (2019). Deciphering the genetic architecture of plant height in soybean using two RIL populations sharing a common M8206 parent. *Plants* 8:373. doi: 10.3390/plants8100373

Cao, Y., Li, S., Wang, Z., Chang, F., Kong, J., Gai, J., et al. (2017). Identification of major quantitative trait loci for seed oil content in soybeans by combining linkage and genome-wide association mapping. *Front. Plant Sci.* 8:1222.

Cerdà, A., and Garcıa-Fayos, P. (2002). The influence of seed size and shape on their removal by water erosion. *Catena* 48, 293–301.

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools-an integrative toolkit developed for interactive analyses of big biological data. *bioRxiv [preprint]*, 13, 1194–1202.

Chen, J., Cao, F., Li, H., Shan, S., Tao, Z., Lei, T., et al. (2020). Genotypic variation in the grain photosynthetic contribution to grain filling in rice. *J. Plant Physiol.* 253:153269. doi: 10.1016/j.jplph.2020.153269

Cheng, X., Wang, S., Wang, L., Wang, S., Zong, X., and Wang, P. (2006). *Descriptors and Data Standard for Adzuki Bean [Vigna angularis (Willd.) Ohwi & Ohashi]*. Beijing: China Agriculture Press.

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971. doi: 10.1093/genetics/138.3.963

Cober, E., Voldeng, H., and Frégeau−Reid, J. (1997). Heritability of seed shape and seed size in soybean. *Crop Sci.* 37, 1767–1769. doi: 10.2135/cropsci1997.0011183x003700060017x

Edwards, C. J. Jr., and Hartwig, E. E. (1971). Effect of seed size upon rate of germination in soybeans 1. *Agron. J.* 63, 429–450. doi: 10.2134/agronj1971.00021962006300030024x

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Geno. Biol.* 18, 1–14.

Funatsuki, H., Kawaguchi, K., Matsuba, S., Sato, Y., and Ishimoto, M. (2005). Mapping of QTL associated with chilling tolerance during reproductive growth in soybean. *Theo. Appl. Gen.* 111, 851–861.

Galal, A., Sharma, S., Abou-Elwafa, S. F., Sharma, S., Kopisch-Obuch, F., Laubach, E., et al. (2014). Comparative QTL analysis of root lesion nematode resistance in barley. *Theor. Appl. Genet.* 127, 1399–1407. doi: 10.1007/s00122-014-2307-x

Hacisalihoglu, G., Burton, A. L., Gustin, J. L., Eker, S., Asikli, S., Heybet, E. H., et al. (2018). Quantitative trait loci associated with soybean seed weight and composition under different phosphorus levels. *J. Int. Plant Biol.* 60, 232–241. doi: 10.1111/jipb.12612

Hacisalihoglu, G., and Settles, A. M. (2017). Quantification of seed ionome variation in 90 diverse soybean (*Glycine max*) lines. *J. Plant Nut.* 40, 2808–2817. doi: 10.1080/01904167.2017.1382528

Haig, D. (2013). Kin conflict in seed development: an interdependent but fractious collective. *Ann. Rev. Cell Dev. Biol.* 29, 189–211. doi: 10.1146/annurev-cellbio-101512-122324

Han, Y., Li, D., Zhu, D., Li, H., Li, X., Teng, W., et al. (2012). QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theor. Appl. Genet.* 125, 671–683. doi: 10.1007/s00122-012-1859-x

Hina, A., Cao, Y., Song, S., Li, S., Sharmin, R. A., Elattar, M. A., et al. (2020). High-resolution mapping in two RIL populations refines major "QTL hotspot" regions for seed size and shape in soybean (*Glycine max* L.). *Int. J. Mol. Sci.* 21:1040. doi: 10.3390/ijms21031040

Hoeck, J. A., Fehr, W. R., Shoemaker, R. C., Welke, G. A., Johnson, S. L., and Cianzio, S. R. (2003). Molecular marker analysis of seed size in soybean. *Crop Sci.* 43, 68–74. doi: 10.2135/cropsci2003.0068

Hu, Z., Zhang, H., Kan, G., Ma, D., Zhang, D., Shi, G., et al. (2013). Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. Merr.). *Genetica* 141, 247–254. doi: 10.1007/s10709-013-9723-8

Hyten, D., Pantalone, V., Sams, C., Saxton, A., Landau-Ellis, D., Stefaniak, T., et al. (2004). Seed quality QTL in a prominent soybean population. *Theor. Appl. Genet.* 109, 552–561. doi: 10.1007/s00122-004-1661-5

Jannink, J.-L., Moreau, L., Charmet, G., and Charcosset, A. (2009). Overview of QTL detection in plants and tests for synergistic epistatic interactions. *Genetica* 136:225. doi: 10.1007/s10709-008-9306-2

Jeong, N., Suh, S. J., Kim, M.-H., Lee, S., Moon, J.-K., Kim, H. S., et al. (2012). Ln is a key regulator of leaflet shape and number of seeds per pod in soybean. *Plant Cell* 24, 4807–4818. doi: 10.1105/tpc.112.104968

Jun, T. H., Freewalt, K., Michel, A. P., and Mian, R. (2014). Identification of novel QTL for leaf traits in soybean. *Plant Breeding* 133, 61–66. doi: 10.1111/pbr.12107

Kajiya-Kanegae, H., Takanashi, H., Fujimoto, M., Ishimori, M., Ohnishi, N., Fiona, W. W., et al. (2020). RAD-seq-based high-density linkage map construction and QTL mapping of biomass-related traits in sorghum using a Japanese Landrace Takakibi NOG. *Plant Cell Physiol.* 61, 1262–1272. doi: 10.1093/pcp/pcaa056

Karikari, B., Li, S., Bhat, J. A., Cao, Y., Kong, J., Yang, J., et al. (2019). Genome-wide detection of major and epistatic effect QTLs for seed protein and oil content in soybean under multiple environments using high-density bin map. *Int. J. Mol. Sci.* 20:979. doi: 10.3390/ijms20040979

Kato, S., Sayama, T., Fujii, K., Yumoto, S., Kono, Y., Hwang, T.-Y., et al. (2014). A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. *Theor. Appl. Genet.* 127, 1365–1374. doi: 10.1007/s00122-014-2304-0

Kaushik, N., Kumar, K., Kumar, S., Kaushik, N., and Roy, S. (2007). Genetic variability and divergence studies in seed traits and oil content of Jatropha (*Jatropha curcas* L.) accessions. *Biomass and Bioenergy* 31, 497–502. doi: 10.1016/j.biombioe.2007.01.021

Khan, N. A., Githiri, S. M., Benitez, E. R., Abe, J., Kawasaki, S., Hayashi, T., et al. (2008). QTL analysis of cleistogamy in soybean. *Theor. Appl. Genet.* 117, 479–487. doi: 10.1007/s00122-008-0792-5

Li, D., Sun, M., Han, Y., Teng, W., and Li, W. (2010). Identification of QTL underlying soluble pigment content in soybean stems related to resistance to soybean white mold (*Sclerotinia sclerotiorum*). *Euphytica* 172, 49–57. doi: 10.1007/s10681-009-0036-z

Li, N., and Li, Y. (2014). Ubiquitin-mediated control of seed size in plants. *Front. Plant Sci.* 5:332. doi: 10.3389/fpls.2014.00332

Li, S., Cao, Y., He, J., Zhao, T., and Gai, J. (2017). Detecting the QTL-allele system conferring flowering date in a nested association mapping population of soybean using a novel procedure. *Theor. Appl. Genet.* 130, 2297–2314. doi: 10.1007/s00122-017-2960-y

Li, W., Zheng, D.-H., Van, K., and Lee, S.-H. (2008). QTL mapping for major agronomic traits across two years in soybean (Glycine max L. Merr.). *J. Crop. Sci. Biotechnol.* 11, 171–190.

Liang, H., Xu, L., Yu, Y., Yang, H., Dong, W., and Zhang, H. (2016). Identification of QTLs with main, epistatic and QTL by environment interaction effects for seed shape and hundred-seed weight in soybean across multiple years. *J. Genet.* 95, 475–477. doi: 10.1007/s12041-016-0648-8

Liu, D., Yan, Y., Fujita, Y., and Xu, D. (2018). Identification and validation of QTLs for 100-seed weight using chromosome segment substitution lines in soybean. *Breeding Science* 68, 442–448. doi: 10.1270/jsbbs.17127

Liu, N., Li, M., Hu, X., Ma, Q., Mu, Y., Tan, Z., et al. (2017). Construction of high-density genetic map and QTL mapping of yield-related and two quality traits in soybean RILs population by RAD-sequencing. *BMC Genom.* 18:466.

Lu, X., Xiong, Q., Cheng, T., Li, Q.-T., Liu, X.-L., Bi, Y.-D., et al. (2017). A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol. Plant* 10, 670–684. doi: 10.1016/j.molp.2017.03.006

Ma, J., Wingen, L. U., Orford, S., Fenwick, P., Wang, J., and Griffiths, S. (2015). Using the UK reference population Avalon× Cadenza as a platform to compare breeding strategies in elite Western European bread wheat. *Mol. Breeding* 35:70.

Mahmoud, A. F., Abou-Elwafa, S. F., and Shehzad, T. (2018). Identification of charcoal rot resistance QTLs in sorghum using association and in silico analyses. *J. Appl. Genet.* 59, 243–251. doi: 10.1007/s13353-018-0446-5

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.A.* 107, 19579–19584. doi: 10.1073/pnas.1014419107

Meng, Y., Chen, F., Shuai, H., Luo, X., Ding, J., Tang, S., et al. (2016). Karrikins delay soybean seed germination by mediating abscisic acid and gibberellin biogenesis under shaded conditions. *Sci. Rep.* 6, 1–12.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T., Yano, M., Bhatia, C., et al. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breeding* 3, 87–103.

Moongkanna, J., Nakasathien, S., Novitzky, W., Kwanyuen, P., Sinchaisri, P., and Srinives, P. (2011). SSR markers linking to seed traits and total oil content in soybean. *Thai. J. Agric. Sci.* 44, 233–241.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* 7:277.

Nyquist, W. E., and Baker, R. (1991). Estimation of heritability and prediction of selection response in plant populations. *Crit. Rev. Plant Sci.* 10, 235–322. doi: 10.1080/07352689109382313

Omokhafe, K., and Alika, J. (2004). Clonal variation and correlation of seed characters in Hevea brasiliensis Muell. Arg. *Industrial Crops Products* 19, 175–184. doi: 10.1016/j.indcrop.2003.09.004

Panthee, D., Pantalone, V., West, D., Saxton, A., and Sams, C. (2005). Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Sci.* 45, 2015–2022. doi: 10.2135/cropsci2004.0720

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery

and genotyping in model and non-model species. *PLoS One* 7:e37135. doi: 10.1371/journal.pone.0037135

Qi, Z.-M., Wu, Q., Han, X., Sun, Y.-N., Du, X.-Y., Liu, C.-Y., et al. (2011). Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. *Euphytica* 179, 499–514. doi: 10.1007/s10681-011-0386-1

Qi, Z., Xiaoying, Z., Huidong, Q., Dawei, X., Xue, H., Hongwei, J., et al. (2017). Identification and validation of major QTLs and epistatic interactions for seed oil content in soybeans under multiple environments based on a high-density map. *Euphytica* 213:162.

Salas, P., Oyarzo-Llaipen, J., Wang, D., Chase, K., and Mansur, L. (2006). Genetic mapping of seed shape in three populations of recombinant inbred lines of soybean (Glycine max L. Merr.). *Theor. Appl. Genet.* 113, 1459–1466. doi: 10.1007/s00122-006-0392-1

Severin, A. J., Woody, J. L., Bolon, Y.-T., Joseph, B., Diers, B. W., Farmer, A. D., et al. (2010). RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol.* 10:160. doi: 10.1186/1471-2229-10-160

Sun, Y.-N., Pan, J.-B., Shi, X.-L., Du, X.-Y., Wu, Q., Qi, Z.-M., et al. (2012). Multi-environment mapping and meta-analysis of 100-seed weight in soybean. *Mol. Biol. Rep.* 39, 9435–9443. doi: 10.1007/s11033-012-1808-4

Tao, Y., Mace, E. S., Tai, S., Cruickshank, A., Campbell, B. C., Zhao, X., et al. (2017). Whole-genome analysis of candidate genes associated with seed size and weight in sorghum bicolor reveals signatures of artificial selection and insights into parallel domestication in cereal crops. *Front. Plant Sci.* 8:1237.

Tasma, I., Lorenzen, L., Green, D., and Shoemaker, R. (2001). Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. *Mol. Breeding* 8, 25–35.

Teng, W., Feng, L., Li, W., Wu, D., Zhao, X., Han, Y., et al. (2017). Dissection of the genetic architecture for soybean seed weight across multiple environments. *Crop Pasture Sci.* 68, 358–365. doi: 10.1071/cp16462

Teng, W., Han, Y., Du, Y., Sun, D., Zhang, Z., Qiu, L., et al. (2009). QTL analyses of seed weight during the development of soybean (*Glycine max* L. Merr.). *Heredity* 102, 372–380. doi: 10.1038/hdy.2008.108

Tewodros, M., and Zelalem, B. (2016). Advances in quantitative trait loci, mapping and importance of markers assisted selection in plant breeding research. *International J. Plant Breeding Genet.* 10, 58–68. doi: 10.3923/ijpbg.2016.58.68

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129.

Tomooka, N., Vaughan, D., Maxted, N., and Moss, H. (2002). *The Asian Vigna: Genus Vigna Subgenus Ceratotropis Genetic Resources*. Berlin: Springer Science & Business Media.

Vieira, A. J. D., Oliveira, D.A.d, Soares, T. C. B., Schuster, I., Piovesan, N. D., Martínez, C. A., et al. (2006). Use of the QTL approach to the study of soybean trait relationships in two populations of recombinant inbred lines at the F7 and F8 generations. *Braz. J. Plant Physiol.* 18, 281–290. doi: 10.1590/s1677-04202006000200004

Wang, D., Bales-Arcelo, C., Zhang, Z., Gu, C., Difonzo, C. D., Zhang, G., et al. (2019). *Sources of Aphid Resistance in Soybean Plants". Google Patents US14/099,469.*

Wang, S., Basten, C., and Zeng, Z. (2006). *Windows QTL Cartographer 2.5. Dep. of Statistics, North Carolina State Univ., Raleigh. Windows QTL cartographer 2.5*. Raleigh, NC: Dep. of Statistics, North Carolina State Univ.

Wu, D., Zhan, Y., Sun, Q., Xu, L., Lian, M., Zhao, X., et al. (2018). Identification of quantitative trait loci underlying soybean (*Glycine max* [L.] Merr.) seed weight including main, epistatic and QTL× environment effects in different regions of Northeast China. *Plant Breeding* 137, 194–202. doi: 10.1111/pbr.12574

Xie, F.-T., Niu, Y., Zhang, J., Bu, S.-H., Zhang, H.-Z., Geng, Q.-C., et al. (2014). Fine mapping of quantitative trait loci for seed size traits in soybean. *Mol. Breeding* 34, 2165–2178. doi: 10.1007/s11032-014-0171-7

Xie, M., Ming, Y., Shao, F., Jian, J., Zhang, Y., and Peng, Z. (2018). Restriction site-associated DNA sequencing for SNP discovery and high-density genetic map construction in southern catfish (Silurus meridionalis). *R. Soc. Open Sci.* 5:172054. doi: 10.1098/rsos.172054

Xin, D., Qi, Z., Jiang, H., Hu, Z., Zhu, R., Hu, J., et al. (2016). QTL location and epistatic effect analysis of 100-seed weight using wild soybean (*Glycine soja* Sieb. & Zucc.) chromosome segment substitution lines. *PLoS One* 11:e0149380. doi: 10.1371/journal.pone.0149380

Xing, G., Zhou, B., Wang, Y., Zhao, T., Yu, D., Chen, S., et al. (2012). Genetic components and major QTL confer resistance to bean pyralid (*Lamprosema indicata* Fabricius) under multiple environments in four RIL populations of soybean. *Theor. Appl. Genet.* 125, 859–875. doi: 10.1007/s00122-012-1878-7

Xu, Y., Li, H.-N., Li, G.-J., Wang, X., Cheng, L.-G., and Zhang, Y.-M. (2011). Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. Merr.). *Theor. Appl. Genet.* 122, 581–594. doi: 10.1007/s00122-010-1471-x

Yamanaka, N., Ninomiya, S., Hoshi, M., Tsubokura, Y., Yano, M., Nagamura, Y., et al. (2001). An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. *DNA Res.* 8, 61–72. doi: 10.1093/dnares/8.2.61

Yang, J., Hu, C., Hu, H., Yu, R., Xia, Z., Ye, X., et al. (2008). QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* 24, 721–723. doi: 10.1093/bioinformatics/btm494

Yang, J., Zhu, J., and Williams, R. W. (2007). Mapping the genetic architecture of complex traits in experimental populations. *Bioinformatics* 23, 1527–1536. doi: 10.1093/bioinformatics/btm143

Yang, K., Moon, J.-K., Jeong, N., Chun, H.-K., Kang, S.-T., Back, K., et al. (2011). Novel major quantitative trait loci regulating the content of isoflavone in soybean seeds. *Genes Genom.* 33, 685–692. doi: 10.1007/s13258-011-0043-z

Yao, D., Wang, P., Zhang, J., Liu, Z., Guan, S., Liu, S., et al. (2014). A QTL mapping analysis of main yield traits in soybean. *J. South China Agric. Univ.* 35, 41–46.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, W293–W297.

Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468. doi: 10.1093/genetics/136.4.1457

Zhang, F., Kang, J., Long, R., Yu, L.-X., Wang, Z., Zhao, Z., et al. (2019). High-density linkage map construction and mapping QTL for yield and yield components in autotetraploid alfalfa using RAD-seq. *BMC Plant Biol.* 19:165.

Zhang, H., Chen, J., Shan, S., Cao, F., Chen, G., Zou, Y., et al. (2020). Proteomic profiling reveals differentially expressed proteins associated with amylose accumulation during rice grain filling. *BMC Genom.* 21:714.

Zhang, L., Xu, X., Zhao, C., Shan, F., Yuan, S., and Sun, H. (2011). QTL analysis of plant height based on doubled haploid (DH) population derived from PTSMS wheat. *Mol. Plant Breeding* 2, 92–97

Zhang, W.-K., Wang, Y.-J., Luo, G.-Z., Zhang, J.-S., He, C.-Y., Wu, X.-L., et al. (2004). QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. *Theor. Appl. Genet.* 108, 1131–1139. doi: 10.1007/s00122-003-1527-2

Zhang, X., Hina, A., Song, S., Kong, J., Bhat, J. A., and Zhao, T. (2019). Whole-genome mapping identified novel "QTL hotspots regions" for seed storability in soybean (*Glycine max* L.). *BMC Genom.* 20:499.

Zhang, Y., Li, W., Lin, Y., Zhang, L., Wang, C., and Xu, R. (2018). Construction of a high-density genetic map and mapping of QTLs for soybean (*Glycine max*) agronomic and seed quality traits by specific length amplified fragment sequencing. *BMC Genom.* 19:641.

Zhao, B., Dai, A., Wei, H., Yang, S., Wang, B., Jiang, N., et al. (2016). ArabidopsisKLU homologue GmCYP78A72 regulates seed size in soybean. *Plant Mol. Biol.* 90, 33–47. doi: 10.1007/s11103-015-0392-0

Zhao, F., and Xu, S. (2012). Genotype by environment interaction of quantitative traits: a case study in barley. *G3 Genes Genom. Genet.* 2, 779–788. doi: 10.1534/g3.112.002980

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 09 June 2021
doi: 10.3389/fgene.2021.684702

Development and Exploitation of KASP Assays for Genes Underpinning Drought Tolerance Among Wheat Cultivars From Pakistan

Shoaib Ur Rehman[1], Muhammad Ali Sher[1], Muhammad Abu Bakar Saddique[1], Zulfiqar Ali[1], Mahmood Alam Khan[1], Xinguo Mao[2], Ahsan Irshad[3], Muhammad Sajjad[4], Rao Muhammad Ikram[5], Mahnoor Naeem[1] and Ruilian Jing[2]**

[1] *Institiute of Plant Breeding and Biotechnology, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan,* [2] *National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China,* [3] *National Engineering Laboratory of Crop Molecular Breeding, National Center of Space Mutagenesis for Crop Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China,* [4] *Department of Biosciences, COMSATS University Islamabad (CUI), Islamabad, Pakistan,* [5] *Department of Agronomy, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan*

High-throughput genotyping for functional markers offers an excellent opportunity to effectively practice marker-assisted selection (MAS) while breeding cultivars. We developed kompetitive allele-specific PCR (KASP) assays for genes conferring drought tolerance in common wheat (*Triticum aestivum* L.). In total, 11 KASP assays developed in this study and five already reported assays were used for their application in wheat breeding. We investigated alleles at 16 loci associated with drought tolerance among 153 Pakistani hexaploid wheat cultivars released during 1953–2016; 28 diploid wheat accessions (16 for AA and 12 for BB) and 19 tetraploid wheat (AABB) were used to study the evolutionary history of the studied genes. Superior allelic variations of the studied genes were significantly associated with higher grain yield. Favored haplotypes of *TaSnRK2.3*-1A, *TaSnRK2.3*-1B, *TaSnRK2.9*-5A, *TaSAP*-7B, and *TaLTPs*-1A predominated in Pakistani wheat germplasm indicating unconscious pyramiding and selection pressure on favorable haplotypes during selection breeding. *TaSnRK2.8*-5A, *TaDreb*-B1, *1-feh w3*, *TaPPH*-7A, *TaMOC*-7A, and *TaPARG*-2A had moderate to low frequencies of favorable haplotype among Pakistani wheat germplasm pointing toward introgression of favorable haplotypes by deploying functional markers in marker-assisted breeding. The KASP assays were compared with gel-based markers for reliability and phenotypically validated among 62 Pakistani wheat cultivars. Association analyses showed that the favorable allelic variations were significantly associated with grain yield-contributing traits. The developed molecular marker toolkit of the genes can be instrumental for the wheat breeding in Pakistan.

Keywords: gel-free markers, KASP markers, drought related genes, Pakistani wheat, genetic diversity

OPEN ACCESS

Edited by:
Awais Rasheed,
Quaid-i-Azam University, Pakistan

Reviewed by:
Faheem Shehzad Baloch,
Sivas Bilim ve Teknoloji Fakültesi, Turkey
Jindong Liu,
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, China

****Correspondence:***
Shoaib Ur Rehman
shoaib.rehman@mnsuam.edu.pk
Ruilian Jing
jingruilian@caas.cn

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *23 March 2021*
Accepted: *22 April 2021*
Published: *09 June 2021*

Citation:
Ur Rehman S, Ali Sher M, Saddique MAB, Ali Z, Khan MA, Mao X, Irshad A, Sajjad M, Ikram RM, Naeem M and Jing R (2021) Development and Exploitation of KASP Assays for Genes Underpinning Drought Tolerance Among Wheat Cultivars From Pakistan. Front. Genet. 12:684702. doi: 10.3389/fgene.2021.684702

INTRODUCTION

Crop improvement strategies have always circumambulated yield-enhancing genes. Therefore, exploitation of gene diversity in breeding germplasm and identification of superior genetic variations are prioritized activities in crop genetic improvement. Such exploitations enable breeders to identify desirable germplasm for breeding and to devise strategies for pyramiding superior genetic variations for targeted traits. Wheat is one of the most important cereal crops, and there is demand for a yield increase of up to 50% by 2050 (Curtis and Halford, 2014). Being a staple food crop in Pakistan, wheat growth and development is severely influenced by abiotic stress, resulting in a significant reduction in grain yield. Moreover, the genetic structure of modern Pakistani wheat cultivars built around only a few cultivars such as Bluebird, Kauz, Kalyansona, and Buho, and there is an urgent need to introduce new diversity for sustainable wheat production in Pakistan. Therefore, the utilization of genes conferring drought tolerance is regarded as an effective way to ensure high and sustainable yield in wheat. Marker-assisted selection (MAS) based on pyramiding superior alleles/haplotypes is considered as a potential strategy to wheat improvement for economically important traits. The challenge is to deploy such strategy in breeding programs in a time- and cost-efficient manner for different scenarios (Richards et al., 2014).

TaSnRK2.3-1A/1B (Miao et al., 2017), *TaSnRK2.9*-5A (Ur Rehman et al., 2019), *TaPARG*-2A (Li B. et al., 2016), *TaSAP*-7B (Wang et al., 2018), *TaPPH*-7A (Wang et al., 2019), and *TaMOC1*-7A (Zhang B. et al., 2015) are associated with higher grain yield under water stress conditions. *TaSnRK2.8*-5A associated with higher seedling biomass under normal conditions and water-soluble carbohydrates under limited irrigation conditions (Zhang et al., 2013), *TaLTPs*-1A associated with ideal plant height under drought conditions (Li Q. et al., 2016), and *TaDreb*-B1 (Wei et al., 2009) and *1-feh w3* (Zhang J. et al., 2015) are also reported as drought tolerance-conferring genes.

Functional markers (FMs) of the aforementioned genes were successfully applied in Chinese wheat cultivars and provided the concept of screening of genotypes for wheat breeding in Pakistan. Various single-marker methods have been developed for single-nucleotide polymorphism (SNP) genotyping, such as cleaved amplified polymorphic sequences (CAPS), and derived cleaved amplified polymorphic sequences (dCAPS). The CAPS and dCAPS markers are relatively low throughput, laborious, and cost ineffective, as they rely upon site-specific cleavage of PCR products with restriction enzymes and require gel electrophoresis to separate products. At present, more than 150 FMs are available for important genes, giving plant breeders a molecular toolkit for the selection of favorable traits (Liu et al., 2012). Although FMs are available for wheat, their deployment retains limited courtesy cost and time needed to exploit larger populations.

Kompetitive allele-specific PCR (KASP) is a uniplex and flexible genotyping platform which achieves high throughput in a time- and cost-effective way (Semagn et al., 2014). Conversion of conventional FMs into KASP assays could greatly speed up improvement in breeding programs. Therefore, the aims of the present study are (i) to develop KASP-based assays of FMs for higher grain yield and drought-conferring genes in wheat, (ii) to perform marker trait association analyses among Pakistani wheat cultivars and to investigate the distribution of FMs in wheat cultivars across Pakistan, and (iii) to know the genetic diversity of given genes among diploid, tetraploid, and hexaploid wheat. The information will be useful in breeding wheat for higher grain yield and drought tolerance by MAS.

MATERIALS AND METHODS

Germplasm

One hundred and fifty-three wheat cultivars from Pakistan released during 1953–2016 were used to identify favorable haplotype frequencies of drought tolerance-conferring genes (**Supplementary Table 1**). The wheat collection comprised four groups based on time of release, i.e., pre-green revolution 1953–1965, post-green revolution 1966–1985, 1986–2005, and post-2005. Besides *Triticum aestivum* germplasm, nine genotypes of *T. urartu* (AA), four genotypes of *T. boeoticum Bioss* (AA), three genotypes of *T. monococcum* (AA), 12 genotypes *Aegilops speltoides* (BB), five genotypes of *T. dicoccum* L. (AABB), three genotypes of *T. persicum* Vav (AABB), five genotypes of *T. dicoccoides* Koern (AABB), four genotypes of *T. polonicum* L., and two genotypes of *T. turgidum* (AABB) (**Supplementary Table 2**) were also used to identify the polymorphic information contents (PIC) and gene diversity of the studied genes. Wheat genotypes such as Chinese Spring, MexiPak-65, and Parwaz-94 were used as controls for the identification of particular alleles. A subset of 62 wheat genotypes (55 modern cultivars and seven landraces) from 153 wheat cultivars were also grown at MNS University of Agriculture, Multan and Quaid-e-Azam University, Islamabad, under two water regimes, i.e., under water stress at flowering stage and under normal conditions. Randomized complete block design was followed with duplicates at both locations. Standard agronomic practices were followed to ensure proper plant stand. Wheat sown under normal conditions was irrigated initially after 25 days of sowing followed by irrigation at flowering and grain filling stages. The wheat grown under the water stress regime was irrigated only once after 25 days of sowing. Precipitation mainly occurred at the end of March at both locations. Water contents of different soil profiles are given in **Supplementary Table 3**. Each experimental plot was 6 m in length with six rows having a row spacing of 30 cm with ∼40 seeds per row. The cultivars were sown in mid-November 2019 and harvested in April of the following year. These 62 genotypes were planted for traits, i.e., plant height (PH), thousand kernel weight (TKW), and grains per spike (GPS), under both water regimes. Association analysis was performed on the average of all the parameters from both water regimes, and a phenotypic comparison of allelic variations was presented.

Genotyping

Five KASP assays were selected from published reports including two SNPs for *TaSnRK2.9*-5A (Ur Rehman et al., 2019) and one each for *TaMOC1*-7A, *TaDreb*-1B, and *1-feh*

w3 (Rasheed et al., 2016). The remaining 11 KASP assays were developed in this study (**Figure 1**). The information on the selected genes, the sequence polymorphism, KASP assays, and their sources are provided in **Supplementary Table 4**.

For the KASP assays developed in this study, the nucleotide sequences of drought tolerance-causing genes were retrieved from the published literature. The diagnostic polymorphic sites were identified, and KASP primers were developed following standard KASP guidelines. The allele-specific primers were designed carrying the standard FAM and HEX tails and with the targeted SNP at the 3′ end. A common primer was designed so that the total amplicon length was less than 120 bp. The primer mixture comprised 46 μl ddH_2O, 30 μl common primer (100 μM), and 12 μl of each tailed primer (100 μM). Assays were tested in 96-well formats and set up as 5 μl reaction mixture (2.4 μl of 25 ng/μl DNA, 2.4 μl of 2 × KASP master mixture, 0.06 μl of primer mixture, 0.04 $MgCl_2$, and 0.1 μl of ddH_2O). PCR cycling was performed using the following protocol: hot start at 95°C for 15 min, followed by 10 touchdown cycles (95°C for 20 s; touchdown at 65°C initially and decreasing by −1°C per cycle for 25 s), followed by 32–35 additional cycles of annealing (95°C for 15 s, 57°C for 1 min). Fluorescence levels were detected and analyzed by using CFX Connect Real-Time PCR detection system (Bio-Rad® laboratories Inc. United States) and QuantStudio 7 Flex Real-Time PCR systems.

Statistical Analyses

Student's *t*-test at $P < 0.05$ was used to check the effect of SNP/haplotype on the studied agronomic traits. Allele/haplotype frequencies were calculated for all loci. PIC and gene diversity (H_e) were calculated for each locus using https://www.gene-calc.pl/pic.

RESULTS

Comparison of KASP Markers and Conventional Gel-Based PCR Markers

The results from KASP markers were compared to contrasting gel-based markers for all the genes. All studied KASP assays produced consistent results when compared to conventional PCR markers in 23 diverse wheat genotypes (**Supplementary Table 5**),

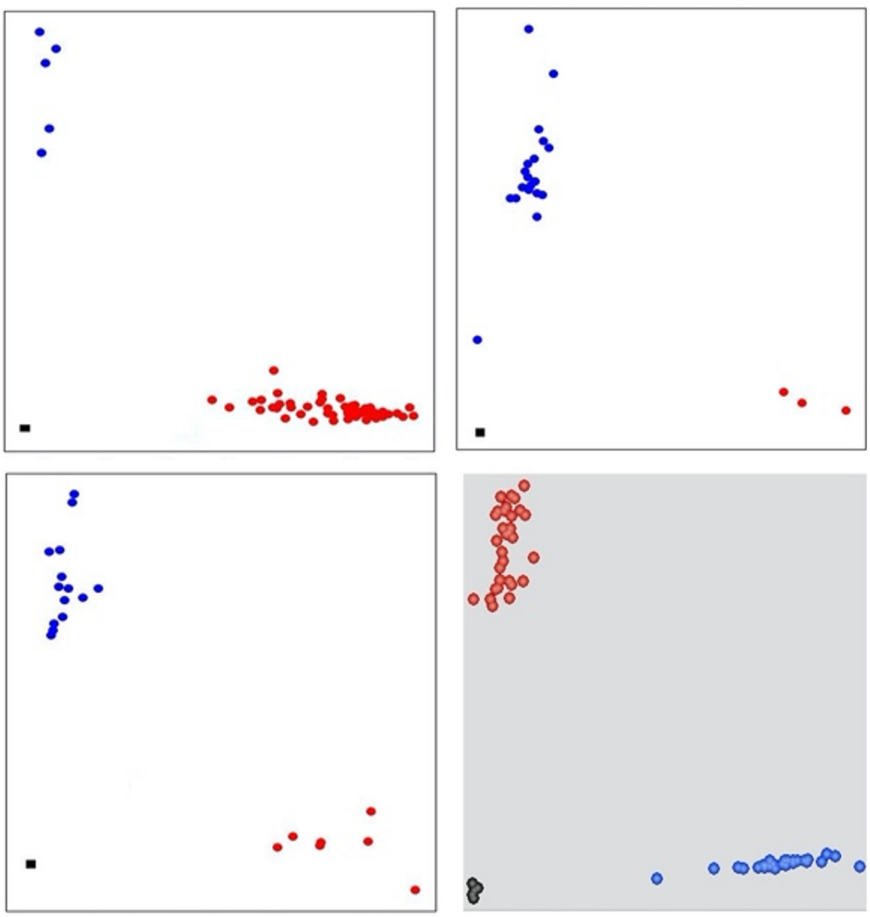


FIGURE 1 | Scatter plot for selected KASP assays showing clustering of genotypes on the *Y*- and *X*-axes. Genotypes colored red have a HEX-type allele; genotypes colored blue have a FAM-type allele; black dots represent non-template control. Top left and right corners, KASP assays for *TaSnRK2.3*-1A and *TaSnRK2.3*-1B, respectively. Bottom left and right corner, KASP assays for *TaSAP*-7B and *TaLTPs*, respectively.

but for *TaPARG*-2A-KASP-10 (C/T), the concentration of each tailed primer increased up to 15 μM to obtain satisfactory results.

KASP Assays for Grain Yield-Contributing Traits

Association analyses of allelic variations of the studied genes showed that the favorable allelic variations were significantly associated with higher grain yield traits among the studied 62 Pakistani wheat germplasm (**Figure 2**). Haplotypes associated with grain-related traits are *Hap*-1 (CA) of *TaSnRK2.3*-1A and *Hap*-1 (CG) of *TaSnRK2.3*-1B which are favored for higher GPS and TKW (**Figure 2**). *Hap*-1/3 (A-A/G allele accessions) of *TaSnRK2.8*-5A showed a non-significant association with higher GPS and TKW. *Hap*-1 (TA) and *Hap*-4 (CA) of *TaSnRK2.9*-5A are also associated with higher GPS and TKW (**Figure 2**). For *TaSAP*-7B, accessions carrying the "C" allele possess ideal plant height (99 cm) and higher GPS and TKW. Accessions carrying *Hap*-H of *TaMOC1*-7A possess higher GPS and TKW than *Hap*-L. *Hap*-3 (*GC*) of *TaLTPs* associated with ideal plant height (100 cm) and higher GPS and TKW. *TaDreb*-B1 and *1-feh w3* accessions carrying the *Hap*-1 "A-allele" and *Hap*-1 "C-allele," respectively, are associated with higher grain yield contributing parameters (**Figure 2**). *Hap*-1 (CC) of *TaPARG*-2A showed an association with higher GPS. Favorable allelic variation of *TaPPH*-7A-1 ("A" allele) also showed an association with higher GPS among Pakistani wheat cultivars (**Figure 2**).

In general, frequencies of favored haplotypes and/or alleles were higher in 153 Pakistani wheat germplasm released from 1953 to 2016. For *TaSnRK2.3*-1A, 127 (83.01%) Pakistani wheat cultivars had the desirable haplotype (*Hap*-1). The favorable haplotype (*Hap*-1) of *TaSnRK2.3*-1B was also present in 127 (83.01%) Pakistani wheat cultivars (**Table 1**). For *TaSnRK2.8*-5A, the frequency of preferred allele "A" was low (33.99%) in given wheat cultivars. The frequency of favored haplotypes for *TaSnRK2.9*-5A (*Hap*-1 and *Hap*-4) was 75.82% and 2.61%, respectively. The favored allele for *TaSAP*-7B was predominant in 129 (84.31%) wheat cultivars; at *TaMOC1*-7A, 33 (21.57%) desirable haplotypes (*Hap*-H). The favorable haplotype (*Hap*-3) of *TaLTPs* was present in 146 (95.42%) wheat cultivars. Superior alleles for *TaDreb*-B1, *1-feh w3*, and *TaPPH-7A* were present in 48 (31.37%), 44 (28.75%), and 67 (43.79%) wheat cultivars. The favorable haplotypes of *TaPARG*-2A (*Hap*-2 (1.31%) and *Hap*-3 absent) were present in very low frequencies among given wheat germplasm (**Table 1**).

Certain combinations of two or more desirable alleles or haplotypes tended to occur in higher frequencies in one group more than the other; for example, TaSnRK2.3-1A + TaSnRK2.3-1B + TaSnRK2.9-5A + TaSAP-7B + TaLTPs-1A were present in 110 (71.89%) wheat cultivars.

Selection Frequencies of Favored Alleles/Haplotypes in Pakistan Since 1953

Since 1953, the frequency distribution of favored haplotypes of given drought tolerance-responsible genes varied among Pakistani wheat germplasm (**Figure 3**). For *TaSnRK2.3*-1A/1B, 83.01% wheat accessions contained favored haplotypes. For *TaSnRK2.9*-5A, *TaLTPs*-1A, and *TaSAP*-7B, 78.41%, 95.42, and 84.31% Pakistani wheat accessions possessed favorable allelic variations, respectively. Based on released time, Pakistani

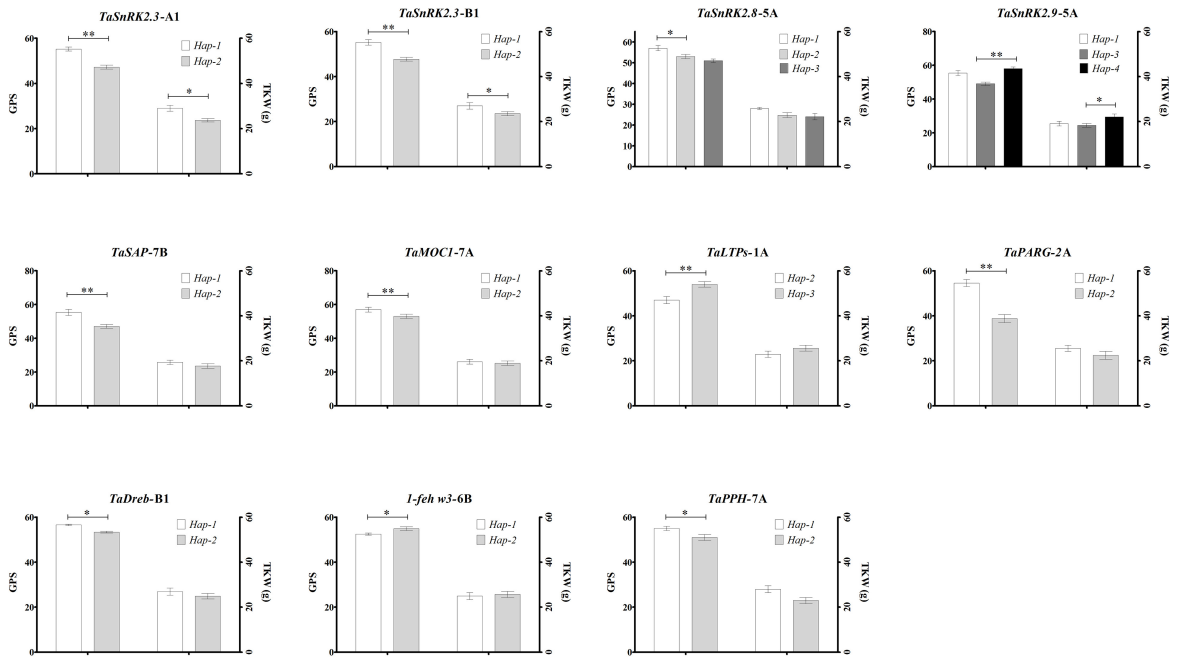


FIGURE 2 | Phenotypic comparison of allelic variations. Traits are thousand kernel weight—TKW, grains per spike—GPS, and plant height—PH. *$P < 0.05$, **$P < 0.01$. Error bar denotes standard deviation.

TABLE 1 | Allelic frequencies in 153 Pakistani wheat cultivars.

Gene	Locus	Haplotype	Genotype	Number of accession	Phenotype	Frequency (%)	Reference (of phenotype)
TaSnRK2.3	*TaSnRK2.3*-1A	*Hap*-1	CA	127	Higher TKW	83.01	Miao et al., 2017
		Hap-2	TG	24		15.69	
		Hap-3	CG	2		1.31	
	TaSnRK2.3-1B	*Hap*-1	CG	127	Higher TKW	83.01	
		Hap-2	TC	7		4.58	
		Hap-3	CC	19		12.42	
TaSnRK2.8	*TaSnRK2.8*-5A	*Hap*-1	A	52	Seedling biomass and water-soluble carbohydrates	33.99	Zhang et al., 2013
		Hap-2	G	69		45.10	
			A/G	32		20.92	
TaSnRK2.9	*TaSnRK2.9*-5A	*Hap*-1	TA	116	Higher TKW	75.82	Ur Rehman et al., 2019
		Hap-2	TC	2		1.31	
		Hap-3	CC	31		20.26	
		Hap-4	CA	4	Higher GPS	2.61	
TaSAP	*TaSAP*-7B	*Hap*-1	C	129	Higher TKW and short PH	84.31	Wang et al., 2018
		Hap-2	T	24		15.69	
TaMOC	*TaMOC1*-7A	*Hap*-H	G	33	Higher grain number	21.57	Zhang B. et al., 2015
		Hap-L	A	120	Lower grain number	78.43	
TaLTPs	*TaLTPs*-1A	*Hap*-1	AC	0		0.00	Li Q. et al., 2016
		Hap-2	GT	7		4.58	
		Hap-3	GC	146	Ideal plant height	95.42	
TaPARG	*TaPARG*-2A	*Hap*-1	CC	151		98.69	Li B. et al., 2016
		Hap-2	TC	2	Lower PH, ETN, and higher TKW	1.31	
		Hap-3	TT	0		0.00	
TaDreb	*TaDreb*-B1	*Hap*-1	A	48	Drought tolerance	31.37	Wei et al., 2009
		Hap-2	C	105		68.63	
1-FEH W3	*1-feh w3*-6B	*Hap*-1	C	44	Drought tolerance	28.76	Zhang J. et al., 2015
		Hap-2	T	109		71.24	
TaPPH	*TaPPH*-7A	*Hap*-1	A	67	Higher TKW and short PH	43.79	Wang et al., 2019
		Hap-2	G	86		56.21	

wheat cultivars were divided into four groups. From 1953 to 2016, the frequency of favored haplotypes for *TaSnRK2.3*-1A, *TaSnRK2.3*-1B, *TaSnRK2.9*-5A, *TaSAP*-7B, and *TaLTPs*-1A increased remarkably. The combined frequencies of favored alleles/haplotypes of aforementioned genes increased from 35.78% in the pre-1965s to 93% in the post-2005s, showing a progressive selection of favored alleles/haplotypes over the years. The frequencies of favored alleles/haplotypes of other drought-conferring genes remain low (<22%) since 1953, suggesting the potential of favored allele/haplotype introgression through FMs developed in this study.

Diversity Pattern in Wheat Germplasm

To survey the evolutionary history of given drought tolerance-responsible genes, we analyzed the given genes in wheat progenitor accessions. The results showed that during polyploidization events, diversity decreased in a given set of genes. Diploid wheat accessions showed an average of 0.265 (PIC) and 0.0.331 (H_e) in the investigated genes. The studied tetraploid (AABB) accessions showed an average 0.361 (PIC) and 0.418 (H_e) in the given genes. Hexaploid Pakistani wheat showed an average 0.270 (PIC) and 0.317 (H_e) in the given genes (**Table 2**). The results depict that the overall diversity for the given genes (except *TaSnRK2.8*-5A, *TaDreb*-B1, and *TaPPH*-7A) among Pakistani wheat accession reduced.

DISCUSSION

Novel genomic tools provide an opportunity in meeting the challenge of enhanced genetic gain to safeguard sustainable production. The application of molecular markers to accelerate MAS has proven successful in wheat breeding programs (Rasheed et al., 2016). Moreover, the concepts of MAS in wheat are now transformed into genomic selection methods to improve genetic gains (Zhao et al., 2019). It has been reported that the use of functional markers for individual genes can significantly improve prediction accuracies (Rutkoski et al., 2014). Using functional markers and genes for wheat breeding, the appropriate breeding material should be selected based on production needs (Hao et al., 2020). Breeder friendliness, high throughput, and cost-effectiveness are the main considerations in selecting an appropriate genotyping platform for genomic selection and MAS (Semagn et al., 2014). Here we have demonstrated

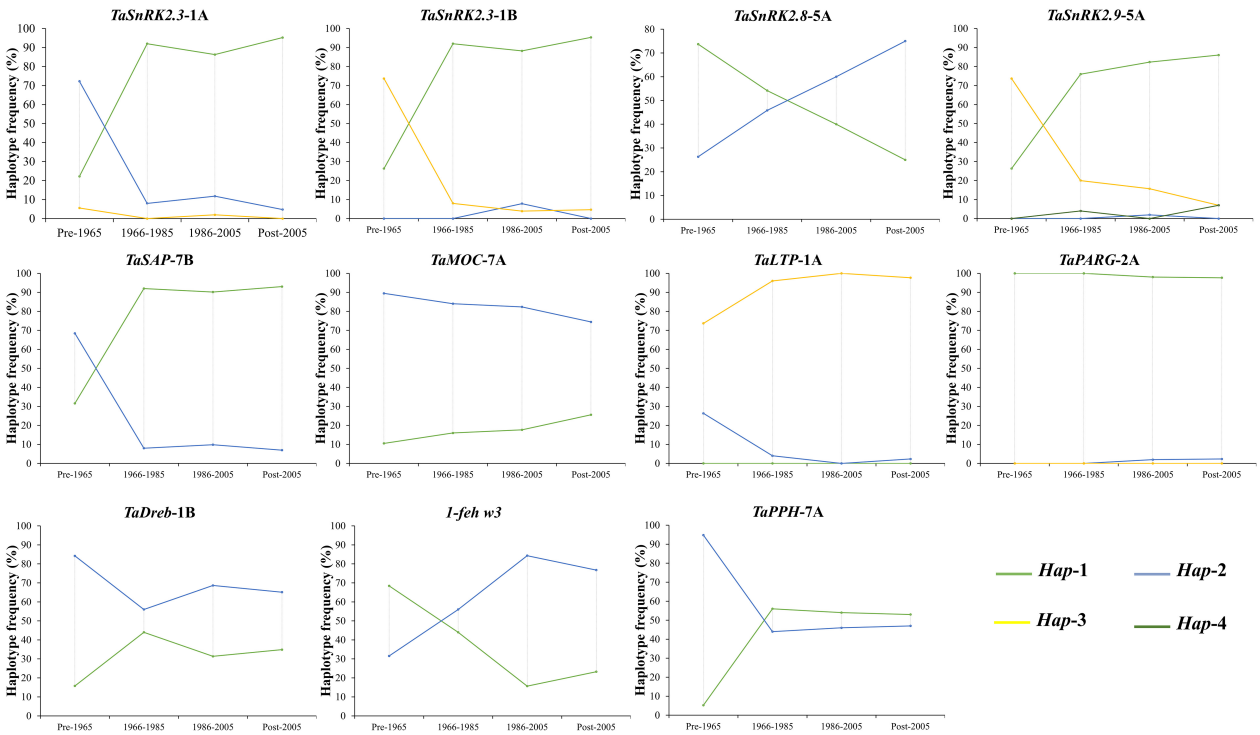


FIGURE 3 | The frequency of drought-tolerant genes' haplotypes in Pakistani wheat cultivars released in different eras.

TABLE 2 | Polymorphic information contents and gene diversity in diploid, tetraploid, and hexaploid Pakistani wheat.

Genes	Diploid wheat			Tetraploid wheat			Hexaploid wheat		
	Alleles/Haplotypes	PIC	H_e	Alleles/Haplotypes	PIC	H_e	Alleles/Haplotypes	PIC	H_e
TaSnRK2.3-1A	2	0.304	0.375	3	0.414	0.498	3	0.250	0.285
TaSnRK2.3-1B				3	0.504	0.570	3	0.270	0.294
TaSnRK2.8-5A	2	0.370	0.490	2	0.357	0.466	3	0.562	0.637
TaSnRK2.9-5A				3	0.512	0.575	4	0.334	0.381
TaSAP-7B				2	0.276	0.331	2	0.232	0.268
TaMOC1-7A				2	0.368	0.487	2	0.284	0.343
TaLTPs-1A				2	0.276	0.331	3	0.090	0.095
TaPARG-2A	2	0.121	0.130	2	0.487	0.368	3	0.019	0.019
TaDreb-B1				2	0.090	0.09	2	0.335	0.427
1-feh w3				2	0.374	0.498	2	0.327	0.411
TaPPH-7A				2	0.310	0.384	2	0.371	0.492
Sum		0.796	0.995	25	3.973	4.607	29	2.707	3.165
Average	2	0.265	0.331	2.27	0.361	0.418	2.64	0.270	0.317

the effectiveness of newly developed KASP assays for genes conferring drought tolerance in wheat. These assays offer fast-track ways to deploy drought tolerance-causing genes in wheat improvement in a cost-effective manner.

Reliability of Developed KASP Assays

Identification and validation of SNPs is a significant challenge in wheat due to the large genome size, polyploidy, and high percentage of repetitive sequences (Ramirez-Gonzalez et al., 2015). Hence, it is necessary to validate the SNPs. The developed KASP assays were validated for reliability. The KASP assays were compared to their equivalent gel-based markers on a small but diverse set of Chinese wheat germplasm and check cultivars with known alleles at each gene. One KASP assay (*TaSnRK2.8*-5A-KASP-7) showed inconsistent outcomes in the form of heterozygous conditions for the given alleles. Overall, the conversion rate for newly developed KASP assays was > 98% and were able to convert gel-based PCR markers into breeder friendly gel-free KASP markers.

Allelic Variation at Loci Influencing Grain Related Traits

MAS of superior alleles in breeding programs is important for the ongoing improvement of wheat. The deployment of superior alleles in improved cultivars could be enhanced if efficient molecular diagnostics are available (Rasheed et al., 2017). TKW, GPS, and PH are important yield-contributing traits in wheat, and recently several genes affecting these traits were cloned. Favorable allelic variations of the genes studied in this work have been reported to be associated with higher grain weight and higher grain number under normal and water stress conditions in Chinese wheat germplasm (**Table 1**). The investigation of Pakistani wheat germplasm for these genes is necessary for assessing the effect of selection pressure on favorable haplotypes and to alert wheat breeders for these favorable variations for grain yield. Our results suggested strong selection pressure on favorable haplotypes at *TaSnRK2.3*-1A/1B, *TaSnRK2.9*-5A, *TaSAP*-7B, and *TaLTPs*-1A among Pakistani wheat accessions. A moderate frequency of favored haplotypes was observed at *TaSnRK2.8*-5A, *TaDreb*-B1, *TaPPH*-7A, and *1-feh w3*, indicating that exploitation of these alleles may be continued to gain a yield increase in Pakistan. This unconscious selection of favorable haplotypes is likely due to the high linkage disequilibrium of important genes selected during selection breeding. The given Pakistani wheat germplasm had high frequencies of unfavorable allelic variations for *TaMOC*-7A and *TaPARG*-2A, suggesting the potential of favorable haplotype introgression through functional markers. Previously, selection pressure was observed on *TaSnRK2.3*-1A/B (Miao et al., 2017), *TaSnRK2.9*-5A (Ur Rehman et al., 2019), *TaSAP*-7B (Wang et al., 2018), *TaMOC*-7A (Zhang B. et al., 2015), *TaLTPs*-1A (Li Q. et al., 2016), *TaPARG*-2A (Li B. et al., 2016), and *TaPPH*-7A (Wang et al., 2019) favorable allelic variations in Chinese wheat cultivars where the frequencies of favored haplotypes had gradually increased from the beginning of the last century.

Both PIC and H_e values were higher in tetraploid wheat as compared to Pakistani hexaploid wheat. Lower PIC and H_e values in Pakistani hexaploid wheat concluded a severe domestication genetic bottleneck. The probable reason for bottlenecking is the genetic structure mainly built around relatively few cultivars such as Bluebird, Kalyansona, Kauz, and Buho, causing serious threat by narrowing the genetic base for drought tolerance-conferring genes. Hence, introgression from geographically different wheat may be a preferred strategy to introduce novel allelic variations at loci conferring drought tolerance for sustainable production.

Finally, our work included a set of genes conferring drought tolerance. We developed a breeding toolkit for high-throughput and cost-effective genotyping for drought-conferring genes in wheat. We believe that this toolkit can accelerate breeding efforts to select diverse and pyramid favorable allelic variations in wheat breeding programs in Pakistan.

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

SU was responsible for conceptualization. SU, MA, MSd, MK, MN, RI, and AI performed the experiments and analyzed the data. SU wrote the manuscript. ZA, XM, MSj, and RJ reviewed the manuscript and assisted in the completion of the experiments. All authors contributed to the article and approved the submitted version.

FUNDING

The study was supported by the Higher Education Commission of Pakistan (SRGP-2629) and the National Key R&D Program of China (2017YFD0300202).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.684702/full#supplementary-material

REFERENCES

Curtis, T., and Halford, N. G. (2014). Food security: the challenge of increasing wheat yield and the importance of not compromising food safety. *Ann. Appl. Biol.* 164, 354–372. doi: 10.1111/aab.12108

Hao, C., Jiao, C., Hou, J., Li, T., Liu, H., Wang, Y., et al. (2020). Resequencing of 145 landmark cultivars reveals asymmetric sub-genome selection and strong founder genotype effects on wheat breeding in China. *Mol. Plant* 13, 1733–1751. doi: 10.1016/j.molp.2020.09.001

Li, B., Li, Q., Mao, X., Li, A., Wang, J., Chang, X., et al. (2016). Two novel AP2/EREBP transcription factor genes TaPARG have pleiotropic functions on plant architecture and yield-related traits in common wheat. *Front. Plant Sci.* 7:1191. doi: 10.3389/fpls.2016.01191

Li, Q., Wang, J. Y., Khan, N., Chang, X. P., Liu, H. M., and Jing, R. L. (2016). Polymorphism and association analysis of a drought-resistant gene TaLTPs in wheat. *J. Integr. Agric.* 15, 1198–1206. doi: 10.1016/S2095-3119(15)61189-3

Liu, Y., He, Z., Appels, R., and Xia, X. (2012). Functional markers in wheat: current status and future prospects. *Theor. Appl. Genet.* 125, 1–10. doi: 10.1007/s00122-012-1829-3

Miao, L., Mao, X., Wang, J., Liu, Z., Zhang, B., Li, W., et al. (2017). Elite haplotypes of a protein kinase gene TaSnRK2.3 associated with important agronomic traits in common wheat. *Front. Plant Sci.* 8:368. doi: 10.3389/fpls.2017.00368

Ramirez-Gonzalez, R. H., Uauy, C., and Caccamo, M. (2015). PolyMarker: a fast polyploid primer design pipeline. *Bioinformatics* 31, 2038–2039. doi: 10.1093/bioinformatics/btv069

Rasheed, A., Wen, W., Gao, F., Zhai, S., Jin, H., Liu, J., et al. (2016). Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theor. Appl. Genet.* 129, 1843–1860. doi: 10.1007/s00122-016-2743-x

Rasheed, A., Hao, Y., Xia, X. C., Khan, A., Xu, Y., Varsheny, R. K., et al. (2017). Crop breeding chips and genotyping platforms: progress, challenges and perspectives. *Mol. Plant* 10, 1047–1064. doi: 10.1016/j.molp.2017.06.008

Richards, R. A., Hunt, J. R., Kirkegaard, J. A., and Passioura, J. B. (2014). Yield improvement and adaptation of wheat to water-limited environments in Australia—a case study. *Crop. Pasture Sci.* 65, 676–689. doi: 10.1071/CP13426

Rutkoski, J. E., Poland, J. A., Singh, R. P., Huerta-Espino, J., Bhavani, S., Barbier, H., et al. (2014). Genomic selection for quantitative adult plant stem rust resistance in wheat. *Plant Genome* 7, 1–10. doi: 10.3835/plantgenome2014.02.0006

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33, 1–14. doi: 10.1007/s11032-013-9917-x

Ur Rehman, S., Wang, J., Chang, X., Zhang, X., Mao, X., and Jing, R. (2019). A wheat protein kinase gene TaSnRK2.9-5A associated with yield contributing traits. *Theor. Appl. Genet.* 132, 907–919. doi: 10.1007/s00122-018-3247-7

Wang, H., Wang, S., Chang, X., Hao, C., Sun, D., and Jing, R. (2019). Identification of TaPPH-7A haplotypes and development of a molecular marker associated with important agronomic traits in common wheat. *BMC Plant Biol.* 19:296. doi: 10.1186/s12870-019-1901-0

Wang, Y., Xu, Q., Chang, X., Hao, C., Li, R., and Jing, R. (2018). A dCAPS marker developed from a stress associated protein gene TaSAP7-B governing grain size and plant height in wheat. *J. Integr. Agric.* 17, 276–284. doi: 10.1016/S2095-3119(17)61685-X

Wei, B., Jing, R., Wang, C., Chen, J., Mao, X., Chang, X., et al. (2009). Dreb1 genes in wheat (*Triticum aestivum* L.): development of functional markers and gene mapping based on SNPs. *Mol. Breed.* 23, 13–22. doi: 10.1007/s11032-008-9209-z

Zhang, B., Liu, X., Xu, W., Chang, J., Li, A., Mao, X., et al. (2015). Novel function of a putative MOC1 ortholog associated with spikelet number per spike in common wheat. *Sci. Rep.* 5:12211. doi: 10.1038/srep12211

Zhang, H., Mao, X., Zhang, J., Chang, X., and Jing, R. (2013). Single-nucleotide polymorphisms and association analysis of drought-resistance gene TaSnRK2.8 in common wheat. *Plant Physiol. Biochem.* 70, 174–181. doi: 10.1016/j.plaphy.2013.04.010

Zhang, J., Xu, Y., Chen, W., Dell, B., Vergauwen, R., Biddulph, B., et al. (2015). A wheat 1-feh-w3 variant underlies enzyme activity for stem WSC remobilization to grain under drought. *New Phytol.* 205, 293–305. doi: 10.1111/nph.13030

Zhao, J., Wang, Z., Liu, H., Zhao, J., Li, T., Hou, J., et al. (2019). Global status of 47 major wheat loci controlling yield, quality, adaptation and stress resistance selected over the last century. *BMC Plant Biol.* 19:5. doi: 10.1186/s12870-018-1612-y

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 June 2021
doi: 10.3389/fgene.2021.682324

Asymmetric Somatic Hybridization Affects Synonymous Codon Usage Bias in Wheat

Wenjing Xu[1], *Yingchun Li*[1], *Yajing Li*[1], *Chun Liu*[1], *Yanxia Wang*[2], *Guangmin Xia*[1] *and Mengcheng Wang*[1]*

[1] The Key Laboratory of Plant Development and Environmental Adaption, Ministry of Education, School of Life Science, Shandong University, Jinan, China, [2] Shijiazhuang Academy of Agriculture and Forestry Sciences, Shijiazhuang, China

Asymmetric somatic hybridization is an efficient strategy for crop breeding by introducing exogenous chromatin fragments, which leads to whole genomic shock and local chromosomal shock that induces genome-wide genetic variation including indel (insertion and deletion) and nucleotide substitution. Nucleotide substitution causes synonymous codon usage bias (SCUB), an indicator of genomic mutation and natural selection. However, how asymmetric somatic hybridization affects SCUB has not been addressed. Here, we explored this issue by comparing expressed sequence tags of a common wheat cultivar and its asymmetric somatic hybrid line. Asymmetric somatic hybridization affected SCUB and promoted the bias to A- and T-ending synonymous codon (SCs). SCUB frequencies in chromosomes introgressed with exogenous fragments were comparable to those in chromosomes without exogenous fragments, showing that exogenous fragments had no local chromosomal effect. Asymmetric somatic hybridization affected SCUB frequencies in indel-flanking sequences more strongly than in non-flanking sequences, and this stronger effect was present in both chromosomes with and without exogenous fragments. DNA methylation-driven SCUB shift was more pronounced than other SC pairs. SCUB shift was similar among seven groups of allelic chromosomes as well as three sub-genomes. Our work demonstrates that the SCUB shift induced by asymmetric somatic hybridization is attributed to the whole genomic shock, and DNA methylation is a putative force of SCUB shift during asymmetric somatic hybridization. Asymmetric somatic hybridization provides an available method for deepening the nature of SCUB shift and genetic variation induced by genomic shock.

Keywords: introgression line, genomic shock, synonymous codon usage bias, somatic hybridization, wheat

OPEN ACCESS

Edited by:
Awais Rasheed,
Quaid-i-Azam University, Pakistan

Reviewed by:
Ratna Prabha,
Indian Council of Agricultural Research, India
Chenggen Chu,
Edward T. Schafer Agricultural Research Center (USDA-ARS), United States
Pengtao Ma,
Yantai University, China

***Correspondence:**
Mengcheng Wang
wangmc@sdu.edu.cn

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *18 March 2021*
Accepted: *07 May 2021*
Published: *11 June 2021*

Citation:
Xu W, Li Y, Li Y, Liu C, Wang Y, Xia G and Wang M (2021) Asymmetric Somatic Hybridization Affects Synonymous Codon Usage Bias in Wheat. Front. Genet. 12:682324. doi: 10.3389/fgene.2021.682324

INTRODUCTION

During domestication and improvement processes, the genetic base and diversity of crops gradually become low. The wild relatives of crops retain genetic diversity, and therefore are a valuable genetic resource for crop breeding. The genetic materials of wild relatives can be introduced intro crops *via* remote sexual hybridization and somatic hybridization (Tanksley and McCouch, 1997; Zamir, 2001). Somatic hybridization comprises symmetric and asymmetric somatic hybridization (Xia, 2009). Symmetric somatic hybridization is that protoplasts of two species are fused to form

fused cells, and fused cells regenerate into seedlings. Asymmetric somatic hybridization is that protoplasts of one species (donor) are irradiated by UV to fragment the genome prior to fusion and then irradiated protoplasts are fused with unirradiated protoplasts (recipient) (Xia, 2009). During asymmetric somatic hybridization, very small amounts of chromatin fragments of the donor are introgressed into the recipient genome (Wang et al., 2004; Cui et al., 2015); thus, asymmetric somatic hybrid is a special allopolyploidy. The transient co-emergence of donor and recipient genomes in fused cell as well as the introgression of exogenous chromatin segments into the genome lead to a strong genomic shock. Similar to diploydization of polyploydies (McClintock, 1984; Chen, 2007), the genomic shock can induce genome-wide genetic variation [mainly insertion and deletion (indel) and single-nucleotide substitution] in somatic hybrids (Liu and Xia, 2014).

Single-nucleotide substitution is a major force of evolution. Single-nucleotide substitution in protein-coding sequences produces either a synonymous codon (SC) or a non-synonymous codon. Except for methionine and tryptophan, all amino acids are encoded by at least two SCs. SCs of an amino acid appear to have a different frequency in the genome, which is referred to as synonymous codon usage bias (SCUB). Nucleotide substitution between SCs is generally considered to be functionally neutral (King and Jukes, 1969; Nei and Gojobori, 1986). In fact, SCs affect recombination rates, splicing regulation, transcription efficiency, RNA secondary structure, mRNA stability, translational efficiency, and accuracy in the regulation of gene expression, protein folding, and so on (Marais et al., 2001; Warnecke and Hurst, 2007; Zhang et al., 2009; Tuller et al., 2010; Presnyak et al., 2015). Thus, SCUB has proved to reflect the mutation, genetic drift, and natural selection (Akashi and Eyre-Walker, 1998; Akashi, 2001; Guo and Yuan, 2009; Wang Z. et al., 2014) and to be closely associated with plant evolution (Qin et al., 2013; Qi et al., 2015; Xu et al., 2015). The induction of genome-scale genetic variation induced by asymmetric somatic hybridization is under selection pressure in wheat (Wang et al., 2018). Given that the nucleotide substitution between SCs suffers from less selection pressure than the substitution between non-SCs, the interesting question is that whether asymmetric somatic hybridization affects SCUB in the recipient genome.

Unlike allopolyploidy that the genomes of progenitors stably co-emerge in the nucleus, the recipient genome is introgressed several donor chromatin fragments in asymmetric somatic hybrids. In wheat, the introgressed exogenous fragments results in both whole genome shock and local chromosomal shock, which induces genome-scale genetic variation and nucleotide substitution in indel-flanking sequence, respectively, in a non-random manner (Wang et al., 2018). This indicates that the introgression of exogenous fragments has an effect on genetic variation in chromosomes possessing chromatin fragments. However, whether the introgressed fragments influence SCUB through local chromosomal shock has not been addressed.

We previously bred a common wheat cultivar Shanrong 3 (SR3) *via* asymmetric somatic hybridization with the wheat cultivar Jinan 177 (JN177) as the recipient and tall wheatgrass (*Thinopyrum elongatum*, wheat's close relative) as the donor (Xia et al., 2003). SR3 genome possesses six exogenous fragments (Wang et al., 2005; Liu et al., 2015) and occurs in genome-scale genetic variation including nucleotide substitution and insertion and deletion (indel) (Feng et al., 2004; Liu et al., 2007, 2009, 2015; Wang et al., 2015, 2018), which contributes to salt tolerance of SR3. Thus, it is necessary to uncover whether this genetic variation was along with SCUB alteration, which will provide evidence for further understanding the characteristics of genetic variation induced by asymmetric somatic hybridization, and how genetic variation accounts for the salt tolerance of SR3 as well as the change in traits of introgression lines. Here, we used the unigenes of SR3 and JN177 (Wang et al., 2015) and found that asymmetric somatic hybridization caused SCUB shift, and introgressed fragments had no stronger effect on SCUB in local chromosome.

MATERIALS AND METHODS

The Introduction of Wheat Introgression Cultivar Shanrong 3

The common bread wheat cultivars Jinan 177 (JN177) and Shanrong 3 (SR3) were used for analysis (Wang et al., 2018), both of which were bred by our lab (Dr. Guangmin Xia) and have not been deposited in a publicly available herbarium. The detail of JN177 and SR3 was introduced in our previous study (Xia et al., 2003). Briefly, SR3 is a wheat cultivar with high salt and drought tolerance bred *via* asymmetric hybridization with the common wheat cultivar JN177 (modest salt and drought tolerance) as the recipient and wheat's close relative tall wheatgrass (*T. elongatum*, topmost salt tolerance) as the donor, and SR3 genome is introgressed with several chromatin fragments of the donor (Xia et al., 2003). The genome of SR3 took place with genome-wide genetic and epigenetic variation (Liu et al., 2015; Wang et al., 2015, 2018).

EST Sequencing, Chromosomal Localization, and Genetic Variation Analysis

The methods of expressed sequence tag (EST) sequencing, chromosomal localization, and genetic variation analysis were present in the previous study. Briefly, the ESTs of JN177 and SR3 were used for constructing cDNA library, which were used for large-scale EST sequencing with the Sanger sequencing method. After sequence cleaning, highly qualified EST sequences (>100 nt) were assembled to produce unigenes, and PCR was conducted to confirm the quality of unigene assembly. The chromosomal localization was determined by BLASTing the unigenes against wheat survey database (http://wheat-urgi.versailles.inra.fr/Seq-Repository). The local BLASTN was carried out to analyze the genetic variation between the unigenes of SR3 and JN177. The local BLASTX against the non-redundant protein database was performed to extract CDS of unigenes.

SCUB Frequency Calculation

The frequency of a SC was calculated by the ratio of the amount of this codon to the amount of 59 SCs encoding 18 amino acids except for three stop codons TAA, TAG and TGA, ATG (methionine) and TGG (tryptophan) in CDS of all unigenes according to our previous study [8]. The SCUB frequency of a given amino acid was defined as the amount of SCs with C and G at the third position to the amount of SCs with A and T at the third position in CDS of all unigenes. For instance, alanine is coded by GCA, GCT, GCC, and GCG, so the SCUB frequency of alanine was defined as the ratio of GCC and GCG amount to GCA and GCT amount. Total SCUB frequency was defined as the ratio of the amount of all SCs with A, T, C, or G at the third position (NNA, NNT, NNC, or NNG) to the amount of all SCs in CDS of all unigenes.

DNA methylation is a major source of DNA variation in the nuclear genome, given that methylated cytosine (5 mC) is readily converted into thymine [21]. Methylation is mainly present in C of CpG, and the conversion of 5 mC produces TpG in sense strand and CpA in antisense strand. Given the lower selection pressure on the third position of codons, the conversion of NCG to NCA (the second-third position) as well as NC|G to NT|G (the third-next codon's first position) would be dominant, which leads to the bias to A- and T-ending codons [6]. Thus, the ratios of NXA/NXG (X = A, T, C, or G) can reflect the effect of the second nucleotide on the conversion from G to A at the third positon, and the ratios of NT|X/NG|X (X = A, T, C, or G at the first position of the next codon) can reflect the first nucleotide of the next codon on the conversion from C to T at the third position. If DNA methylation contributes to SCUB, the ratio of NCA/NCG would be higher than those of NXA/NXG (X = A, G, T), and the ratio of NT|G/NC|G would be higher than those of NT|X/NG|X (X = A, T, C). Based on this, the difference between the ratios of NCA/NCG and NAA/NAG, NGA/NGG, and NTA/NTG as well as the difference between the ratios of NT|G/NC|G and NT|A/NC|A, NT|C/NC| NT|G/NC|G, and NT|T/NG|T were calculated to assess the potential association between DNA methylation and SCUB.

The association between DNA methylation and SCUB was also analyzed using the ratio of NXA/NXG of amino acids (Ala, Arg, Gln, Glu, Gly, Leu, Lys, Pro, Ser, Thr, and Val) that are encoded by SCs including A and G at the third position. The ratio of NXA/NXG ratio of SR3 to NXA/NXG ratio of JN177 was used to access the putative contribution of DNA methylation to the change in SCUB induced by asymmetric somatic hybridization.

Indel-Flanking and Indel-Remote Sequence Extraction

Fifteen codons (45 nt) of 5′- and 3′-flanking sequences of indels (insertions and deletions) were extracted for calculating SCUB frequency. The indels with instance >45 nt from start and stop codons were used for analysis to avoid the terminal effect. Sequences with length <90 nt between two indels were not considered as flanking sequences to avoid the effect of adjacent indels. The sequences with distance >45 nt from indels were extracted as non-flanking sequences.

Statistical Analysis

The chi square (χ^2) test of the cross-table analysis was performed to establish the significance of differences in the SCUB frequency based on SCs in CDS sequences of all unigenes, mapped unigenes, introgressed unigenes, and non-introgressed unigenes between SR3 and JN177. The difference in SCUB frequencies of each of the 18 amino acids between SR3 and JN177 was performed using the amount of A- and T-ending SCs and C- and G-ending SCs of a given amino acid. The difference in total SCUB frequencies between SR3 and JN177 was performed using the amount of all SCs with A, T, C, or G at the third position (NNAs, NNTs, NNCs, or NNGs). The significance of differences in SCUB frequency related to the third nucleotide position concerning DNA methylation between SR3 and JN177 was also analyzed with the χ^2 test of the cross-table analysis. For example, the difference in NCA/NCG ratio (the second–third nucleotide combination) between SR3 and JN177 was analyzed by the amounts of NCA and NCG of SR3 and JN177; the difference in NT|G/NC|G ratio (the third nucleotide and the first nucleotide of next codon combination) between SR3 and JN177 was analyzed by the amounts of NT|G and NC|G of SR3 and JN177. Besides, the difference of SCUB frequencies in other comparisons such as those between introgressed and non-introgressed sequences as well as those between indel 5′-flanking and 3′-flanking sequence were also conducted using the χ^2 test of the cross-table analysis. The difference in NXA/NXG SCs of an amino acid encoding by A- and G-ending SCs (Ala, Arg, Gln, Glu, Gly, Leu, Lys, Pro, Ser, Thr, and Val) between SR3 and JN177 was calculated with the χ^2 test of the cross-table analysis, and the amounts of NXA and NXG of SR3 and JN177 were used for calculation. The difference in the ratios of NCG/NCA of Ala, Pro, Ser, and Thr between SR3 and JN177 from the ratios of N(G/T)G/N(G/T)A of Arg, Gln, Glu, Gly, Leu, Lys, and Val between SR3 and JN177 was calculated with the *t*-test. The fluctuation was assessed by the coefficient of variation (CV), which is calculated as the ratio of standard deviation to mean. The consistency of SCUB frequency was detected *via* reliability analysis (model was set as alpha), and Cronbach's alpha value was used to indicate the consistency.

RESULTS

SCUB Appeared to Be Different Between Shanrong 3 and Jinan 177

We previously large-scale sequenced the cDNA libraries and obtained 9,634 and 7,107 unigenes from Shanrong 3 (SR3) and Jinan 177 (JN177), respectively (Wang et al., 2015). SR3 genome took place at a high frequency of genetic variation such as nucleotide substitution (Liu et al., 2015; Wang et al., 2015, 2018); therefore, we analyzed here to know the SCUB alteration. Approximately 77.1% of SR3 and 79.2% of JN177 in length were coding sequences (CDS) (**Supplementary Figure 1A**), which included ∼1.4 × 10^6 and 1.3 × 10^6 codons (**Supplementary Figure 1B**). Among the 59 SCs that encode 18 amino acids, the frequencies of both codons with A/T at the third position (NNAs/NNTs) and NNCs/NNGs were various, and NNCs and NNGs were more dominant than NNAs

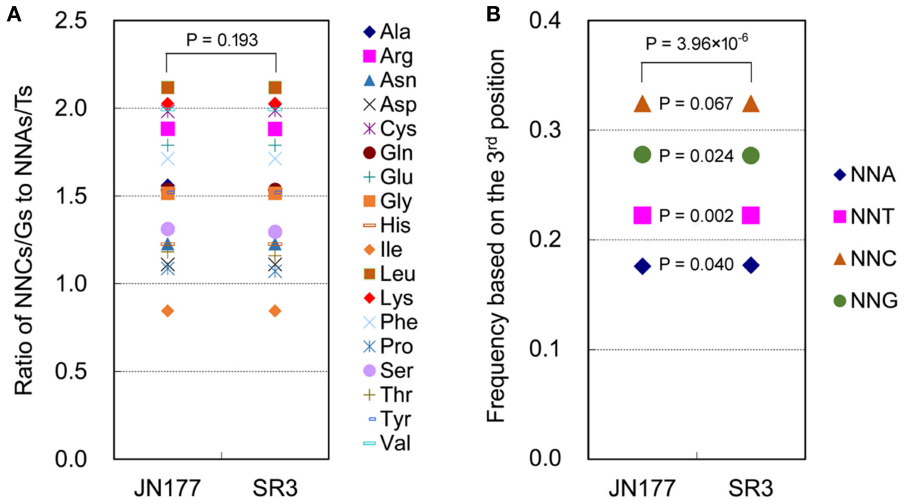


FIGURE 1 | SCUB was shifted in the genome of SR3. **(A)** The ratios of the frequencies of NNC/NNG SCs to NNA/NNT SCs. NNCs/Gs: SCs with C or G as their final bases; NNAs/Ts: SCs with A or T as their final base. **(B)** The total frequency of NNA, NNT, NNC, and NNG codons. The total frequency was calculated as the ratio between the number of all SCs ending with A, T, C, or G and the amount of all SCs. The statistical comparison was conducted with chi square (χ^2) test of cross-table analysis.

and NNTs; the frequency of each of the codons was comparable among JN177 and SR3 (**Supplementary Figure 2**).

SCUB frequency of a given amino acid encoded by SCs was calculated as the ratio of NNCs/NNGs amount to NNAs/Ts amount (**Figure 1A**). The SCUB frequencies of 18 amino acids were 0.844 to 2.118 in JN177 and SR3, and they were different among each other [coefficient of variation (CV) = 0.069–0.501] (**Supplementary Table 1**). The SCUB frequencies of amino acids was almost comparable between two cultivars (**Figure 1A**), except that the frequencies of Ala, Pro, Ser, and Thr were different between JN177 and SR3 ($P < 0.05$, χ^2 test) (**Supplementary Table 1**).

Total SCUB frequencies of four types of codons with A, T, C, and G at the third positon (NNA, NNT, NNC, and NNG), calculated as the ratios of the amounts of four types of codons to the amount of all codons, were used to compare SCUB between JN177 and SR3. NNC and NNG were more pronounced than NNA and NNT (**Figure 1B**). Generally, the frequencies of NNA, NNT, NNC, and NNG showed a significant difference between JN177 and SR3 ($P = 3.96 \times 10^{-6}$, χ^2 test). In comparison with JN177, SR3 had higher frequencies of NNA and NNT ($P = 0.040$ and 0.002), but lower frequencies of NNC and NNG ($P = 0.024$ and 0.067). The results indicate that asymmetric somatic hybridization affected SCUB.

Introgression of Exogenous Fragments had No Genome-Wide Effect on SCUB

In the genome of SR3, six exogenous fragments were found to be introgressed in chromosomes 1BL, 1DL, 2AL, 2DL, 5BS, and 6DS (Wang et al., 2004). We previously found that genetic variation in these chromosomes had no significant difference to that in the other chromosomes without introgressed fragments (Wang et al., 2018). To know the effect of exogenous fragments on SCUB in introgressed chromosomes, the SCUB frequencies of unigenes mapped to 21 chromosomes were compared. In general, the SCUB frequencies of the amino acids encoded by SCs were similar between JN177 and SR3 when mapped (unigenes mapped to chromosomes), introgressed (unigenes mapped to chromosomes introgressed with exogenous fragments), or non-introgressed (unigenes mapped to chromosomes without exogenous fragments) unigenes were used for calculation, respectively ($P > 0.05$, χ^2 test), except for Ala and Pro of mapped unigenes ($P = 0.023$ and 0.043, χ^2 test) (**Figure 2A**; **Supplementary Table 2**). However, the total SCUB frequencies of either mapped or non-introgressed unigenes were significantly different between JN177 and SR3 ($P = 3.70 \times 10^{-5}$ and 3.09×10^{-4}, χ^2 test), but the difference was weakened in introgressed unigenes ($P = 0.196$, χ^2 test) (**Figure 2B**). Moreover, each of the NNA, NNT, NNC, and NNG frequencies were obviously different between SR3 and JN177 when total, mapped, and non-introgressed unigenes were calculated, and SR3 had higher NNA and NNT ratios but lower NNC and NNG ratios. Although the ratio values were similar to non-introgressed unigenes, the difference in each of the NNA, NNT, NNC, and NNG ratios of introgressed unigenes between JN177 and SR3 were statistically not obvious ($P = 0.281$–0.652, χ^2 test) (the reason was that the codon amounts of introgressed unigenes were drastically fewer than those of non-introgressed unigenes). The findings showed that SCUB in chromosomes introgressed with and without exogenous fragments was both shifted.

The SCUB frequencies of the amino acids encoded by SCs were generally comparable among total unigenes, mapped unigenes, non-introgressed unigenes, and introgressed unigenes in either JN177 (CV = 6.94×10^{-5} to 0.0064, Cronbach's alpha = 1.00) and SR3 (CV = 1.79×10^{-4} to 0.0059, Cronbach's alpha = 1.00) (**Figure 2A**; **Supplementary Table 3**). The SCUB

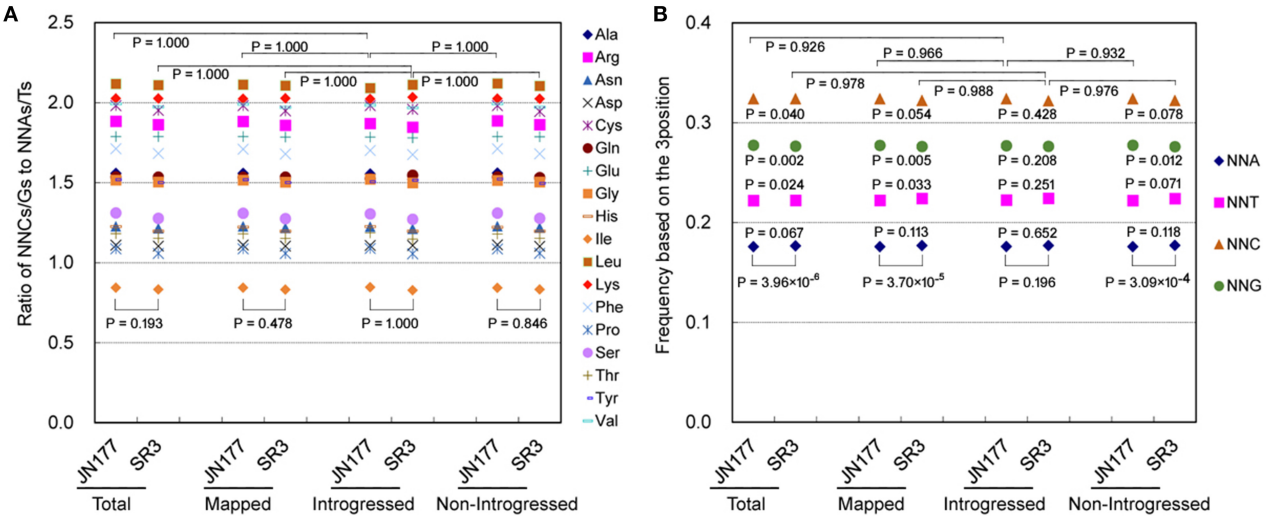


FIGURE 2 | SCUB was shifted to a similar extent in chromosomes introgressed with and without exogenous fragments. **(A)** The ratios of the frequencies of NNC/NNG SCs to NNA/NNT SCs. NNCs/Gs: SCs with C or G as their final bases; NNAs/Ts: SCs with A or T as their final base. **(B)** The total frequency of NNA, NNT, NNC, and NNG codons. The total frequency was calculated as the ratio between the number of all SCs ending with A, T, C, or G and the amount of all SCs. Total: all unigenes; Mapped: unigenes mapped to chromosomes; Introgressed: unigenes from chromosomes introgressed with exogenous fragments; Non-introgresed: unigenes from chromosomes without exogenous fragments. The statistical comparison was conducted with chi square (χ^2) test of fourfold cross-table analysis.

frequencies of these amino acids in introgressed unigenes had no significant difference from those in non-introgressed unigenes as well as mapped unigenes and total unigenes in either JN177 or SR3 (**Figure 2A**; **Supplementary Table 2**). Moreover, the total SCUB frequencies of introgressed unigenes, illustrated by the ratios of NNA, NNT, NNC, and NNG, were also similar to those of non-introgressed unigenes, mapped unigenes, and total unigenes in JN177 (P = 0.926–0.966) and SR3 (P = 0.976–0.988) (**Figure 2B**). These results indicated that SCUB in chromosomes introgressed with exogenous fragments was not affected more strongly than SCUB in chromosomes without exogenous fragments.

SCUB Was Promoted in Indel-Flanking Sequences

We previously found that nucleotide substitution preferred sequences adjacent to indels in SR3 genome (Wang et al., 2018). To clarify whether SCUB had a similar effect, the CDS sequences aligned between JN177 and SR3 were extracted, and sequences flanking to indels between JN177 and SR3 as well as sequences not flanking to indels were used for calculating SCUB frequency. The SCUB frequencies of amino acids encoded by SCs were also similar between introgressed sequences and non-introgressed sequences as well as aligned and mapped sequences (P = 1.000, χ^2 test) (**Supplementary Figure 3A**; **Supplementary Table 4**); the total SCUB frequencies were also comparable in these comparison combinations (P = 0.123–0.484, χ^2 test) (**Supplementary Figure 3B**).

The total SCUB frequencies of flanking and non-flanking (remote) sequences of indels were calculated. For mapped sequences, the total SCUB frequencies showed a significant difference between JN177 and SR3 when the whole aligned sequences were compared (P = 0.003, χ^2 test) (**Figure 3A**). Consistently, the total SCUB frequencies of indel-flanking sequences (two-sides) and non-flanking sequences of SR3 were also obviously different from those of JN177 (P = 0.030 and 0.041, χ^2 test). In comparison with JN177, SR3 had slightly higher frequencies of NNA and NNT but lower frequencies of NNC and NNG in whole and non-flanking sequences, while an increase of NNA and NNT frequencies and a decrease of NNC and NNG frequencies were stronger in SR3 (**Figure 3A**; **Supplementary Table 5**). To further evaluate the effect of indels on SCUB, we calculate the ratio of NNC/G to NNA/T (**Figure 3D**). When compared with JN177, SR3 had significantly lower ratios in whole, indel flanking (5′-, 3′-, and two sides), and non-flanking sequences (P = 0.0003–0.066, χ^2 test). On the other hand, for mapped sequences, the total SCUB frequencies were comparable among whole, indel flanking, and non-flanking sequences in JN177 (**Figure 3A**). In SR3, the total SCUB frequencies of whole sequences and indel non-flanking sequences were comparable (P = 0.927, χ^2 test), and they were both higher than that of indel-flanking sequences (P = 0.091 and 0.045, χ^2 test); two sides of indel-flanking sequences had similar total SCUB frequency with either 5′- or 3′-side. Consistently, the ratio of NNC/G to NNA/T of indel-flanking sequences was lower than those of whole and non-flanking sequences in SR3 (P = 0.012 and 0005, χ^2 test), as was not found in JN177 (P = 0.775 and 0.860, χ^2 test) (**Figure 3D**). These data were also found in introgressed sequences and non-introgressed sequences (**Figures 3B,C,E,F**). Note that, in introgressed sequences, although difference values of the ratios were similar to those of mapped and non-introgressed sequences, some P-values of

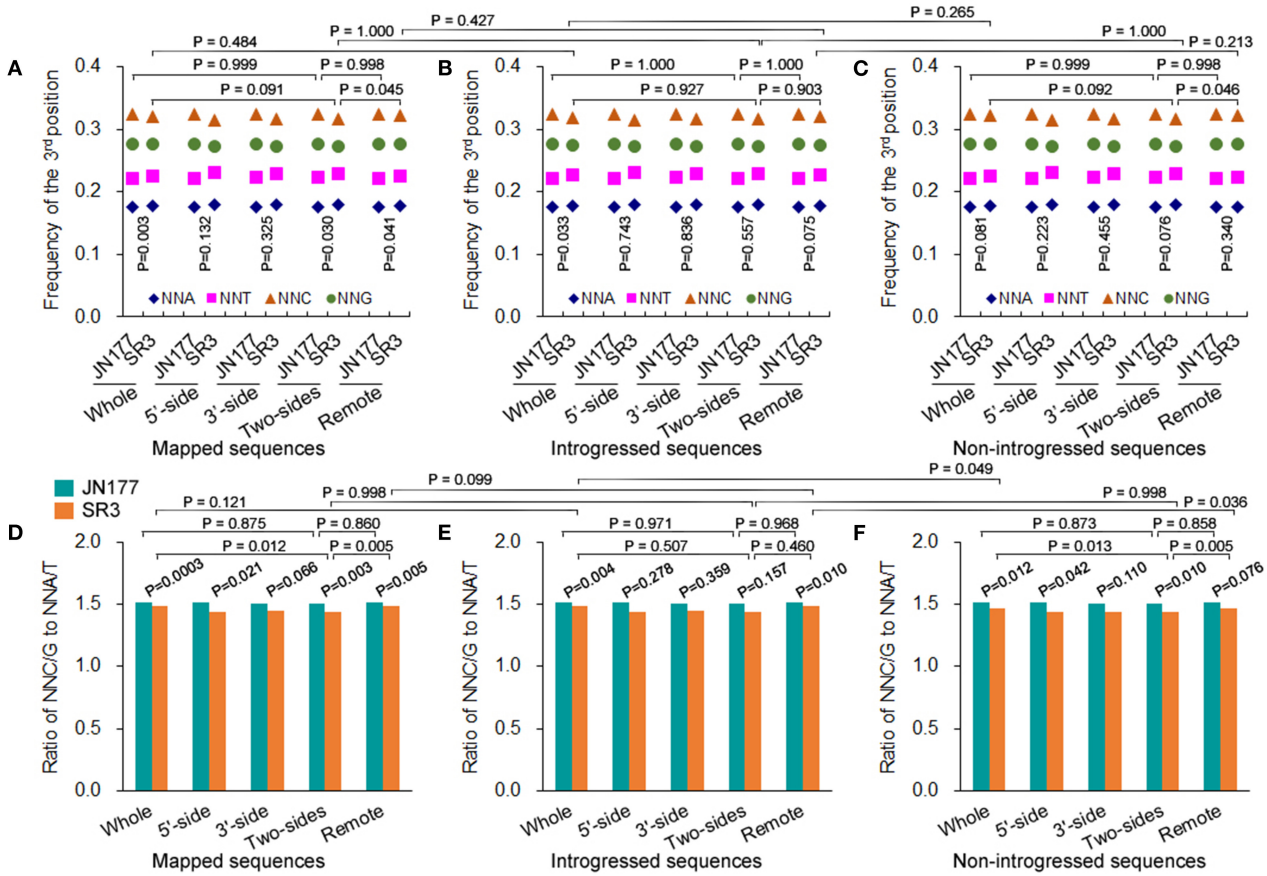


FIGURE 3 | SCUB shift was promoted close to indels. **(A–C)** The total frequencies of NNA, NNT, NNC, and NNG codons in indel-flanking and non-flanking sequences. **(D–F)** The ratio of NNC and NNG frequency to NNA and NNT frequency. Whole: the whole part of sequences; 5′-side: 5′-side sequences flanking to indels; 3′-side: 3′-side sequences flanking to indels; Two-sides: 5′- and 3′-side sequences flanking to indels; Remote: the sequences non-flanking to indels. The statistical comparison was conducted with chi square (χ^2) test of fourfold cross-table analysis.

the comparisons were more than 0.05 (**Figures 3B,E**), which resulted from the SC amount of indel-flanking sequences being drastically fewer than whole and non-flanking sequences. These results indicate that SCUB shift was promoted in indel-flanking sequences.

The total SCUB frequencies of whole sequences were comparable between introgressed sequences and non-introgressed sequences/mapped sequences in SR3 ($P = 0.265$ and 0.484, χ^2 test) (**Figures 3A–C**). Similarly, the total SCUB frequencies of both indel-flanking sequences and non-flanking sequences were almost the same in these comparisons. The ratios of NNC/G to NNA/T of indel-flanking, non-flanking, and whole sequences in introgressed sequences were also not different from those in mapped sequences of SR3 ($P = 0.121$, χ^2 test) (**Figures 3D,E**). In comparison between introgressed and non-introgressed sequences, the ratios of NNC/G to NNA/T of indel-flanking sequences were similar ($P = 0.998$, χ^2 test) but those of whole and non-flanking sequences appeared to be slightly different ($P = 0.049$ and 0.036, χ^2 test) (**Figures 3E,F**). These data show that the preference of SCUB in the sequences adjacent to indels was not improved in the chromosomes with exogenous fragments.

The Association Between DNA Methylation and SCUB

Asymmetric somatic hybridization induced genome-scale epigenetic variation (DNA methylation) in SR3 genome (Liu et al., 2015). DNA methylation mediated C-to-T conversion and partially accounts for genetic variation induced by asymmetric somatic hybridization (Wang et al., 2018); thus, we analyzed whether it is also associated with SCUB in SR3 genome. The conversion produces C to T and G to A in sense and antisense strands, respectively, so the conversions lead to the NCG-to-NCA shift (the second–third position) and the NC|G-to-NT|G shift (the third-next codon's first position).

In mapped, introgressed, and non-introgressed sequences, the NCA/NCG ratio (illustrating the conversion of C to T in antisense strand) was higher in SR3 than in JN177 ($P = 3.99 \times 10^{-9}$, χ^2 test), but the NGA/NGG, NAA/NAG, and NTA/NTG ratios were comparable between two cultivars ($P = 0.893$–0.990,

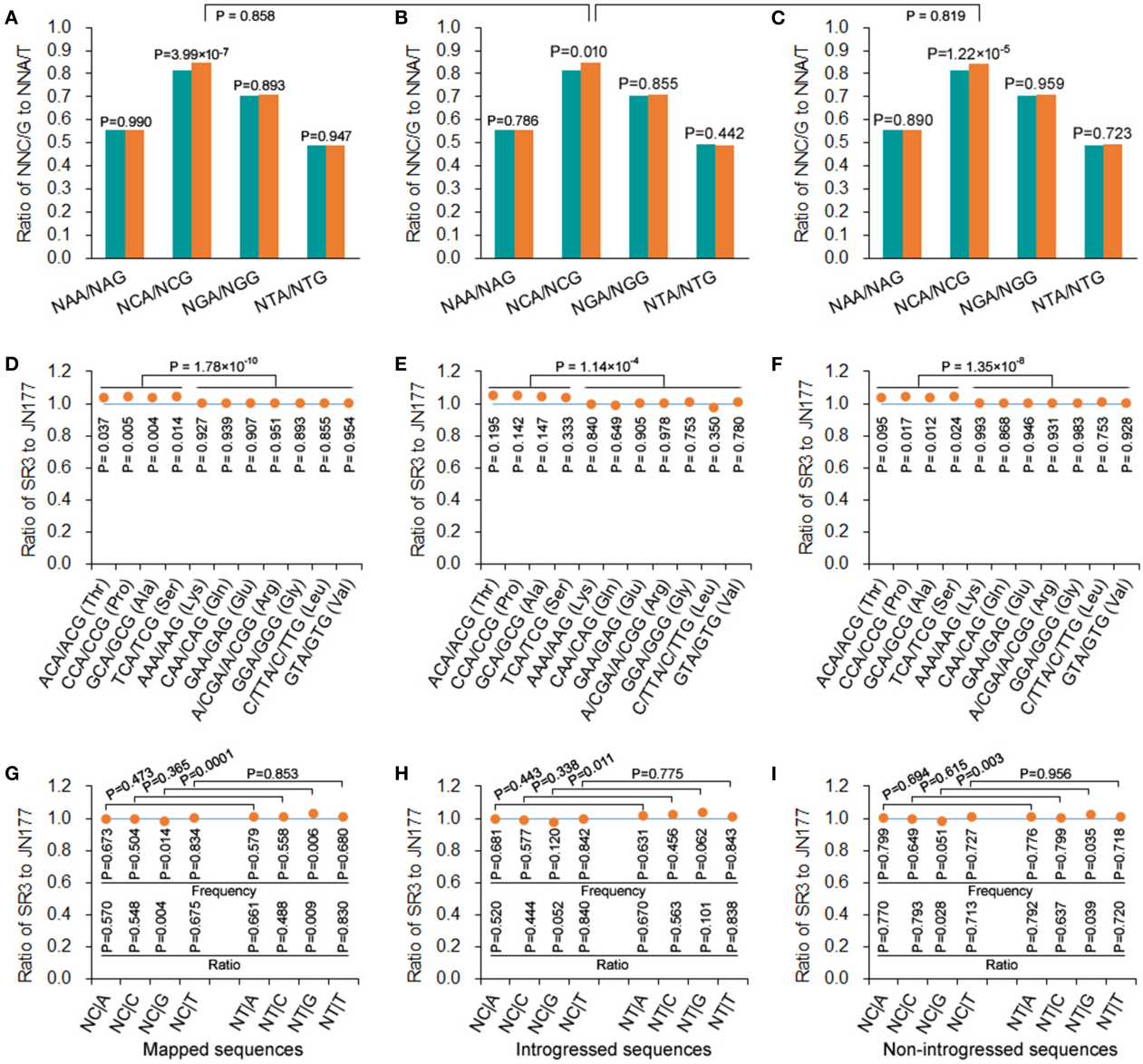


FIGURE 4 | DNA methylation-driven SCUB shift was more pronounced. **(A–C)** The ratio of NNA frequency to NNG frequency. **(D–F)** The ratio of the number of NNA to NNG codons encoding a given amino acid. **(G–I)** The ratio of the frequency of NC|N adjacent codons and NT|N adjacent codons between SR3 and JN177. In **(D)** to **(F)**, the difference in the ratios of Thr, Pro, Ala, and Ser from the ratios of other amino acids was calculated using *t*-test. The other statistical comparison was conducted with chi square (χ^2) test of fourfold cross-table analysis.

χ^2 test) (**Figure 4A**). On the other hand, we calculated the NT|G and NC|G frequencies to evaluate the effect of C-to-T conversion of sense strand. In mapped sequences, the NC|G frequency of SR3 was significantly lower than JN177's frequency ($P = 0.014$, χ^2 test), and the ratio of SR3's NC|G frequency to JN177's NC|G frequency was <1 ($P = 0.004$, χ^2 test); the frequencies of NA|G, NG|G, and NT|G were comparable between SR3 and JN177 ($P = 0.504$–0.834, χ^2 test), and the ratios of the frequencies were all around 1 ($P = 0.488$–0.830, χ^2 test) (**Figure 4G**). Oppositely, the NT|G frequency of SR3 was significantly higher than JN177's frequency ($P = 0.006$, χ^2 test), and the ratio of SR3's NC|G frequency to JN177's NC|G frequency was more than 1 ($P = 0.009$, χ^2 test); the frequencies of NA|G, NG|G, and NT|G were similar between SR3 and JN177 ($P = 0.504$–0.834, χ^2 test), and the ratios of the frequencies were also near to 1 ($P = 0.488$–0.830, χ^2 test) (**Figure 4G**). The NT|G/NC|G ratio exhibited a significant difference between JN177 and SR3 ($P = 0.0001$), but NT|A/NC|A, NT|C/NC|C, and NT|T /NC|T ratios were similar

($P = 0.365$–0.853). These results were repeated in introgressed and non-introgressed sequences (**Figures 4B,C,H,I**). Moreover, the NCA/NCG ratio in introgressed sequences was similar to non-introgressed and mapped sequences ($P = 0.819$ and 0.858, χ^2 test) (**Figures 4A–C**); the NT|G/NC|G ratios were also comparable between introgressed sequences and non-introgressed/mapped sequences (**Figures 4G–I**). These results indicated that DNA methylation affected SCUB in SR3, and the effect appeared to be similar in introgressed and non-introgressed sequences.

To further confirm this effect, A- and G-ending SC pairs encoding a given amino acid sharing the same nucleotides in their first and second positions were analyzed. In mapped sequences, the frequencies of NCA/NCG pairs (encoding alanine, proline, serine, and threonine) of SR3 were higher than those of JN177 ($P = 0.004$–0.037, χ^2 test), while the frequencies of N(A/G/T)A/N(A/G/T)G pairs (encoding Lys, Gln, Glu, Arg, Gly, Leu, and Val) were similar between SR3 and JN177 ($P = 0.855$–0.954, χ^2 test) (**Figure 4D**). The ratios of the frequencies of NCA/NCG of SR3 to JN177 were larger than 1, and significantly larger than those of N(A/G/T)A/N(A/G/T)G pairs (around 1) ($P = 1.78 \times 10^{-10}$, *t*-test). In introgressed and non-introgressed sequences, SR3 also had higher frequencies of NCA/NCG pairs but comparable frequencies of other pairs in comparison with JN177 (**Figures 4E,F**).

SCUB in Allelic Chromosomes and Sub-genomes

Allohexaploid wheat has seven groups of allelic chromosomes originating from A, B, and D sub-genomes. The genetic variation exhibited different frequencies and patterns among seven allelic chromosomes as well as three sub-genomes in SR3 (Wang et al., 2018), so we further analyzed the heterogeneity of their SCUB alteration. Here, the SC frequencies of 18 amino acids in each of seven allelic chromosomes totally had no statistical difference between JN177 and SR3 (**Figure 5A**), and the frequencies among seven allelic chromosomes were drastically consistent in both JN177 (CV = 0.001–0.009, Cronbach's alpha = 1.00) and SR3 (CV = 0.001–0.012, Cronbach's alpha = 1.00). The total SCUB frequencies of NNA, NNT, NNC, and NNG in each of seven allelic chromosomes were also similar between JN177, although the difference values were similar to those calculated using all unigenes (**Figure 5B**). On the other hand, the SC frequencies of 18 amino acids in each of the three sub-genomes were comparable between JN177 and SR3 (**Figure 5C**), and the consistency of the frequencies among three sub-genomes was also very strong in two cultivars (CV = 0.001–0.007 in JN177 and 0.0003–0.013 in SR3, Cronbach's alpha was 1.00). The total frequencies of NNA, NNT, NNC, and NNG in sub-genomes of SR3 were different from those of JN177 ($P = 0.040$–0.073) (**Figure 5D**). However, the difference in each of the NNA, NNT, NNC, and NNG frequencies between JN177 and SR3 appeared to be weak. The weaker difference was majorly due to the fewer amounts of SCs for statistical analysis.

DISCUSSION

Genetic variation serves as a driver of SCUB. Asymmetric somatic hybridization induces high frequency of genome-scale genetic variation such as nucleotide substitutions and indels (Feng et al., 2004; Liu et al., 2007, 2009, 2015; Wang et al., 2015, 2018). Here, we found that the SCUB frequencies exhibited difference between Shanrong 3 (SR3) and Jinan 177 (JN177) (**Figure 1**), showing that asymmetric somatic hybridization can affect SCUB. Moreover, genetic variation is suffered from selection pressure during plant evolution, and the coding sequence of the gene is under stronger selection pressure than untranslated regions (UTR) (Vinogradov, 2004). We previously found that genetic variation being stably reserved in SR3 is under selection pressure (Wang et al., 2018). SCUB is associated with a balance between mutation, genetic drift, and natural selection (Akashi, 2001; Adams and Wendel, 2005). Thus, the shift of SCUB in the genome of SR3 mirrors the selection pressure of genetic variation induced by asymmetric somatic hybridization.

Genomic shock has proved to induce genetic variation during natural evolution and dipolyploidization of polyploidies (McClintock, 1984; Chen, 2007). Introgression lines are specific polyploidies, but unlike polyploidies, chromosome rearrangement and large fragment deletion that are the force of genomic shock do not usually take place (Xia, 2009). However, our previous studies found that the genomic shock induced by the introgression of exogenous fragments led to the high frequency of genetic variation at the whole-genome scale rather than local chromosomal scale (Wang et al., 2018). Consistently, the shift of SCUB also occurred at the whole-genome scale, and the frequency in chromosome introgressed with and without exogenous fragments was similar (**Figure 2**), further confirming that the introgression of exogenous fragments predominantly leads to whole genomic shock so that high frequency of genetic variation is induced at the whole-genome scale. The possible causes are as follows: (1) end-joining of fragments as the mechanism of the introgression of donor chromatin segments usually results in point mutations and deletions during repair (Grundy et al., 2014); (2) besides visible fragments by GISH, small invisible exogenous fragments may contribute to the genomic shock.

There has a close association between indel and nucleotide substitution. Nucleotide substitution has higher frequency in flanking sequences of indel than non-flanking sequences (Tian et al., 2008), and substitution frequency appears to increases following the distance decrease to indels (Zhang et al., 2008; Guo et al., 2016). Thus, indel serves as a local "mutator" (Tian et al., 2008; Conrad et al., 2010; De and Babu, 2010; Hollister et al., 2010). In SR3 genome, indel-flanking sequences has higher nucleotide substitution frequency, and the frequency increases close to indels (Wang et al., 2018), showing that indel is also a local "mutator" under the whole genomic shock induced by the introgression of exogenous fragments. Here, SCUB frequency was also altered more strongly in indel-flanking sequences than non-flanking sequences in SR3 genome (**Figure 3**), consistent with the rule of indel as a local "mutator." Moreover, the introgression of exogenous fragments performs a local

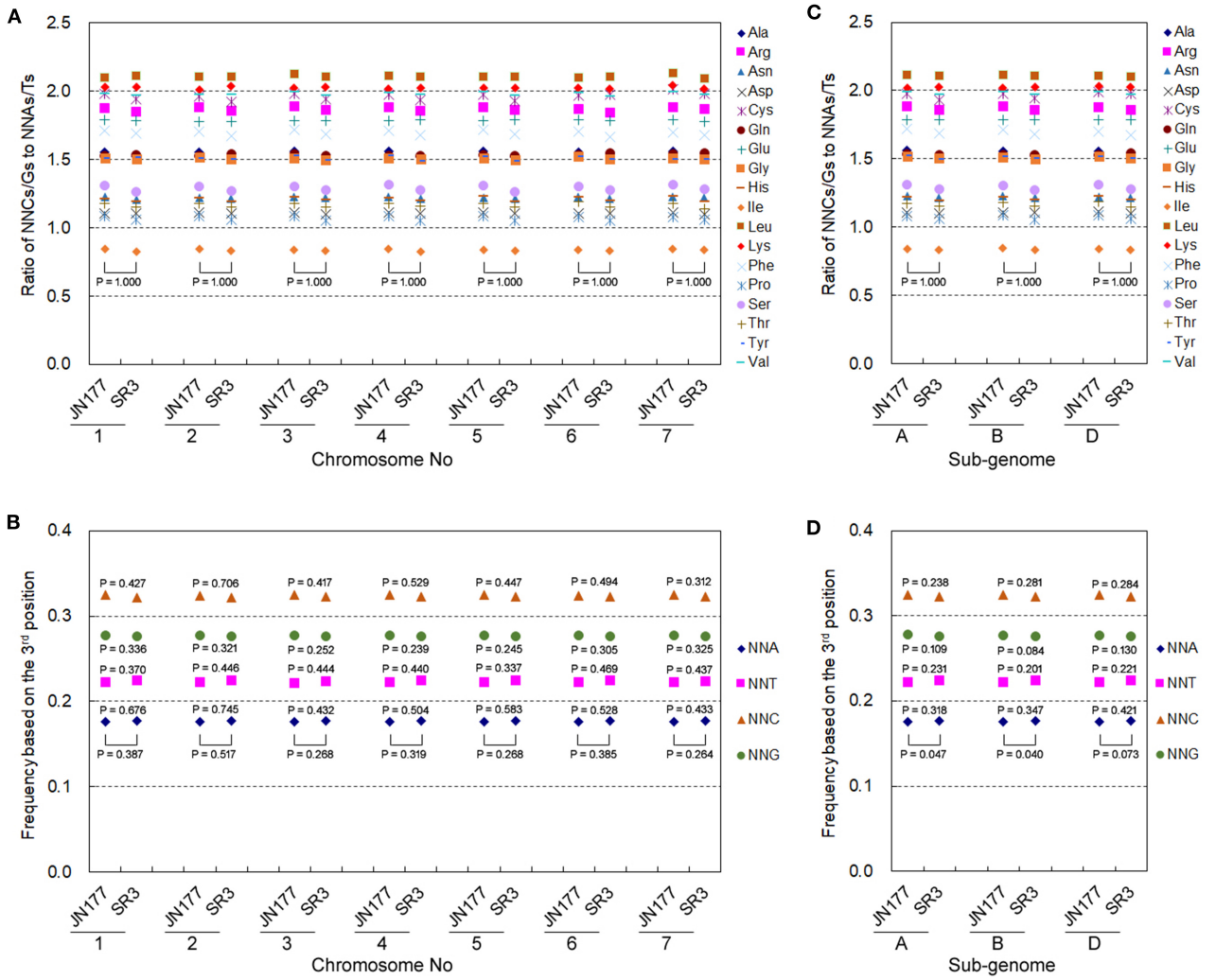


FIGURE 5 | The patterns of SCUB shift based on allelic chromosomes and sub-genomes. **(A,C)** The ratios of the frequencies of NNC/NNG SCs to NNA/NNT SCs in seven groups of allelic chromosomes **(A)** and three sub-genomes **(C)**. NNCs/Gs: SCs with C or G as their final bases; NNAs/Ts: SCs with A or T as their final base. **(B,D)** The total frequency of NNA, NNT, NNC, and NNG codons in seven groups of allelic chromosomes **(B)** and three sub-genomes **(D)**. The total frequency was calculated as the ratio between the amounts number of all SCs ending with A, T, C, or G and the amount of all SCs. The statistical comparison was conducted with chi square (χ^2) test of fourfold cross-table analysis.

chromosomal shock to promote higher nucleotide substitution frequency at 5′-side flanking sequences of indels in chromosomes with exogenous fragments than that in chromosomes without exogenous fragments (Wang et al., 2018). However, the SCUB frequency of two-side flanking sequences of indels was similar between chromosomes with and without exogenous fragments (**Figure 3**), indicating the difference in the patterns of nucleotide substitution and SCUB.

The alteration of cytosine methylation, a kind of variation induced by genomic shock, often occurs in the genomes of allopolyploidies (Comai, 2000; Shaked et al., 2001; Kashkush et al., 2002, 2003) and newly synthesized allohexaploid wheat (Shaked et al., 2001). Similarly, asymmetric somatic hybridization also alters cytosine methylation patterns in wheat (Wang M. et al., 2014; Liu et al., 2015). Methylated cytosines can be converted to thymine (Ossowski et al., 2010), and therefore is a major source of SNP formation (C→ T, and G→ A in complementary strand) (Laird, 2010). We previously proved that epigenetic modification-mediated nucleotide substitution is one of the major forces of genetic variation induced by asymmetric somatic hybridization in wheat (Wang et al., 2018). Here, DNA methylation was also closely associated with SCUB shift in SR3 genome (**Figure 4**). In line with the association of DNA methylation and SCUB differentiation during plant evolution (Qin et al., 2013; Qi et al., 2015; Xu et al., 2015), it could be concluded that epigenetic variation may play crucial roles in genetic variation induced by genomic shock, and there have been a close association between genetic and epigenetic variation.

As two major types of genetic variation, both nucleotide substitution and small indels suffer from selection pressure (McNally et al., 2006). In wheat, asymmetric somatic hybridization induces similar genetic variation among non-allelic chromosomes that suffer from similar selection pressure, but different genetic variation among three sub-genomes that suffer from different selection pressure, showing that the induction of genetic variation is a non-random but regulatory process (Wang et al., 2018). However, the SCUB frequencies in seven non-allelic as well as three sub-genomes chromosomes exhibited similar difference values of SCUB frequencies between SR3 and JN177, but the difference values were similar to those calculated using all unigenes (**Figure 5**), indicating the consistency and specificity between nucleotide substitution and SCUB induced by asymmetric somatic hybridization.

SR3 is a salt-tolerant cultivar with stronger salt tolerance capacity and higher yield than JN177 (Xia et al., 2003). Our previous work found that there was a remarkable difference in transcriptomic and proteomic profiles between these two cultivars under normal and saline conditions (Wang et al., 2008; Peng et al., 2009; Liu et al., 2012), which is partially attributed to the whole-genome scale genetic and epigenetic variation in SR3, including transposon activation, DNA methylation, as well as indels and nucleotide substitution in the promoters. For instance, the variation in promoter of a salt-tolerant associated gene *TaCHP* offers its transcription increase in SR3. Besides, large-scale EST comparison showed that a set of genes possessing genetic variation in CDS exhibits differential transcriptional profile between SR3 and Jn177. More importantly, the variation in CDS of some salt-tolerant associated genes alters the function of encoding products in salt response, of which the amino acid substitutions in TaSRO1, the putative salt-tolerant major QTL candidate gene, and TaSOD2 enhances their PARP and SOD activities, respectively. On the other hand, given the role of SCs in transcription efficiency, mRNA stability, translational efficiency, and accuracy and other aspects (Marais et al., 2001; Warnecke and Hurst, 2007; Zhang et al., 2009; Tuller et al., 2010; Presnyak et al., 2015), the SCUB shift may also account for the superior salt tolerance of SR3. Although there are no good or established tests for functional SCs so far, it is worthy of studying the biological performance in asymmetric somatic hybridization, natural evolution, and other events.

CONCLUSIONS

This work firstly addressed the patterns of asymmetric somatic hybridization-induced SCUB shift. Asymmetric somatic hybridization induced a whole-genome scale shift of SCUB, and introgressed exogenous fragments did not induce a stronger shift of SCUB in introgressed chromosomes, showing that SCUB was shifted *via* whole genomic shock rather than local chromosomal shock. Asymmetric somatic hybridization induced indels that promoted SCUB shift in flanking sequences. DNA methylation was a driver of SCUB shift, indicating the complicated association between genetic and epigenetic variation induced by asymmetric somatic hybridization.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: Sequences used for analysis in this work were submitted to Genbank (Accession number: JZ881292–JZ892704).

AUTHOR CONTRIBUTIONS

MW and GX designed the research. MW, WX, and CL performed the experiments. MW, YiL, YaL, and YW analyzed the data. MW and GX wrote the paper. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Key Research and Development Project (2016YFD0102003, MW), the National Natural Science Foundation of China (31870242, MW; 31800214, WX), the Shandong Province's Key Research and Development (2018GNC110010, MW), and the Natural Science Foundation of Shandong Province (ZR2019BC061, WX).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.682324/full#supplementary-material

Supplementary Figure 1 | The statistical of codons of unigenes. **(A)** The length of total unigenes as well as their CDS and UTR. **(B)** The codon amounts in CDS.

Supplementary Figure 2 | The frequencies of amino acid encoding codons.

Supplementary Figure 3 | SCUB frequencies in chromosomes introgressed with and without exogenous fragments using aligned sequences between JN177 and SR3. **(A)** The ratios of the frequencies of NNC/NNG SCs to NNA/NNT SCs. NNCs/Gs: SCs with C or G as their final bases; NNAs/Ts: SCs with A or T as their final base. **(B)** The total frequency of NNA, NNT, NNC, and NNG codons. The total frequency was calculated as the ratio between the amounts number of all SCs ending with A, T, C, or G and the amount of all SCs. Total: all unigenes; Mapped: unigenes mapped to chromosomes; Introgressed: unigenes from chromosomes introgressed with exogenous fragments; Non-introgresed: unigenes from chromosomes without exogenous fragments. The statistical comparison was conducted with chi square ($\chi 2$) test of fourfold cross-table analysis.

Supplementary Table 1 | The statistical analysis of SCUB frequencies of 18 amino acids.

Supplementary Table 2 | The statistical analysis of SCUB frequencies of 18 amino acids in chromosomes with and without exogenous fragments.

Supplementary Table 3 | The consistency of SCUB frequencies of 18 amino acids.

Supplementary Table 4 | The statistical analysis of SCUB frequencies of 18 amino acids in chromosomes with and without exogenous fragments based on aligned sequences.

Supplementary Table 5 | The statistical analysis of total SCUB frequencies between JN177 and SR3.

REFERENCES

Adams, K. L., and Wendel, J. F. (2005). Novel patterns of gene expression in polyploid plants. *Trends Genet*. 2005, 539–543. doi: 10.1016/j.tig.2005.07.009

Akashi, H. (2001). Gene expression and molecular evolution. *Curr. Opin. Genet. Dev*. 11, 660–666. doi: 10.1016/S0959-437X(00)00250-1

Akashi, H., and Eyre-Walker, A. (1998). Translational selection and molecular evolution. *Curr. Opin. Genet. Dev*. 8, 688–893. doi: 10.1016/S0959-437X(98)80038-5

Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Ann. Rev. Plant Biol*. 58, 377–406. doi: 10.1146/annurev.arplant.58.032806.103835

Comai, L. (2000). Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol. Biol*. 43, 387–399. doi: 10.1023/A:1006480722854

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., et al. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704–712. doi: 10.1038/nature08516

Cui, H., Sun, Y., Deng, J., Wang, M., and Xia, G. (2015). Chromosome elimination and introgression following somatic hybridization between bread wheat and other grass species. *Plant Cell Tissue Organ Cult*. 120, 203–210. doi: 10.1007/s11240-014-0594-1

De, S., and Babu, M. M. (2010). A time-invariant principle of genome evolution. *Proc. Natl. Acad. Sci. U.S.A*. 107, 13004–13009. doi: 10.1073/pnas.0914454107

Feng, D. S., Xia, G. M., Zhao, S. Y., and Chen, F. G. (2004). Two quality-associated HMW glutenin subunits in a somatic hybrid line between *Triticum aestivum* and *Agropyron elongatum*. *Theor. Appl. Genet*. 110, 136–144. doi: 10.1007/s00122-004-1810-x

Grundy, G. J., Moulding, H. A., Caldecott, K. W., and Rulten, S. L. (2014). One ring to bring them all–the role of Ku in mammalian non-homologous end joining. *DNA Rep*. 17, 30–38. doi: 10.1016/j.dnarep.2014.02.019

Guo, C., Du, J., Wang, L., Yang, S., Mauricio, R., Tian, D., et al. (2016). Insertions/deletions-associated nucleotide polymorphism in *Arabidopsis thaliana*. *Front. Plant Sci*. 7:1792. doi: 10.3389/fpls.2016.01792

Guo, F. B., and Yuan, J. B. (2009). Codon usages of genes on chromosome, and surprisingly, genes in plasmid are primarily affected by strand-specific mutational biases in *Lawsonia intracellularis*. *DNA Res*. 16, 91–104. doi: 10.1093/dnares/dsp001

Hollister, J. D., Ross-Ibarra, J., and Gaut, B. S. (2010). Indel-associated mutation rate varies with mating system in flowering plants. *Mol. Biol. Evol*. 27, 409–416. doi: 10.1093/molbev/msp249

Kashkush, K., Feldman, M., and Levy, A. A. (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics*. 160, 1651–1659. doi: 10.1093/genetics/160.4.1651

Kashkush, K., Feldman, M., and Levy, A. A. (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet*. 33, 102–106. doi: 10.1038/ng1063

King, J., and Jukes, T. (1969). Non-Darwinian evolution. *Science*. 165, 788–798. doi: 10.1126/science.164.3881.788

Laird, P. W. (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nat. Rev. Genet*. 11, 191–203. doi: 10.1038/nrg2732

Liu, C., Li, S., Wang, M., and Xia, G. (2012). A transcriptomic analysis reveals the nature of salinity tolerance of a wheat introgression line. *Plant Molecular Biology*. 78, 159–169. doi: 10.1007/s11103-011-9854-1

Liu, H., Liu, S., and Xia, G. (2009). Generation of high frequency of novel alleles of the high molecular weight glutenin in somatic hybridization between bread wheat and tall wheatgrass. *Theor. Appl. Genet*. 118, 1193–1198. doi: 10.1007/s00122-009-0973-x

Liu, S., Li, F., Kong, L., Sun, Y., Qin, L., Chen, S., et al. (2015). Genetic and Epigenetic Changes in Somatic Hybrid Introgression Lines Between Wheat and Tall Wheatgrass. *Genetics*. 199, 1035–1045. doi: 10.1534/genetics.114.174094

Liu, S., and Xia, G. (2014). The place of asymmetric somatic hybridization in wheat breeding. *Plant Cell Rep*. 33, 595–603. doi: 10.1007/s00299-013-1552-9

Liu, S., Zhao, S., Chen, F., and Xia, G. (2007). Generation of novel high quality HMW-GS genes in two introgression lines of *Triticum aestivum/Agropyron elongatum*. *BMC Evol. Biol*. 7:76. doi: 10.1186/1471-2148-7-76

Marais, G., Mouchiroud, D., and Duret, L. (2001). Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc. Natl. Acad. Sci. U.S.A*. 98, 5688–5692. doi: 10.1073/pnas.091427698

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260

McNally, K. L., Bruskiewich, R., Mackill, D., Buell, C. R., Leach, J. E., and Leung, H. (2006). Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. *Plant Physiol*. 141, 26–31. doi: 10.1104/pp.106.077313

Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol*. 3, 418–426.

Ossowski, S., Schneeberger, K., Lucas-Lledó J. I., Warthmann, N., Clark, R. M., Shaw, R. G., et al. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327, 92–94 doi: 10.1126/science.1180677

Peng, Z., Wang, M., Li, F., Lv, H., Li, C., and Xia, G. (2009). A proteomic study of the response to salinity and drought stress in an introgression strain of bread wheat. *Mol. Cell. Proteomics* 8, 2676–2686. doi: 10.1074/mcp.M900052-MCP200

Presnyak, V., Alhusaini, N., Chen, Y., Martin, S., Morris, N., Kline, N., et al. (2015). Codon optimality is a major determinant of mRNA stability. *Cell* 160, 1111–1124. doi: 10.1016/j.cell.2015.02.029

Qi, Y., Xu, W., Xing, T., Zhao, M., Li, N., Yan, L., et al. (2015). Synonymous codon usage bias in the plastid genome is unrelated to gene structure and shows evolutionary heterogeneity. *Evol. Bioinform*. 11, 65–77. doi: 10.4137/EBO.S22566

Qin, Z., Cai, Z., Xia, G., and Wang, M. (2013). Synonymous codon usage bias is correlative to intron number and shows disequilibrium among exons in plants. *BMC Genomics*. 14:56. doi: 10.1186/1471-2164-14-56

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., and Levy, A. A. (2001). Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell*. 13, 1749–1759. doi: 10.1105/TPC.010083

Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*. 277, 1063–1066. doi: 10.1126/science.277.5329.1063

Tian, D., Wang, Q., Zhang, P., Araki, H., Yang, S., Kreitman, M., et al. (2008). Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. *Nature* 455, 105–108. doi: 10.1038/nature07175

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., et al. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141, 344–354. doi: 10.1016/j.cell.2010.03.031

Vinogradov, A. E. (2004). Compactness of human housekeeping genes: selection for economy or genomic design? *Trends Genet*. 20, 248–253. doi: 10.1016/j.tig.2004.03.006

Wang, J., Xiang, F., Xia, G., and Chen, H. (2004). Transfer of small chromosome fragments of Agropyron elongatum to wheat chromosome via asymmetric somatic hybridization. *Sci. China Ser. C Life Sci*. 47, 434–441. doi: 10.1360/03yc0150

Wang, J., Xiang, F. N., and Xia, G. M. (2005). Agropyron elongatum chromatin localization on the wheat chromosomes in an introgression line. *Planta* 221, 277–286. doi: 10.1007/s00425-004-1443-y

Wang, M., Ji, Y., Feng, S., Liu, C., Xiao, Z., Wang, X., et al. (2018). The non-random patterns of genetic variation induced by asymmetric somatic hybridization in wheat. *BMC Plant Biol*. 18:244. doi: 10.1186/s12870-018-1474-3

Wang, M., Liu, C., Xing, T., Wang, Y., and Xia, G. (2015). Asymmetric somatic hybridization induces point mutations and indels in wheat. *BMC Genomics*. 16:807. doi: 10.1186/s12864-015-1974-6

Wang, M., Qin, L., Xie, C., Li, W., Yuan, J., Kong, L., et al. (2014). Induced and constitutive DNA methylation in a salinity-tolerant wheat introgression line. *Plant Cell Physiol*. 55, 1354–1365. doi: 10.1093/pcp/pcu059

Wang, M.-C., Peng, Z.-Y., Li, C.-L., Li, F., Liu, C., and Xia, G.-M. (2008). Proteomic analysis on a high salt tolerance introgression strain of *Triticum aestivum/Thinopyrum ponticum*. *Proteomics* 8, 1470–1489. doi: 10.1002/pmic.200700569

Wang, Z., Lucas, F., Qiu, P., and Liu, Y. (2014). Improving the sensitivity of sample clustering by leveraging gene co-expression networks in variable selection. *BMC Bioinform*. 15:153. doi: 10.1186/1471-2105-15-153

Warnecke, T., and Hurst, L. (2007). Evidence for a trade-off between translational efficiency and splicing regulation in determining synonymous

codon usage in *Drosophila melanogaster*. *Mol. Biol. Evol.* 24, 2755–2762. doi: 10.1093/molbev/msm210

Xia, G. (2009). Progress of chromosome engineering mediated by asymmetric somatic hybridization. *J. Genet. Genomics.* 36, 547–556. doi: 10.1016/S1673-8527(08)60146-0

Xia, G. M., Xiang, F. N., Zhou, A. F., Wang, H. A., and Chen, H. M. (2003). Asymmetric somatic hybridization between wheat (*Triticum aestivum* L.) and Agropyron elongatum (Host) Nevishi. *Theor. Appl. Genetics.* 107, 299–305. doi: 10.1007/s00122-003-1247-7

Xu, W., Xing, T., Zhao, M., Yin, X., Xia, G., and Wang, M. (2015). Synonymous codon usage bias in plant mitochondrial genes is associated with intron number and mirrors species evolution. *PLoS ONE.* 10:e0131508. doi: 10.1371/journal.pone.0131508

Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* 2, 983–989. doi: 10.1038/35103590

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* 16, 274–280. doi: 10.1038/nsmb.1554

Zhang, W., Sun, X., Yuan, H., Araki, H., Wang, J., and Tian, D. (2008). The pattern of insertion/deletion polymorphism in *Arabidopsis thaliana*. *Mol. Genet. Genomics.* 280, 351–361. doi: 10.1007/s00438-008-0370-1

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 09 July 2021
doi: 10.3389/fgene.2021.697294

Identification of Single Nucleotide Polymorphism in *TaSBEIII* and Development of KASP Marker Associated With Grain Weight in Wheat

*Ahsan Irshad[1], Huijun Guo[1], Shoaib Ur Rehman[2], Xueqing Wang[1], Jiayu Gu[1], Hongchun Xiong[1], Yongdun Xie[1], Linshu Zhao[1], Shirong Zhao[1], Chaojie Wang[1] and Luxiang Liu[1]**

[1] *Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, National Engineering Laboratory of Crop Molecular Breeding, National Center of Space Mutagenesis for Crop Improvement, Beijing, China,* [2] *Institute of Plant Breeding and Biotechnology, Muhammad Nawaz Sharif University of Agriculture, Multan, Pakistan*

OPEN ACCESS

Edited by:
Awais Rasheed, Quaid-i-Azam University, Pakistan

Reviewed by:
Pablo Federico Roncallo, National University of the South, Argentina
Giriraj Kumawat, ICAR Indian Institute of Soybean Research, India
Suyash Patil, International Rice Research Institute, IRRI, India

***Correspondence:**
Luxiang Liu
liuluxiang@caas.cn

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *19 April 2021*
Accepted: *08 June 2021*
Published: *09 July 2021*

Citation:
Irshad A, Guo H, Ur Rehman S, Wang X, Gu J, Xiong H, Xie Y, Zhao L, Zhao S, Wang C and Liu L (2021) Identification of Single Nucleotide Polymorphism in TaSBEIII and Development of KASP Marker Associated With Grain Weight in Wheat. Front. Genet. 12:697294. doi: 10.3389/fgene.2021.697294

Manipulation of genes involved in starch synthesis could significantly affect wheat grain weight and yield. The starch-branching enzyme (SBE) catalyzes the formation of branch points by cleaving the α-1,4 linkage in polyglucans and reattaching the chain *via* an α-1,6 linkage. Three types of SBE isoforms (SBEI, SBEII, and SBEIII) exist in higher plants, with the number of SBE isoforms being species-specific. In this study, the coding sequence of the wheat *TaSBEIII* gene was amplified. After the multiple sequence alignment of *TaSBEIII* genome from 20 accessions in a wheat diversity panel, one SNP was observed in *TaSBEIII-A*, which formed the allelic marker *allele-T*. Based on this SNP at 294 bp (C/T), a KASP molecular marker was developed to distinguish allelic variation among the wheat genotypes for thousand grain weight (TGW). The results were validated using 262 accessions of mini core collection (MCC) from China, 153 from Pakistan, 53 from CIMMYT, and 17 diploid and 18 tetraploid genotypes. Association analysis between *TaSBEIII-A* allelic variation and agronomic traits found that *TaSBEIII-A* was associated with TGW in mini core collection of China (MCC). The accessions possessing *Allele-T* had higher TGW than those possessing *Allele-C*; thus, *Allele-T* was a favorable allelic variation. By analyzing the frequency of the favorable allelic variation *Allele-T* in MCC, it increased from pre-1950 (25%) to the 1960s (45%) and increased continuously from 1960 to 1990 (80%). The results suggested that the KASP markers can be utilized in grain weight improvement, which ultimately improves wheat yield by marker-assisted selection in wheat breeding. The favorable allelic variation *allele-T* should be valuable in enhancing grain yield by improving the source and sink simultaneously. Furthermore, the newly developed KASP marker validated in different genetic backgrounds could be integrated into a breeding kit for screening high TGW wheat.

Keywords: wheat, association analysis, KASP, *TaSBEIII*, polymorphisms, molecular marker

INTRODUCTION

Wheat is the staple food for ∼33% of the population across the globe (Su et al., 2011). According to a worldwide survey, food demand will increase by 40% in the post-2020 era due to the rapid increase in global population (Su et al., 2011; Wang et al., 2019). To ensure this food security issue, there is a need to develop high-yield varieties through advanced molecular breeding (Wang et al., 2019). The valuable source in plants for energy and carbon is starch. The production of starch takes place in green leaves of plants during photosynthesis and surplus glucose is produced (Irshad et al., 2019). There is a relation of source and sink at the physiological level that helps to determine the yield of the crops (Rossi et al., 2015). Sink capacity is more important in wheat as compared to source so there is a need to explore the starch synthesis enzymes to accelerate the wheat yield through breeding (Hou et al., 2014). Starch mainly consists of amylose and amylopectin in which amylopectin contributed 75% in the starch granule (Pfister and Zeeman, 2016). The metabolism process of different enzymes [ADP-glucose pyrophosphorylases, starch synthases, starch-branching enzymes (SBEs), and starch-debranching enzymes] helps in the formation of starch (Bertoft, 2017).

Metabolism of starch significantly affects the yield and quality of wheat because it accounts for about 65–80% of the grain endosperm (Xia et al., 2020). Functional changes in starch genes dramatically influence the starch content, amylose content, and other agronomic traits. One of the most important yield contributing trait in wheat is the thousand grain weight (TGW). Different starch synthesis genes played a role in TGW in wheat such as decreasing the expression of *GWD via* RNAi and significantly increasing the grain number per plant and TGW (Ral et al., 2012). Similarly, three haplotypes were detected in *TaSSIV-A*, and these haplotypes were associated with TGW. *Hap-2-1A* showed a significant difference from the other two haplotypes and had higher TGW (Irshad et al., 2019). It can be suggested that optimizing starch metabolism might improve the TGW, and its value can be increased with significant genetic improvement in wheat grain yield (Zheng et al., 2011).

The main function of SBE is the formation of branch points through cleavage of the polyglucan chains from α-1,4 linkage and reattachment of these polyglucan chains *via* α-1,6 linkage. Based on physiological and biochemical properties, SBE has three types of isoforms (SBEI, SBEII, and SBEIII) (Yan et al., 2009). SBEIIa and SBEIIb are two further subclades of SBEII. SBEI is present in wheat, rice, maize, and other plants except for *Arabidopsis* (Kang et al., 2013). Different research work has been done on SBE proteins especially on SBEI, SBEIIa, and SBEIIb (Stamova et al., 2009; Jeon et al., 2010). There is limited information about SBEIII due to the challenging tasks in isolation and purification of the coded protein of this gene (Yan et al., 2009). In wheat, the *TaSBEIII* CDS sequence, which consists of 3780 bp with an open reading frame of 2748 bp, was identified through the RACE method from common wheat, which reveals the existence of the *SBEIII* gene in common wheat. SBEIII has special characteristics based on the predicted protein of 916 amino acids with four highly conserved domains (Kang et al., 2013). The *SBEIII* gene is reported in many higher plants but there remains no information in lower plants (Han et al., 2007). Based on this information, it is depicted that *SBEIII* is different from *SBEI*, and when higher plants were separated from lower plants, then this gene could have arisen during the evolutionary process of gene divergence and duplication. The expression of *TaSBEIII* was consistent during grain filling period in wheat, giving the idea that its function is different from other SBE enzymes and its main function is the formation of A and B granules in the grains of wheat (Kang et al., 2013).

To accelerate the process of the wheat breeding program, the potential approach is the marker-assisted selection (MAS) (Zheng et al., 2017). Single nucleotide change and deletion or insertion in a nucleotide sequence is referred to as single nucleotide polymorphism (SNP) (Liu et al., 2012). SNP markers are more reliable as compared to other markers (RFLP, RAPD, AFLP, SSR, and ISSR) due to their high relative stability and high-throughput scoring (Wang et al., 2015). Similarly, it has been reported that SNPs are effective markers in fine mapping, association analysis, and functional marker development (Zhang et al., 2013; Qi et al., 2014). By using the SNPs in association analysis, it has been documented that it is an effective tool for the identification of relationship between the polymorphic site of target gene and quantitative traits. These analyses have been widely used in many crops such as in *Arabidopsis*, maize, rice, and wheat (Hayashi et al., 2004; Li et al., 2010, 2016; Nemri et al., 2010; Irshad et al., 2019). Different methodologies have been used in the development of markers for different genes such as high-resolution melting (HRM) and cleaved amplified polymorphism (CAPS) (Chen et al., 2014; Luo et al., 2014). However, due to some limitations, the utilization of these markers is limited because the HRM method needs multiple PCR cycles for their unique PCR product and CAPS markers also need special enzymes for special pair digestion. Due to these reasons, the breeder needs a friendly and high-throughput marker methodology for diverse breeding programs. The utilization of the newly developed Kompetitive Allele Specific PCR (KASP), which is a gel-free assay that has an allele-specific PCR, has addressed this issue. These markers are suitable for high-throughput genotyping of SNPs and also for insertion/deletion (He et al., 2014). Multiple alleles arise due to the presence of SNPs and InDels in the sequence of nucleotides (Rasheed et al., 2016). These SNPs contribute directly to the phenotypic variation, and this type of polymorphism is important for the development of functional markers. The efficiency of selection in wheat breeding and speed can be improved by conversion of the functional markers into KASP (Rasheed et al., 2018a). The *TaSBEIII* is present in chromosome seven in the A, B, and D genome and plays an important role in granule formation during the grain filling stage, but still, polymorphism of *TaSBEIII* is unclear in wheat.

The main purposes of this study were to (a) identify the polymorphic site and development of the high-throughput KASP marker; (b) identify variation in favorable alleles; (c) identify the geographic distribution of the allelic variations in Chinese, CIMMYT, and Pakistani wheat germplasms; (d) identify gene diversity in diploid, tetraploid, and hexaploid wheat germplasm; and (e) analyze the association between polymorphic sites

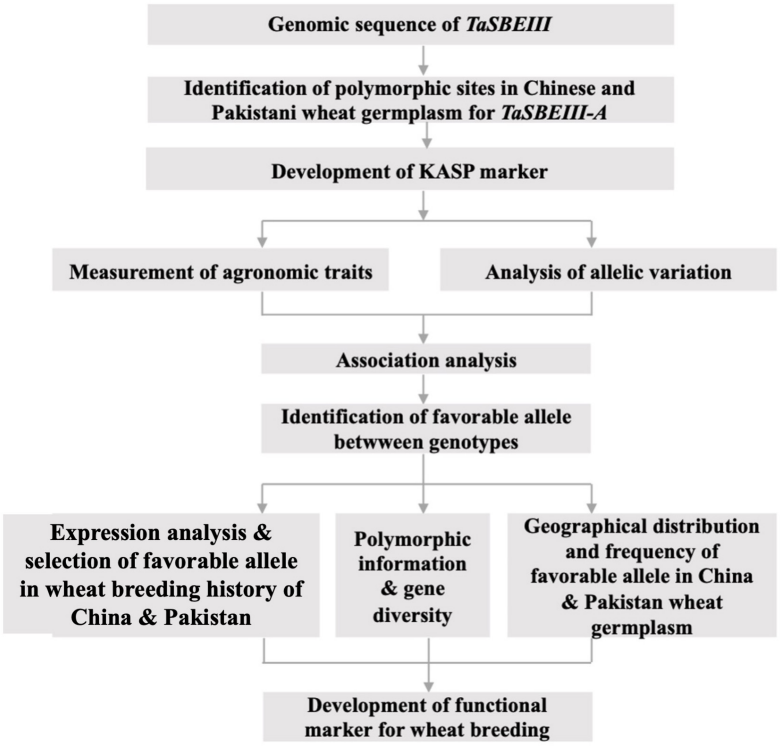


FIGURE 1 | The schematic display of research work.

and phenotypic traits. The research work scheme is shown in **Figure 1**.

MATERIALS AND METHODS

Plant Material and Morphological Data

A diversity panel of 20 accessions was selected, which consists of Chinese wheat landraces (7), Chinese modern wheat cultivars (7)m and wheat accessions from Pakistan (6) to detect the polymorphism in *TaSBEIII* (A, B, and D sub-genomes). The Chinese wheat mini core collection (MCC) consisted of 262 accessions representing more than 70% genetic diversity of wheat germplasm in China (**Supplementary File 1**; Hao et al., 2008), 153 accessions from Pakistan, 53 CIMMYT accessions, and 16 diploid and 19 tetraploid wheat accessions, which were used to evaluate the newly developed molecular markers and their distribution and frequency in different zones. The MCC collection was planted at the Chinese Academy of Agricultural Sciences (CAAS) experimental station in Beijing 2017–2018 (111.6°E, 33.8°N). The plants were grown with two replications, and each single-row plot was 4 m long with a 75-cm space between rows. The plant spacing within the row was 10 cm. Field management was done according to the normal agricultural practices. The data of yield-related traits (TGW, number of grains per spike, plant height, and spike length) were collected from the MCC 262 accessions and used in the present study. Grain numbers were calculated per spike with three replications and TGW was calculated after harvesting the crop by counting the 1,000 seeds from each accessions with three replications. Plant height and spike length were measured at the maturing stage of the crop with the help of a centimeter scale.

Isolation of DNA, Amplification of PCR, Sequencing, and Alignment

DNA of all the accessions was extracted by using the CTAB method (Gawel and Jarret, 1991). The purification of DNA was done by "RNase A" treatment and isoamyl alcohol precipitation (Sambrook, 1987). Agarose gel was used to check the quality and quantity of the DNA. The gene primer *TaSBEIII-F/R* was used to amplify all three genomes simultaneously for the *TaSBEIII* gene covering the CDS sequence (**Table 1**). PCR reaction was carried out in a total volume of 15 μl as initial denaturation at 95°C for 3 min, followed by 32 cycles at 95°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 3 min, with a final extension at 72°C for 10 min. Agarose gel of ∼1% was used to resolve the PCR product by electrophoresis. The resulting PCR fragment was directly sequenced from both directions by a commercial company (Sangon Biotech Co., Ltd., China) to find out the polymorphisms. MEGA software[1] was used to align this 20-genotypes sequence, and Chinese Spring sequence was used as a reference sequence that was downloaded from the NCBI database with accession no. JQ34619.

[1]https://www.megasoftware.net/

TABLE 1 | Primers used in the experiments.

Primer name	Primer sequence 5 to 3	Purpose
TaSBEIII-F/R	F: GGACCTAGAGTCCGCAAGAAGT R: CTACCCTTCACCTCTGCAGTGCT	Genomic fragment amplification
KASP	F: GAAGGTGACCAAGTTCATGCTACTGCTCTCCGCCGGTCGGCTGC**C** F: GAAGGTCGGAGTCAACGGATTACTGCTCTCCGCCGGTCGGCTGC**T** R: CCACAACTTCCTGCCAGGGGCCAC	KASP assay for SNP at 294 nt (C/T)
qRT-PCR-T-Allele	F: GATGGGGACATGCCTAGCAATA R: CTGGACACGGGCTCCACCCAC	*TaSBEIII-A* expression in different tissues of wheat plants
qRT-PCR-C-Allele	F: GCCACGGTAACGCGTAACTGCT R: TCCGGACGGAGAAGACTATGCT	
qRT-PCR-TaActin	F: CTCCCTCACAACAACAACCGC R: TACCAGGAACTTCCATACCAAC	Control for endogenous

Total RNA Extraction

RNAprep pure Plant Kit was used for the extraction of RNA by following the instruction of the manufacturer. Similarly, by using FastQuant RT Kit (Tiangen Beijing), cDNA was synthesized.

Expression Analysis

Two genotypes (Chinese Spring and Aikang-58) were planted on the basis of their *C* and *T* alleles in the field at the experimental station of the Institute of Crop Sciences, CAAS*Chinese Academy of Agricultural Sciences* with normal management. Samples were collected at different stages of both genotypes. The stages that were selected for sampling were leaves/shoot, seedling, spikes at the vegetative growth phase, grains at the reproductive phase (15 days after pollination), roots at the reproductive phase, leaves/shoot at the vegetative stage, spikes at the vegetative phase, spikes at the reproductive phase (after formation of spikes), and root seedling and roots at the vegetative growth phase.

One microgram of RNA was used to synthesize the cDNa with a TranScript Kit (TransGen, AT341) by following the manufacturer's instructions. The concentration of cDNA was quantified with the help of NanoDrop and all the samples were made uniform for qRT-PCR. Two sets of primers were designed for qRT-PCR according to C and T alleles (**Table 1**). TransStart SuperMix Kit (TransGen, AQ131) was used for qRT-PCR and ran on a CFX96 system (Bio-Rad Co., United States). The process of amplification was initiated at 94°C for 30 s, followed by 43 cycles of denaturing for 5 s, annealing for 15 s, and extension for 10 s and then a melting curve stage. Both genotypes were used for qRT-PCR with three biological replications and three technical replications, and actin was used as a housekeeping gene for internal control. The relative expression value was calculated by Microsoft Excel.

KASP Marker Development

On the variant of TaSBEIII-A, KASP primer was designed for high-throughput genotyping by following the standard KASP guidelines[2] (**Table 1**). The specificity of primer was developed based on the standard FAM (5′-GAAGGTGACCAAGTTCATGCT-3′) and HEX (5′-GAAGGTCGGAGTCAACGGATT-3′) tails with a targeted SNP at the 3′ end. To check the diversity, the developed KASP marker was applied across the entire studied population. The methodology of KASP assay was followed as reported by Rehman et al. (2019). The clustering of accessions was shown on the scatter plot at the x (FAM) and y (HEX) signal.

Association Analysis Between SNPs and Yield-Related Traits

Descriptive statistics and estimates of variance were done by using Microsoft Excel 2016. To check the effect of allelic variants on yield-related traits, Student's *t*-test was used at $p < 0.05$ (even 0.01). The polymorphic information content (PIC) and gene diversity (H_e) were measured online: https://www.gene-calc.pl/pic.

RESULTS

Identification of Novel SNP in the CDS Sequence of *TaSBEIII-A*

PCR amplification and sequencing of *TaSBEIII-A* in 20 accessions allowed the identification of a SNP in the exon region of *TaSBEIII-A* at 294 bp. The transversion of SNP was "*C*" substituted into "*T*". The SNP was non-synonymous and proline changed into serine amino acid. The alignment of sequences with Chinese Spring accession revealed that SNP is novel and was reported for the first time (**Supplementary Figure 1**). The sequence had been submitted to the NCBI GenBank and the accession number is MZ261926.

High-Throughput KASP Marker for *TaSBEIII-A*

A KASP marker was developed for the identified polymorphic site in the CDS sequence of *TaSBEIII-A* and named as a KASP1 marker. The frequency of this KASP1 marker was more than 5% in the MCC. In the scatter plot, the accession in blue circles

[2]http://www.lgcgenomics.com

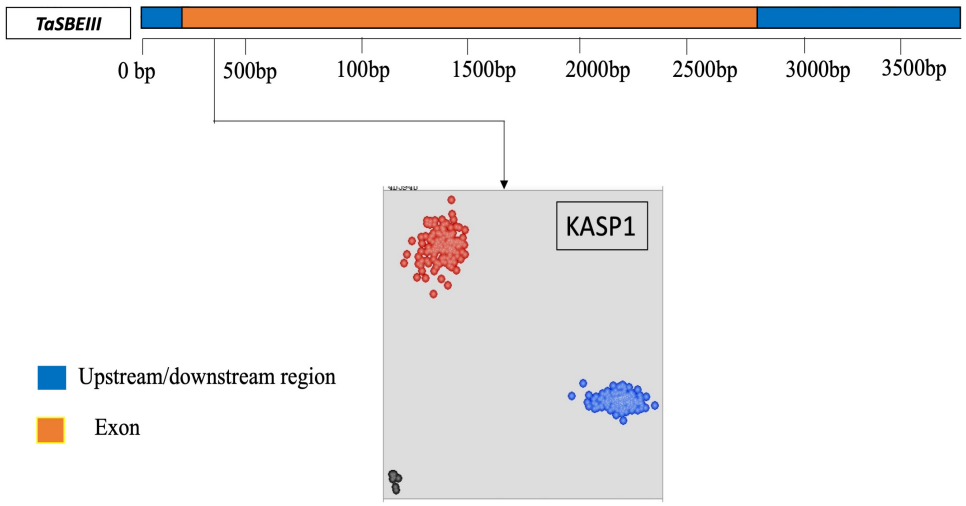


FIGURE 2 | Gene structure and developed KASP marker for *TaSBEIII-A*. Gene structure of *TaSBEIII-A* consists of an upstream and downstream region with blue color and orange blocks representing the exon region. Bp represents the base pairs. The clusters of accessions are represented on the scatter plot on the *x*-axis (FAM) and *y*-axis (HEX). Accessions with blue color represent the FAM type of allele *C* while accession in red color denotes the Hex type of allele *T*. Black dots represent the non-template control (NTC).

have a "*C*" allele while the accessions in red circles have "*T*" alleles for the *TaSBEIII-A* gene (**Figure 2**). The polymorphism for this marker was identified in MCC, Pakistani wheat accessions, CIMMYT wheat accessions, and diploid and tetraploid wheat accessions (**Figure 2**).

Association Between Allele and Yield-Related Ttraits

Mini core collection consisted of 262 accessions that were used to detect the association of *TaSBEIII-A* alleles with yield-related traits. Different yield-related traits data were collected such as TGW, number of grains per spike, spike length, and plant height. The data were collected with three replications and the average of these replications was used for association analysis. *Allele-T* was significantly associated with higher TGW (**Figure 3**). All other yield-related traits (spike length, number of grains per spike, and plant height) showed non-significant association between these two alleles. Therefore, on the basis of this result, it can be said that *Allele-T* significantly associated with yield-related traits and has the potential as a functional marker and that it can be used in MAS from improving starch content, which ultimately affects the yield of the grain.

Expression Analysis of *TaSBEIII-A*

Two wheat accessions (Chinese Spring and Aikang-58) were selected based on allelic variation (*Allele-C* and *Allele-T*, respectively) to test the expression level of *TaSBEIII-A* in different stages of wheat plant. Primer specificity was checked on the basis of the melting curve of qRT-PCR for *TaSBEIII-A*, and *TaActin* gene was used as an internal control. qRT-PCR revealed that Aikang-58 with *Allele-T* had higher relative expression levels as compared with Chinese Spring having *Allele-C* at the spike reproductive stage and grain reproductive stage (**Figure 4**). These findings suggested that *Allele-T* might be responsible for high TGW and help to maintain a high starch content.

PIC and Gene Diversity in Chinese and Non-Chinese Wheat Germplasm

To study the evolutionary history of *TaSBIII-A*, we analyzed *TaSBIII-A* in wheat ancestors. The results illustrated that during polyploidization events, diversity in *TaSBIII-A* increased. Diploid (AA) wheat accessions showed no diversity for *TaSBIII-A*. Tetraploid accessions (AABB) showed 0.3108 PIC and 0.3848 H_e for *TaSBIII-A*. PIC and H_e ranged between 0.3318 and 0.420 in hexaploid Chinese wheat landraces, respectively. For Chinese

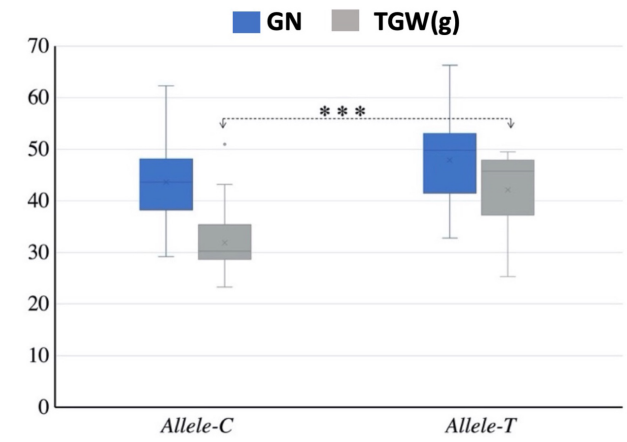


FIGURE 3 | Interaction of favorable allele with yield-related traits. TGW, thousand grain weight; GN, grain number; *$p < 0.05$, **$p < 0.01$. The 2-year data of 2017 and 2018.

modern wheat cultivars, the PIC and H_e ranged between 0.3701 and 0.4902, respectively (**Table 2**). Among non-Chinese wheat germplasm, CIMMYT wheat accession showed 0.3749 PIC and 0.4998 H_e, while Pakistani wheat accessions showed 0.3648 PIC and 0.48 H_e.

Geographic Distribution of *TaSBEIII-A* Allelic Variation

In China, wheat zones are divided on the basis of temperature, growing season, moisture contents, varietal response to photoperiod, and biotic and abiotic stress (Zhang et al., 2015). In this study, MCC was used to survey allelic variation of *TaSBEIII-A* in China, Pakistan, and CIMMYT wheat collections. Based on cultivation and production, zones I–IV are the major zones and cover 75% of the wheat area of China. The frequency of the favored allele *Allele-T* was lower in landraces and *Allele-C* was dominant in all the major zones of China. However, the frequency of favored *Allele-T* in modern cultivars was >50% in four major zones of wheat. The frequency of *Allele-T* showed a significant increment from 35 to 65% in zone I, 27 to 85% in zone II, 20 to 66% in zone III, and 19 to 85% in zone IV from landraces to modern cultivars, respectively (**Figure 5**). These results support the idea that favorable allelic variation was positively selected in all major zones and other zones of China with the passage of time. Therefore, this allele can be used further in breeding programs in China to increase the grain yield of wheat.

The geographic distribution of *TaSBEIII-A* was also investigated among Pakistani wheat accessions. The frequency of *Allele-T* was higher in Pakistan major zones such that its frequency was 64% in the Punjab irrigated zone and 55% in the Punjab rainfed zone. Similarly, the frequency in Khyber-Pakhtunkhwa was 65% and that in Sindh was 70%. There was a positive selection of the favorable allele in all zones of Pakistan (**Figure 6**). Similarly, in the CIMMYT germplasm, the frequency of *Allele-T* was higher than that of *Allele-C*.

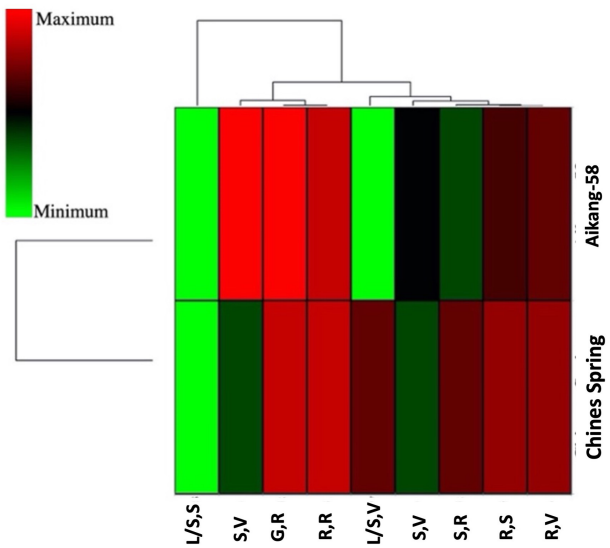


FIGURE 4 | A heatmap of two genotypes having different alleles and their gene expression profile at different stages by using hierarchical clustering. XLSTAT was used to create a heatmap. R,S: root seedling; R,V: roots at the vegetative growth phase; R,R: roots at the reproductive phase; L/S,S: leaves/shoot seedling; L/S,V: leaves/shoot, vegetative; S,V: spikes at the vegetative growth phase; SR: spikes at the reproductive phase; GR: grains at the reproductive phase.

TABLE 2 | Polymorphic information content (PIC) and gene diversity (H_e) values in studied wheat germplasms.

	262MCC	CL	CM	CIMMYT	Tetra	PAK
H_e	0.455	0.42	0.4902	0.4998	0.3848	0.48
PIC	0.315	0.3318	0.3701	0.3749	0.3108	0.3648

MCC, Mini core collection of China; CL, Chinese landraces; CM, Chinese modern cultivars; Tetra, Tetraploidy accessions; PAK, Pakistan.

Positive Selection of *Allele-T* of *TaSBEIII-A* in Wheat Breeding History of China and Pakistan

To evaluate the favorable allelic variation of *Allele-T*, MCC was used with known released dates and was divided into six groups (pre-1950, 1950s, 1960s, 1970s, 1980s, and 1990s). In general, accessions that were released before 1950 possessed *Allele-C* and few accessions had *Allele-T* (25%). The frequency of the favorable allele (*Allele-T*) increases from 1950 to 1960 (up to 38%) but remained stable from the 1960–1970 era. The frequency of the favorable allele increased from 1971 to 1990 (38 to 70%) and it became 80% in 2000 (**Figure 7**). From the 1960s onward, TGW also showed a continuous increasing trend (**Supplementary Figure 2**). These results indicated that this favorable variation is valuable and could be selected to further improve TGW in Chinese wheat germplasm.

Similar results had been depicted in the Pakistan accessions, which consist of 153 genotypes and divided into two groups (1953–1980 and 1981–2016). The two groups were determined on the basis of pre-green revolution (1953–1980) and post-green revolution (1981–2016). The initiation of green revolution started in Pakistan from the early 1970s. The frequency of *Allele-T* was 15% in 1953, and the frequency of *Allele-C* was 80%, but with passage, the frequency of the favorable allele increased. From 1981 to 2016, the frequency of the favorable allele was 55% (**Figure 7**). So, the favorable allele was also selected positively with the passage of time in Pakistani wheat accessions.

DISCUSSION

Due to domestication, evolution and breeding in wheat help to create a lot of genetic diversity in wheat germplasm. The level of polymorphism in wheat was 1 SNP/540 bp based on the bioinformatic analysis of large wheat EST database of 12 accessions (Somers et al., 2003). Similarly, genomic sequences consisting of coding and non-coding regions have 1 SNP/334 bp in the coding region and 1 SNP/267 bp in the genomic region (Ravel et al., 2006). In this study, *TaSBEIII* CDS was sequenced in the 20 diverse cultivars of wheat and polymorphism was

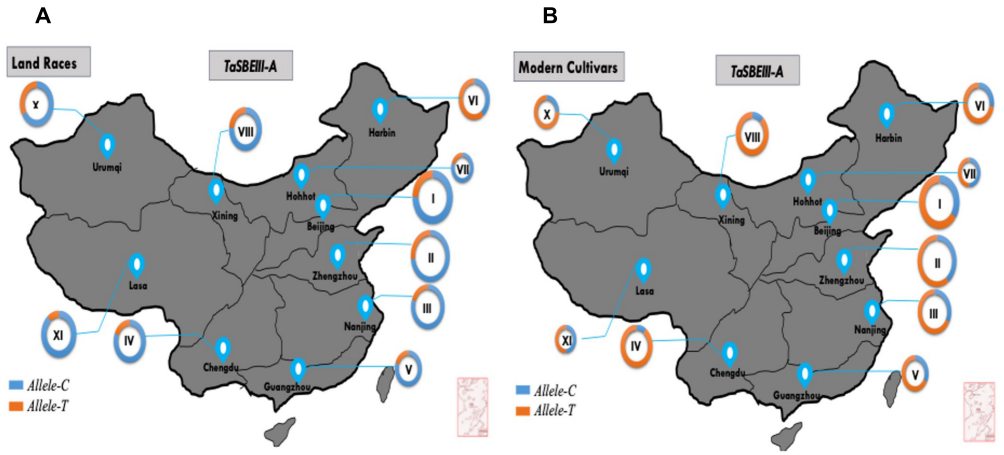


FIGURE 5 | *TaSBEIII-A* alleles. Geographic distribution in China. **(A)** Landraces, **(B)** modern cultivars. I, Northern winter wheat zone; II, Yellow, and Huai River valleys winter wheat zone; III, Middle and low Yangtze valleys winter wheat zone; IV, Southwestern winter wheat zone; V, Southern winter wheat zone; VI, Northeastern spring wheat zone; VII, Northern spring wheat zone; VIII, Northwestern spring wheat zone; IX, Qinghai-Tibetan spring-winter wheat zone; X, Xinjiang winter-spring wheat zone. Pie chart size is directly proportional to the number of genotypes.

detected. The SNP was detected in the exon region of *TaSBEIII-A* (**Figure 2**). There was no other SNP detected in the CDS sequence in the rest of the sub-genomes (B and D sub-genomes) for *TaSBEIII*. These results suggested that *TaSBEIII* is a conserved gene during evolution. Wheat D genome had a narrow genetic background with a lower level of polymorphism. This might be due to the fact that no SNPs in the CDS sequence of these sub-genomes have been reported yet (Rasheed et al., 2018b). Allele fixation during domestication and low genetic diversity in the wheat panel can also be other reasons for less polymorphism in this gene. There is a need to sequence diverse wheat germplasm to investigate these probabilities.

According to different research, it has been predicted that in wheat, per year genetic gain is ∼0.8 to 1% (Shearman et al., 2005; Zhou et al., 2007). Grain number per square meter plays a significant role for achieving genetic gain with a slight change in the grain weight (Gaju et al., 2009; Zhang et al., 2015). In this study, a SNP was identified for *TaSBEIII-A* in the CDS sequence of gene at 294 bp position and showed a significant association with TGW, which might be beneficial for improving grain yield. The expression of *TaSBEIII* was consistent at the grain filling stage, and the function of this gene might be different from other *SBE* genes (*SBEI*, *SBEIIa*, and *SBEIIb*) and might help to improve wheat yield. The function of this gene may be associated with the formation of A and B starch granules in the grains of wheat plant, which ultimately help in the wheat yield (Kang et al., 2013). Similarly in wheat, an open excess browser[3] was developed to check the expression of the gene at different stages. The expression of this gene is also observed at the grain stage by using this web browser, which supports the results that show its function in grains of wheat (Borrill et al., 2016). The final dry weight of the grain contains 65 to 80% starch (Hurkman et al., 2003). The endosperm works as a storage tissue and contributes significantly to the yield of the grain. Therefore, the differential effects of the *TaSBEIII-A* allele on grain weight detected in the present study might be caused by different contributions to starch biosynthesis and hence to endosperm development. Generally, all over the world, a higher grain weight is the main objective of wheat breeders (Zhou et al., 2007).

The development of new genetic tools promises to address the challenges by improving the genetic gains of different crops and help to meet the world's food production demand. The use of functional markers in wheat breeding programs through MAS is a successful approach to increase yield (Liu et al., 2014; Rasheed et al., 2018a). Recently the concept of MAS has been changed by shifting to whole-genome methods to attain maximum genetic gains in wheat breeding by using different complex traits (Zhao et al., 2014). The KASP1 marker was developed based on the SNP present at 294 bp in the CDS sequence of *TaSBEIII-A* (**Figure 3**).

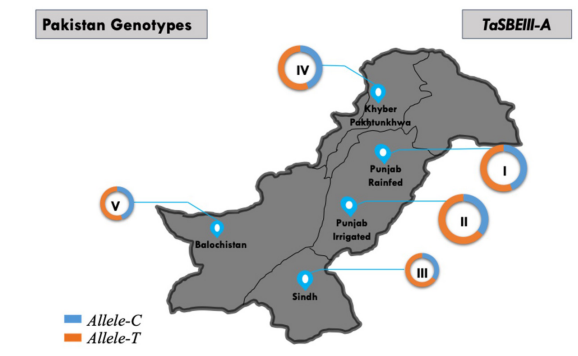


FIGURE 6 | Distribution of *TaSBEIII-A* alleles among Pakistan accessions. I, Punjab rainfed zone; II, Punjab irrigated zone; III, Sindh region; IV, Balochistan zone; V, Khyber Pakhtunkhwa zone. Pie chart size is directly proportional to the number of accessions.

[3]http://www.wheat-expression.com

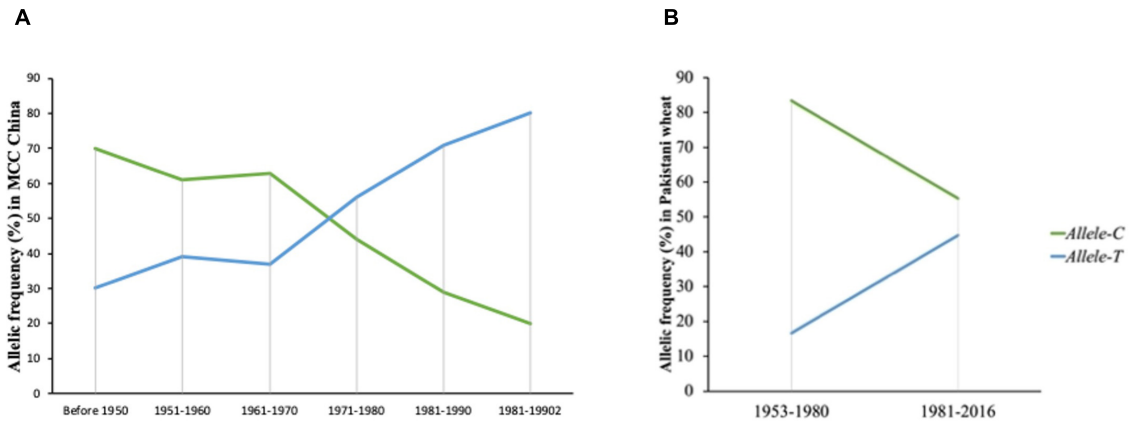


FIGURE 7 | Selection of favored allele of *TaSBEIII-A* in wheat breeding history of China and Pakistan. **(A)** The frequency of China accessions. **(B)** The frequency of Pakistan accessions.

Two allelic variations (*Allele-C* and *Allele-T*) were observed in different wheat populations by using this KASP marker. By using further association analysis of these alleles, it was observed that *Allele-T* showed a significant association with high TGW in MCC. Based on this, it is depicted that this molecular marker can be instrumental in MAS to improve the yield of the wheat.

Polymorphic information content helps to understand the detailed knowledge of the level of polymorphism between accessions. On the basis of previous reports, PIC can be divided into three categories: (1) the marker is considered to highly polymorphic if the value is more than 0.5; (2) similarly, values between 0.25 and 0.4 indicate that the marker is moderately informative; and (3) the marker with a 0.25 PIC value is a low informative marker. In the present study, the average PIC value for landrace and modern cultivars was 0.32, while the value was increasing from tetraploid to hexaploid wheat accessions. This value is in agreement with previous studies using bi-allelic markers such as SNP or DArT in either common or durum wheat. In common wheat, a PIC value of 0.24 has been found for the WAMI population genotyped with the 9K SNP array (Rasheed et al., 2018a).

The favorable allele was also investigated in landraces and modern cultivars in China. The selection of the favorable allele of *TaSBEIII-A* showed an increasing trend from landraces to modern cultivars in wheat breeding in China. In major wheat-producing areas of China (Zones I, II, III, and IV), the frequency of the favorable *Allele-T* showed higher trends. Zones I, II, and III contribute about 64% of the total national area of China (Irshad et al., 2019), and accessions in these zones have a higher frequency of favorable alleles. The average grain weight in these regions is 42–44 g, especially in zone II (Barrero et al., 2011). Wheat yield mainly depends on the increase in TGW (Zhang et al., 2012). The frequency of favorable alleles increased from 1960 onward, and during that time, Chinese wheat varieties experienced a boom in their yield. In China, the main objective was to increase TGW in wheat varieties before 1960 (Wang et al., 2019). With the passage of time from 1970 to 1990, the breeding objective was also changed by adding other traits such as plant height, grain number per spike, and quality traits to improve the yield of Chinese wheat accessions (He et al., 2018). The change in breeding objective from 1950 to 1980 assisted in the positive selection of favorable alleles in Chinese wheat with a rapid increase in TGW before 1980 in wheat accessions (**Figure 6**). The frequency of favorable alleles increases by about 80% from the 1970s to the 1990s with an increase in TGW, which may be the reason for selecting the other favorable genes (Wang et al., 2018). Additionally, the frequency of favorable alleles increases in all 10 zones of China from landraces to modern cultivars. So, it can be said that this allele has large potential to increase TGW, which ultimately increases the yield of wheat crop.

Wheat accessions from Pakistan were also selected to evaluate the favorable allele diversity in the different wheat zones. In all major zones of Pakistan, the favorable allele was positively selected and with high frequency. From 1953 to 2016, the favorable allele was positively selected with the passage of time, showing that there is a positive selection of the favorable allele in the wheat breeding program of Pakistan. The population structure of Pakistan and China is different, but the positive selection of alleles in both germplasms is likely due to the high linkage disequilibrium of wheat in major yield genes, and these genes were elected during selection breeding (Semagn et al., 2014; Rossi et al., 2015; Irshad et al., 2019; Rehman et al., 2019). To confirm these results, there is a need to analyze this favorable allele in other wheat cultivars such as Europe, United States, and Australia. Based on the frequency result of these regions, the conclusion can be made that the selection of the favorable allele of *TaSBEIII-A* is due to the major yield-related genes.

In conclusion, high-throughput genotyping for MAS is of great importance. The molecular marker that was developed on the SNP *TaSBEIII-A* at 294 bp in which *Allele-T* was significantly associated with TGW and the frequency of this favorable allele increased about 80% from 1960 to 1990 in Chinese MCC. This allele can be used in future studies as selection criteria for improving yield traits. Thus, it is depicted that favorable alleles are valuable and could be selected to increase grain yield, and a

gel-free KASP marker approach can help to improve the speed of wheat breeding.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the NCBI Repository, accession number MZ261926 (https://www.ncbi.nlm.nih.gov).

AUTHOR CONTRIBUTIONS

AI, HG, JG, and LZ conceptualized the study. AI, SU, HG, XW, JG, HX, and CW performed the experiments and analyzed the data. AI, YX, LZ, SZ, and HG wrote the manuscript. LL reviewed the manuscript and assisted in the completion of the experiments. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by the NSFC project (31771791), the National Key Research and Development Program (2016YFD0102100), and the China Agriculture Research System (CARS-03).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.697294/full#supplementary-material

REFERENCES

Barrero, R. A., Bellgard, M., and Zhang, X. (2011). Diverse approaches to achieving grain yield in wheat. *Funct. Integr. Genomics* 11, 37–48. doi: 10.1007/s10142-010-0208-x

Bertoft, E. (2017). Understanding starch structure: recent progress. *Agronomy* 7:56. doi: 10.3390/agronomy7030056

Borrill, P., Ricardo, R. J., and Cristobal, U. (2016). expVIP: a customizable RNA-seq data analysis and visualization platform. *Plant physiol.* 170, 2172–2186. doi: 10.1104/pp.15.01667

Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* 46, 714–721. doi: 10.1038/ng.3007

Gaju, O., Reynolds, M. P., Sparkes, D. L., and Foulkes, M. J. (2009). Relationships between large-spike phenotype, grain number, and yield potential in spring wheat. *Crop Sci.* 49, 961–973. doi: 10.2135/cropsci2008.05.0285

Gawel, N. J., and Jarret, R. L. (1991). A modified CTAB DNA extraction procedure for Musa and Ipomoea. *Plant Mol. Biol. Rep.* 9, 262–266.

Han, Y., Sun, F. J., Rosales-Mendoza, S., and Korban, S. S. (2007). Three orthologs in rice, *Arabidopsis*, and Populus encoding starch branching enzymes (SBEs) are different from other SBE gene families in plants. *Gene* 401, 123–130. doi: 10.1016/j.gene.2007.06.026

Hao, C., Dong, Y., Wang, L., You, G., Zhang, H., Ge, H., et al. (2008). Genetic diversity and construction of core collection in Chinese wheat genetic resources. *Chin. Sci. Bull.* 53, 1518–1526. doi: 10.1007/s11434-008-0212-x

Hayashi, K., Hashimoto, N., Daigen, M., and Ashikawa, I. (2004). Development of PCR-based SNP markers for rice blast resistance genes at the Piz locus. *Theor. Appl. Genet.* 108, 1212–1220. doi: 10.1007/s00122-003-1553-0

He, C., Holme, J., and Anthony, J. (2014). SNP genotyping: the KASP assay. *Methods Mol. Biol.* 1145, 75–86. doi: 10.1007/978-1-4939-0446-4_7

He, Z. H., Zhuang, Q. S., Cheng, S. H., Yu, Z. W., Zhao, Z. D., and Liu, X. (2018). Wheat production and technology improvement in China. *J. Agric.* 8, 107–114.

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

Hurkman, W. J., McCue, K. F., Altenbach, S. B., Korn, A., Tanaka, C. K., Kothari, K. M., et al. (2003). Effect of temperature on expression of genes encoding enzymes for starch biosynthesis in developing wheat endosperm. *Plant Sci.* 164, 873–881. doi: 10.1016/S0168-9452(03)00076-1

Irshad, A., Guo, H., Zhang, S., Gu, J., Zhao, L., Xie, Y., et al. (2019). EcoTILLING Reveals Natural Allelic Variations in Starch Synthesis Key Gene TaSSIV and Its Haplotypes Associated with Higher Thousand Grain Weight. *Genes* 10:307. doi: 10.3390/genes10040307

Jeon, J., Ryoo, N., Hahn, T., 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

Kang, G., Li, S., Zhang, M., Peng, H., Wang, C., Zhu, Y., et al. (2013). Molecular cloning and expression analysis of the starch-branching enzyme III gene from common wheat (*Triticum aestivum*). *Biochem. Genet.* 51, 377–386. doi: 10.1007/s10528-013-9570-4

Li, B., Li, Q., Mao, X., Li, A., Wang, J., Chang, X., et al. (2016). Hao, C.; Zhang, X. Jing, R. Two novel AP2/EREBP transcription factor genes TaPARG have pleiotropic functions on plant architecture and yield-related traits in common wheat. *Front Plant Sci.* 7:1191. doi: 10.3389/fpls.2016.01191

Li, Q., Li, L., Yang, X., Warburton, M. L., Bai, G., Dai, J., et al. (2010). Relationship, evolutionary fate and function of two maize co-orthologs of rice GW2associated with kernel size and weight. *BMC Plant Biol.* 10:143. doi: 10.1186/1471-2229-10-143

Liu, S. Y., Rudd, J. C., Bai, G. H., Haley, S. D., Ibrahim, A. M. H., Xue, Q. W., et al. (2014). Molecular markers linked to important genes in hard winter wheat. *Crop Sci.* 54, 1304–1321. doi: 10.2135/cropsci2013.08.0564

Liu, Y. N., He, Z. H., Appels, R., and Xia, X. C. (2012). Functional markers in wheat: current status and future prospects. *Theor. Appl. Genet.* 125, 1–10. doi: 10.1007/s00122-012-1829-3

Luo, W. L., Guo, T., Yang, Q. Y., Wang, H., Liu, Y. Z., Zhu, X. Y., et al. (2014). Stacking of five favorable alleles for amylase content, fragrance and disease resistance into elite lines in rice (*Oryza sativa*) by using four HRM-based markers and a linked gel-based marker. *Mol. Breed.* 34, 805–815. doi: 10.1007/s11032-014-0076-5

Nemri, A., Atwell, S., Tarone, A. M., Huang, Y. S., Zhao, K., Studholme, D. J., et al. (2010). Genome-wide survey of *Arabidopsis* natural variation in downy mildew resistance using combined association and linkage mapping. *Proc. Natl. Acad. Sci. U. S. A.* 107, 10302–10307. doi: 10.1073/pnas.0913160107

Pfister, B., and Zeeman, S. C. (2016). Formation of starch in plant cells. *Mol. Life Sci.* 73, 2781–2807. doi: 10.1007/s00018-016-2250-x

Qi, Z., Huang, L., Zhu, R., Xin, D., Liu, C., Han, X., et al. (2014). A high-density genetic map for soybean based on specific length amplified fragment sequencing. *PLoS One* 9:104871. doi: 10.1371/journal.pone.0104871

Ral, J. P., Bowerman, A. F., Li, Z., Sirault, X., Furbank, R., Pritchard, J. R., et al. (2012). Down-regulation of Glucan, Water-Dikinase activity in wheat endosperm increases vegetative biomass and yield. *Plant Biotechnol. J.* 10, 871–882.

Rasheed, A., Mujeeb-Kazi, A., Ogbonnaya, F. C., He, Z., and Rajaram, S. (2018a). Wheat genetic resources in the post-genomics era: promise and challenges. *Ann. Bot.* 121, 603–616. doi: 10.1093/aob/mcx148

Rasheed, A., Ogbonnaya, F. C., Lagudah, E., Appels, R., and He, Z. (2018b). The goat grass genome's role in wheat improvement. *Nat. Plant* 4, 56–58. doi: 10.1038/s41477-018-0105-1

Rasheed, A., Wen, W., Gao, F., Zhai, S., Jin, H., Liu, J., et al. (2016). Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theor. Appl. Genet.* 129, 1843–1860. doi: 10.1007/s00122-016-2743-x

Ravel, C., Praud, S., Murigneux, A., Canaguier, A., Sapet, F., Samson, D., et al. (2006). Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). *Genome* 49, 1131–1139. doi: 10.1139/g06-067

Rehman, S. U., Wang, J., Chang, X., Zhang, X., Mao, X., and Jing, R. (2019). A wheat protein kinase gene TaSnRK2. 9-5A associated with yield contributing traits. *Theor. Appl. Genet.* 132, 907–919. doi: 10.1007/s00122-018-3247-7

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., et al. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* 524, 230–233. doi: 10.1038/nature14580

Sambrook, J. (1987). Commonly used techniques in molecular cloning. *Mol. Cloning* 3, 1–39.

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using kompetitve allele specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33, 1–14. doi: 10.1007/s11032-013-9917-x

Shearman, V. J., Bradley, R. S., Scott, R. K., and Foulkes, M. J. (2005). Physiological processes associated with wheat yield progress in UK. *Crop Sci.* 45, 175–185. doi: 10.2135/cropsci2005.0175a

Somers, D. J., Kirkpatrick, R., Moniwa, M., and Walsh, A. (2003). Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. *Genome* 46, 431–437. doi: 10.1139/g03-027

Stamova, B. S., Laudencia-Chingcuanco, D., and Beckles, D. M. (2009). Transcriptomic analysis of starch biosynthesis in the developing grain of hexaploid wheat. *Int. J. Plant Genomics* 2009:407426. doi: 10.1155/2009/407426

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, B., Tan, H. W., Fang, W., Meinhardt, L. W., Mischke, S., Matsumoto, T., et al. (2015). Developing single nucleotide polymorphism (SNP) markers from transcriptome sequences for identification of longan (*Dimocarpus longan*) germplasm. *Hortic. Res.* 2, 1–10. doi: 10.1038/hortres.2014.65

Wang, H., Wang, S., Chang, X., Hao, C., Sun, D., and Jing, R. (2019). Identification of TaPPH-7A haplotypes and development of a molecular marker associated with important agronomic traits in common wheat. *BMC Plant Biol.* 19:296. doi: 10.1186/s12870-019-1901-0

Wang, Y. X., Xu, Q. F., Chang, X. P., Hao, C. Y., Li, R. Z., and Jing, R. L. (2018). A dCAPS marker developed from a stress associated protein gene TaSAP7-B governing grain size and plant height in wheat. *J. Integr. Agr.* 17, 276–284. doi: 10.1016/S2095-3119(17)61685-X

Xia, J., Zhu, D., Chang, H., Yan, X., and Yan, Y. (2020). Effects of water-deficit and high-nitrogen treatments on wheat resistant starch crystalline structure and physicochemical properties. *Carbohydr. Polym*. 234:115905.

Yan, H. B., Pan, X. X., Jiang, H. W., and Wu, G. J. (2009). Comparison of the starch synthesis genes between maize and rice: copies, chromosome location and expression divergence. *Theor. Appl. Genet.* 119, 815–825. doi: 10.1007/s00122-009-1091-5

Zhang, B., Liu, X., Xu, W., Chang, J., Li, A., Mao, X., et al. (2015). Novel function of a putative MOC1 ortholog associated with spikelet number per spike in common wheat. *Sci. Rep*. 5, 1–13. doi: 10.1038/srep12211

Zhang, D., Hao, C., Wang, L., and Zhang, X. (2012). Identifying loci influencing grain number by microsatellite screening in bread wheat (*Triticum aestivum* L.). *Planta* 236, 1507–1517. doi: 10.1007/s00425-012-1708-9

Zhang, Y. X., Wang, L. H., Xin, H. G., Li, D. H., Ma, C. X., and Xia, D. (2013). Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biol.* 13:141. doi: 10.1186/1471-2229-13-141

Zhao, Y., Mette, M. F., Gowda, M., Longin, C. F., and Reif, J. C. (2014). Bridging the gap between marker-assisted and genomic selection of head- ing time and plant height in hybrid wheat. *Heredity* 112, 638–645. doi: 10.1038/hdy.2014.1

Zheng, S., Li, Y., Lu, L., Liu, Z., Zhang, C., Ao, D., et al. (2017). Evaluating the contribution of Yr genes to stripe rust resistance breeding through marker-assisted detection in wheat. *Euphytica* 213:50. doi: 10.1007/s10681-016-1828-6

Zheng, T. C., Zhang, X. K., Yin, G. H., Wang, L. N., Han, Y. L., Chen, L., et al. (2011). Genetic gains in grain yield, net photosynthesis and stomatal conductance achieved in Henan Province of China between 1981 and 2008. *Field Crops Res.* 122, 225–233. doi: 10.1016/j.fcr.2011.03.015

Zhou, Y., He, Z. H., Sui, X. X., Xia, X. C., Zhang, X. K., and Zhang, G. S. (2007). Genetic improvement of grain yield and associated traits in the northern China winter wheat region from 1960 to 2000. *Crop Sci.* 47, 245–253. doi: 10.2135/cropsci2006.03.0175

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 19 July 2021
doi: 10.3389/fgene.2021.713446

Rapid Identification of QTL for Mesocotyl Length in Rice Through Combining QTL-seq and Genome-Wide Association Analysis

*Yamei Wang[1,2], Jindong Liu[1,2], Yun Meng[2,3], Hongyan Liu[2,3], Chang Liu[1,2] and Guoyou Ye[1,2,4]**

[1] Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, [2] CAAS-IRRI Joint Laboratory for Genomics-Assisted Germplasm Enhancement, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, [3] College of Tropical Crops, Hainan University, Haikou, China, [4] Rice Breeding Innovation Platform, International Rice Research Institute, Metro Manila, Philippines

OPEN ACCESS

Edited by:
Awais Rasheed,
Quaid-i-Azam University, Pakistan

Reviewed by:
Jingguang Chen,
Sun Yat-sen University, China
Nisha Singh,
Cornell University, United States
Shahzad Amir Naveed,
Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, China

***Correspondence:**
Guoyou Ye
G.Ye@irri.org

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 23 May 2021
Accepted: 15 June 2021
Published: 19 July 2021

Citation:
Wang Y, Liu J, Meng Y, Liu H, Liu C and Ye G (2021) Rapid Identification of QTL for Mesocotyl Length in Rice Through Combining QTL-seq and Genome-Wide Association Analysis. Front. Genet. 12:713446. doi: 10.3389/fgene.2021.713446

Mesocotyl is a crucial organ for pushing buds out of soil, which plays a vital role in seedling emergence and establishment in direct-seeded rice. Thus, the identification of quantitative trait loci (QTL) associated with mesocotyl length (ML) could accelerate genetic improvement of rice for direct seeding cultivation. In this study, QTL sequencing (QTL-seq) applied to 12 F_2 populations identified 14 QTL for ML, which were distributed on chromosomes 1, 3, 4, 5, 6, 7, and 9 based on the Δ(SNP-index) or *G*-value statistics. Besides, a genome-wide association study (GWAS) using two diverse panels identified five unique QTL on chromosomes 1, 8, 9, and 12 (2), respectively, explaining 5.3–14.6% of the phenotypic variations. Among these QTL, seven were in the regions harboring known genes or QTLs, whereas the other 10 were potentially novel. Six of the QTL were stable across two or more populations. Eight high-confidence candidate genes related to ML were identified for the stable loci based on annotation and expression analyses. Association analysis revealed that two PCR gel-based markers for the loci co-located by QTL-seq and GWAS, *Indel-Chr1:18932318* and *Indel-Chr7:15404166* for loci *qML1.3* and *qML7.2* respectively, were significantly associated with ML in a collection of 140 accessions and could be used as breeder-friendly markers in further breeding.

Keywords: candidate gene, GWAS, mesocotyl, QTL-seq, rice

INTRODUCTION

Rice (*Oryza sativa*) is one of the most important food crops in the world, providing more than 21% of the food for the world's population. Maintaining a higher and stable yield is of great importance for food security, especially in developing countries in Asia[1]. Transplanting and direct seeding are two major rice planting patterns. Direct seeding refers to the process of establishing seedlings into

[1] www.fao.org

puddled or submerged soil without the transplanting process (Kumar and Ladha, 2011; Zhan et al., 2020). Compared with traditional rice transplanting, direct seeding is water-efficient and labor-saving (Kato and Katsura, 2014; Liu et al., 2015; Ohno et al., 2018). However, direct seeding is also facing problems such as low seedling emergence rate, poor seedling establishment, weed infestation, and high crop lodging rate (Mahender et al., 2015; Lee et al., 2017). Mesocotyl, an organ between the coleoptile node and the basal part of the seminal root in rice seedlings, plays a key role in pushing buds out of deep water or soil for successful seedling establishment during germination (Zhan et al., 2020). Thus, varieties with longer mesocotyl can be used to partially overcome the problems faced by direct seeding (Lee et al., 2017; Zhan et al., 2020).

Mesocotyl length (ML) is a quantitative trait controlled by multiple minor effect genes (Li et al., 2017; Sun et al., 2018; Liu et al., 2020; Zhan et al., 2020). Up to now, over 40 quantitative trait loci (QTL) on 12 rice chromosomes have been identified, which could explain 5.7–27.8% of the phenotypic variation (Cao et al., 2002; Liu et al., 2020; Rohilla et al., 2020; Zhan et al., 2020). Four genes were cloned: *OsGY1* (Xiong et al., 2017), *OsGSK2* (Sun et al., 2018), *OsSMAX1* (Zheng et al., 2020), and *OsPAO5* (Lv et al., 2021). Recently, genome-wide association study (GWAS) based on linkage disequilibrium (LD) has been widely applied to identify marker–trait associations (MTAs) for complex agronomic traits (Meng et al., 2016; Zhan et al., 2020; Liu et al., 2021). Compared with traditional biparental linkage mapping, GWAS provides a more representative gene pool because all the historical meiotic events can be counted from a diverse panel and is an efficient tool that bypasses the time and expand to the developing population (Flint-Garcia et al., 2003; Breseghello and Sorrells, 2006; Zhu et al., 2008). Now, GWAS has been adopted to investigate a range of complex traits in crops, including disease resistance (Liu et al., 2017; Resende et al., 2017; Prodhomme et al., 2020), grain quality (Yang et al., 2020), yield-related traits (Meng et al., 2016), salt tolerance (Zhang et al., 2020; Ponce et al., 2021), and microelements (Chen et al., 2019; Liu et al., 2020).

To rapidly identify QTL in plants, QTL sequencing (QTL-seq), an effective and economic approach combining the traditional bulk segregant analysis (BSA) and high-throughput whole-genome resequencing, has been developed (Takagi et al., 2013). For QTL-seq, a mapping population was firstly generated by crossing two cultivars showing the extreme target phenotypes, and then two DNA pools from individuals with extreme phenotype in the population and two pools from the parents were constructed and sequenced (Takagi et al., 2013). This method has been successfully used to rapidly identify QTL for a number of traits in rice, such as blast disease resistance and seedling vigor (Takagi et al., 2013), cold tolerance (Luo et al., 2018), cooked grain elongation (Arikit et al., 2019), and low phosphorus tolerance (Nishida et al., 2018).

In this study, QTL-seq (applied to 12 F_2 populations) and GWAS (applied to three diverse panels) were used to rapidly identify QTL for ML in rice. Candidate genes for the important QTL were investigated and two breeder-friendly molecular markers were developed for the loci co-located by GWAS and QTL-seq.

MATERIALS AND METHODS

Plant Materials

The rice variety "IR 145" with short mesocotyl (0.18 cm) was crossed with 12 accessions with long mesocotyl (ranging from 3.39 to 5.13 cm) (**Table 1** and **Figure 1**) to develop the F_2 populations for QTL-seq. These populations were named as Pop1 to Pop12, respectively. Two diverse panels, XI-1A (147 accessions) and AUS (171 accessions) of the 3K Resequencing Project (Wang et al., 2018), were used for GWAS (**Supplementary Tables 1, 2**). Most of the accessions in XI-1A originated from China (**Supplementary Table 1**), whereas accessions in the AUS group mainly came from Bangladesh, India, and Pakistan (**Supplementary Table 2**). One diverse panel consisting of 140 accessions originated from the XI-1B of 3K Resequencing Project (**Supplementary Table 3**), mainly from China, which were used to validate the effectiveness of the markers developed based on QTL-seq and GWAS (Wang et al., 2018).

Phenotyping of ML

To evaluate ML, high-quality seeds were sown in a plastic tray containing 6-cm-deep soil according to Liu et al. (2020). The plastic tray has the following specifications: 50 holes, each with a size of 9.5 cm in depth, 4.5 cm in top diameter, and 2.1 cm in bottom diameter. After sowing, the seeds were covered with nutrient soil until the hole was filled up. Then, the tray was placed in a plastic pallet with 3-cm-deep nutrient soil and the system was kept in a dark incubator (30°C/65% relative humidity). The soil in each pallet was kept water saturated for seed germination and seedling growth. Three days after all of the seeds germinated (about 7–10 days after sowing), the seedlings from each hole were carefully excavated and rinsed with double-distilled water (ddH_2O) for ML measurement with ImageJ software[2]. For QTL-seq, a total of 720 individuals of each population were sown. Meanwhile, two parents of the corresponding population were also planted for ML measurement and sampling. For GWAS, 15

[2]https://imagej.en.softonic.com/

TABLE 1 | Details of the male parents of the 12 F_2 populations.

Population name	Male parent name	Mesocotyl length (cm)	Origin
Pop1	79	5.13	Pakistan (XI-1A)
Pop2	Bamla Suffaid 32	4.80	– (XI-1A)
Pop3	M 136-20	4.67	India (XI-1A)
Pop4	Cash	4.75	Bangladesh (AUS)
Pop5	Balam 2	4.83	Bangladesh (AUS)
Pop6	Kalasu	4.20	Philippines (XI-1A)
Pop7	Black 28-573	3.39	Bangladesh (AUS)
Pop8	Bhahuri	4.18	Bangladesh (AUS)
Pop9	Changai	4.74	Bangladesh (AUS)
Pop10	Basmati 385	4.68	India (XI-1A)
Pop11	IR64	3.48	Philippines (AUS)
Pop12	BR11	4.24	Bangladesh (AUS)

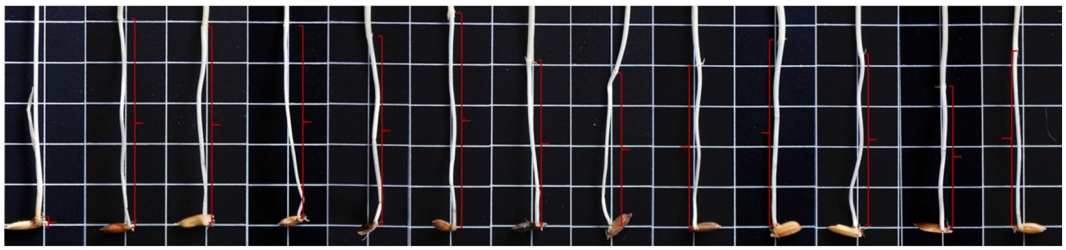


FIGURE 1 | Mesocotyl (*red brackets*) length of "IR 145" and the 12 male parents of the F_2 populations. The 13 individuals were "IR 145" (0.18 cm) and the corresponding male parents of Pop1 to Pop12 (from *left* to *right*) were as follows: 79 (5.13 cm), Bamla Suffaid 32 (4.80 cm), M 136–20 (4.67 cm), Cash (4.75 cm), Balam 2 (4.83 cm), Kalasu (4.20 cm), Black 28–573 (3.39 cm), Bhahuri (4.18 cm), Changai (4.74 cm), Basmati 385 (4.68 cm), IR 64 (3.48 cm), and BR 11 (4.24 cm).

seeds from each accession were used to evaluate the ML, and mean of two replications was used for statistical analysis.

Genotyping

For QTL-seq, the genomic DNA from seedling leaf was isolated according to the modified cetyl trimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). For a specific F_2 population, DNA samples with equal amounts from 50 individuals with extremely long mesocotyl were selected to generate the long mesocotyl (LM) bulk, whereas those from 50 individuals with extremely short mesocotyl were used to form the short mesocotyl (SM) bulk. The "IR 145" library and 12 libraries from the male parents, as well as 24 extreme libraries, were sequenced using the Illumina HiSeq 2500 platform (Illumina, Inc., San Diego, CA, United States) by Berry Genomics Corporation, Beijing, China[3]. The paired-end read data (PE150) with a sequencing depth of approximately 50× of the rice genome (∼400 Mb) were generated. The clean reads were obtained by removing reads with adaptors, with a ratio of *N* larger than 10%, and those of low quality (the number of bases with $Q \leq 3$ accounts for more than 50% of the whole reads).

The clean reads were aligned to Nipponbare RefSeq (IRGSP-1.0)[4] using BWA-MEM (release 0.7.10) (Li and Durbin, 2009). Then, the mapped reads were sorted and the duplicate ones were removed by Picard tools[5]. The variants for each accession were called for single nucleotide polymorphism (SNP) detection and annotation by GATK Best Practices (release 3.2-2;[6]). For the GWAS panel, the genotype data for XI-1A, AUS, and XI-1B were obtained from 3K Resequencing Project with Nipponbare IRGSP-1.0 as the reference genome and at about 11.5× coverage with an average mapping coverage of 92% (Wang et al., 2018).

QTL-seq Analysis

Reads from the LM bulk and SM bulk of an F_2 population were aligned to the "IR 145" variants, respectively. The SNP-index was calculated at each SNP site according to Takagi et al. (2013). The polymorphic sites were then filtered according to the following criteria: (1) the SNP-index values in both bulks were < 0.2 and the SNP sequencing depths were < 7; (2) the SNP-index in either bulk was missing; and (3) GQ < 20 in either bulk. For both bulks, the average SNP-index in each chromosome was presented using a sliding window approach with a 200-kb window size to generate the SNP-index plots for all chromosomes. The Δ(SNP-index) was then calculated by subtracting the SNP-index of LM bulk with that of the SM bulk. The regions in which the average Δ(SNP-index) at a locus was significantly greater than the surrounding regions at 95% confidence interval were considered as candidate regions (Takagi et al., 2013).

Population Structure and LD Decay Analysis

The population structures for the AUS and XI-1A panels were analyzed using 10,000 polymorphic SNP markers with Admixture 1.3.0 (Alexander et al., 2009). Five independent runs for each *K* value from 2 to 7 were performed based on an admixture model. An *ad hoc* quantity statistic, Δ*K*, based on the rate of change in log probability of data between successive *K* values was used to predict the real number of subpopulations. Principal component analysis (PCA) and neighbor-joining (NJ) trees for AUS, XI-1A, and XI-1B were also used to validate population stratification with Tassel v5.0 (Bradbury et al., 2007). The LD among markers was calculated using the full matrix and sliding window options in Tassel v5.0 with 10,000 evenly distributed SNP markers. The squared allele frequency correlation, r^2, values were plotted against physical distance and a LOESS (locally weighted smoothing) curve was fitted to the plot to show the association between LD decay and physical map distance. The intersection of the fitted curve of r^2 values with threshold of 95th percentile in the distribution of r^2 was considered as the estimate of the LD range (Breseghello and Sorrells, 2006).

Genome-Wide Association Analysis

Previous studies have reported the details of the genotype data, LD decay, PCA, and the population structure for the XI-1A and AUS panels (Wang et al., 2018). Associations between the genotypic and phenotypic data were analyzed using the kinship matrix in an MLM (mixed linear model) by GAPIT[7] based on

[3]https://www.berrygenomics.com/

[4]http://rice.plantbiology.msu.edu/

[5]http://broadinstitute.github.io/picard/

[6]https://gatk.broadinstitute.org/hc/en-us

[7]http://www.zzlab.net/GAPIT/

R 3.6.1 (Lipka et al., 2012) to control background variations and eliminate spurious MTAs. Since the Bonferroni–Holm correction for multiple testing ($\alpha = 0.05$) was too conserved, markers with an adjusted $-\log_{10}(P\text{-value}) \geq 4.0$ were regarded as the significant ones.

Candidate Gene Identification and Gene Expression Analysis

Candidate genes for the loci consistently identified in two or more populations were identified. The following steps were conducted to identify the candidate genes for important QTL. Firstly, excavate all the genes located in the LD block region around the peak SNP (±150 kb based on previous LD decay analysis) of each important QTL from the MSU Rice Genome Annotation Project[8]. Then, all available SNPs located inside of these genes were searched. The genes (except for the expressed protein, hypothetical protein, transposon protein, and retrotransposon protein) with SNPs in the coding region that could further lead to sense mutations were considered as candidate genes. A candidate gene with identical SNPs or InDels in the male parent of the co-localized population was further selected. As mesocotyl elongation is highly regulated by various phytohormones, including strigolactones (SLs), cytokinins (CTKs), brassinosteroids (BRs), abscisic acid (ABA), jasmonates (JAs), gibberellins (GAs), and auxins (IAAs) (Watanabe and Takahashi, 1999; Cao et al., 2005; Hu et al., 2014; Xiong et al., 2017; Sun et al., 2018; Zheng et al., 2020; Lv et al., 2021), those genes involved in phytohormone metabolism were regarded as high-confidence candidate genes for mesocotyl elongation.

Quantitative real-time PCR (qRT-PCR) was conducted to test expression differences of the candidate genes in parents of the corresponding F_2 population. The mesocotyl section of "IR 145" and the male parent were sampled for RNA extraction at 52 h after germination before the coleoptile was unearthed. Total RNA was extracted according to the Trizol method. Complementary DNA (cDNA) was synthesized with the HiScript II 1st Strand cDNA Synthesis Kit (Vazyme, Nanjing, China) and then diluted 5–10 times with sterile double distilled water. The primers were designed with Primer

[8]http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/

TABLE 2 | QTL-seq for mesocotyl length in 12 F_2 populations.

Name	Chromosome	Start (Mb)	End (Mb)	Length (Mb)	Position of max. *G*-value (Mb)	Position of max. Δ(SNP-index) (Mb)	Population	References
qML1.1	1	6.57	8.12	1.55	7.21	7.32	5	–
qML1.2	1	14.21	17.32	3.11	14.22	14.51	4	Lu et al., 2016
qML1.3	1	18.59	20.15	1.56	–	19.93	3	Ouyang et al., 2005
	1	18.51	20.36	1.85	20.01	20.12	8	
	1	19.88	21.12	1.24	–	19.83	6	
	1	20.04	21.12	1.08	20.34	20.34	2	
qML1.4	1	36.59	39.01	2.42	37.92	38.57	12	Xiong et al., 2017
	1	38.32	39.67	1.35	39.26	38.39	4	
	1	37.58	39.30	1.82	38.05	38.05	9	
qML3.1	3	25.12	27.5	2.38	25.24	26.53	3	–
qML3.2	3	27.21	30.13	2.92	29.50	29.84	8	Zhao et al., 2018
	3	29.51	32.3	2.79	31.32	31.36	12	
qML3.3	3	35.72	36.15	0.43	35.86	36.04	9	–
qML4.1	4	19.59	21.92	2.33	21.37	21.25	10	–
qML4.2	4	25.50	27.13	1.63	25.58	25.92	8	–
qML5.1	5	8.99	11.03	2.04	10.16	10.21	8	–
qML6.1	6	2.59	5.12	2.53	4.99	3.87	8	–
qML7.1	7	4.98	8.52	3.54	7.90	8.84	4	–
	7	4.35	7.56	3.21	5.63	6.03	9	–
	7	5.52	8.65	3.13	7.31	7.92	6	–
qML7.2	7	13.69	16.25	2.56	15.82	15.88	11	Ouyang et al., 2005; Zhao et al., 2018; Liu et al., 2020
	7	14.99	17.12	2.13	15.85	17.13	12	
	7	14.89	16.52	1.63	16.37	16.12	8	
	7	15.45	16.98	1.53	15.84	15.03	9	
	7	15.58	18.24	2.66	17.23	16.92	6	
	7	16.37	18.49	2.12	17.68	17.21	4	
qML9.1	9	11.53	13.54	2.01	12.29	12.25	12	–
	9	11.99	15.28	3.29	15.13	13.56	5	–

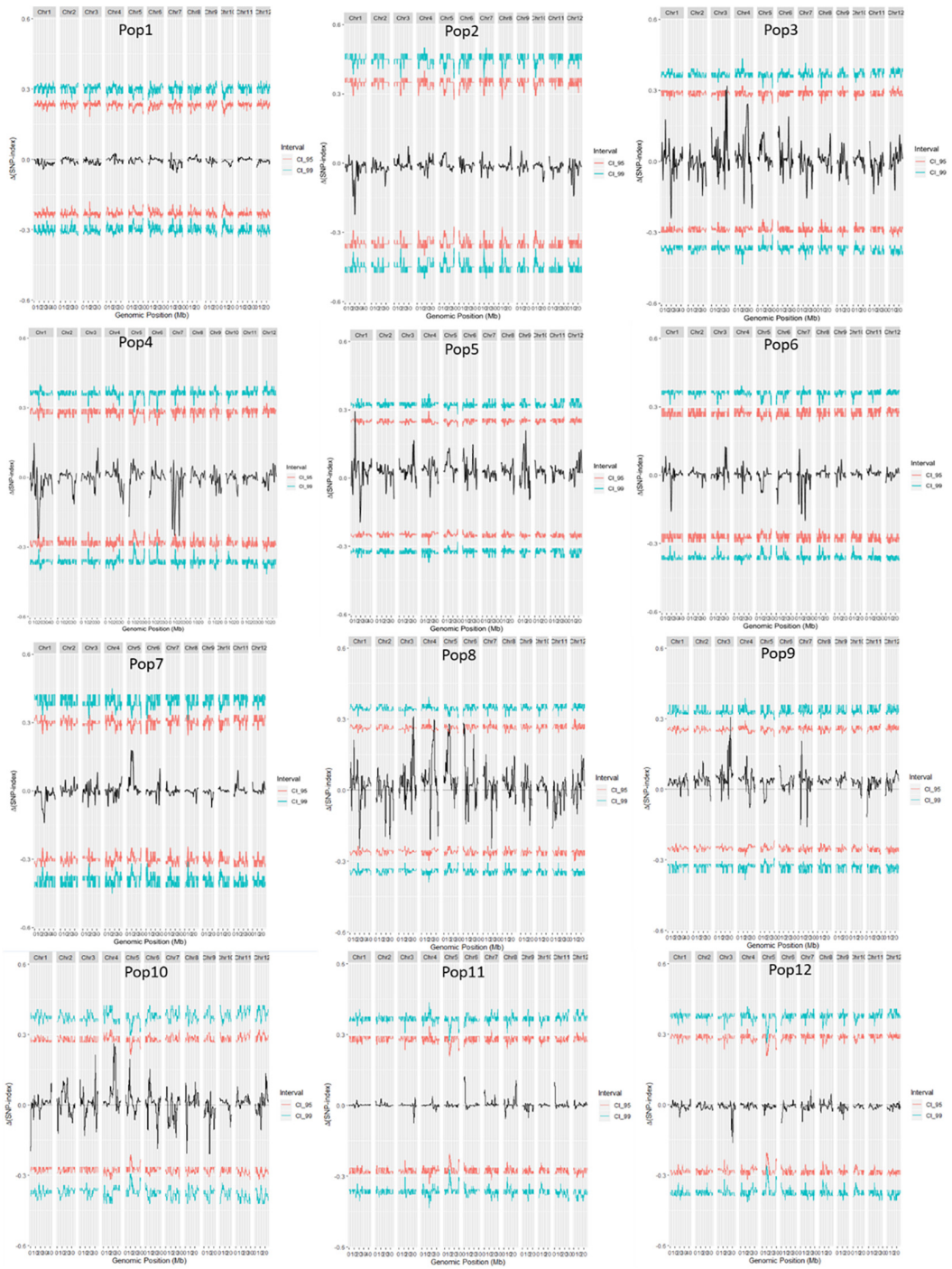


FIGURE 2 | Δ(SNP-index) of 12 F_2 populations. The female parents of the 12 F_2 populations were: 79 (Pop1), Bamla Suffaid 32 (Pop2), M 136–20 (Pop3), Cash (Pop4), Balam 2 (Pop5), Kalasu (Pop6), Black 28–573 (Pop7), Bhahuri Kalasu (Pop8), Changai Kalasu (Pop9), Basmati 385 Kalasu (Pop10), IR 64 Kalasu (Pop11), and BR 11 Kalasu (Pop12).

Premier 5.0 software[9]. PCR procedure was conducted in a volume of 20 μl, containing 2 μl cDNA, 0.4 μl of each primer (in micromolars), and 10 μl ChamQ Universal SYBR qPCR Master Mix. The reaction was conducted in the ABI StepOnePlus Real-Time PCR System with Tower (ABI, Waltham, MA, United States). The gene expression level was analyzed with $2^{-\Delta\Delta CT}$ method. *OsActin1* was used as internal control to normalize the expression levels of different samples. All assays

[9]http://www.premierbiosoft.com/

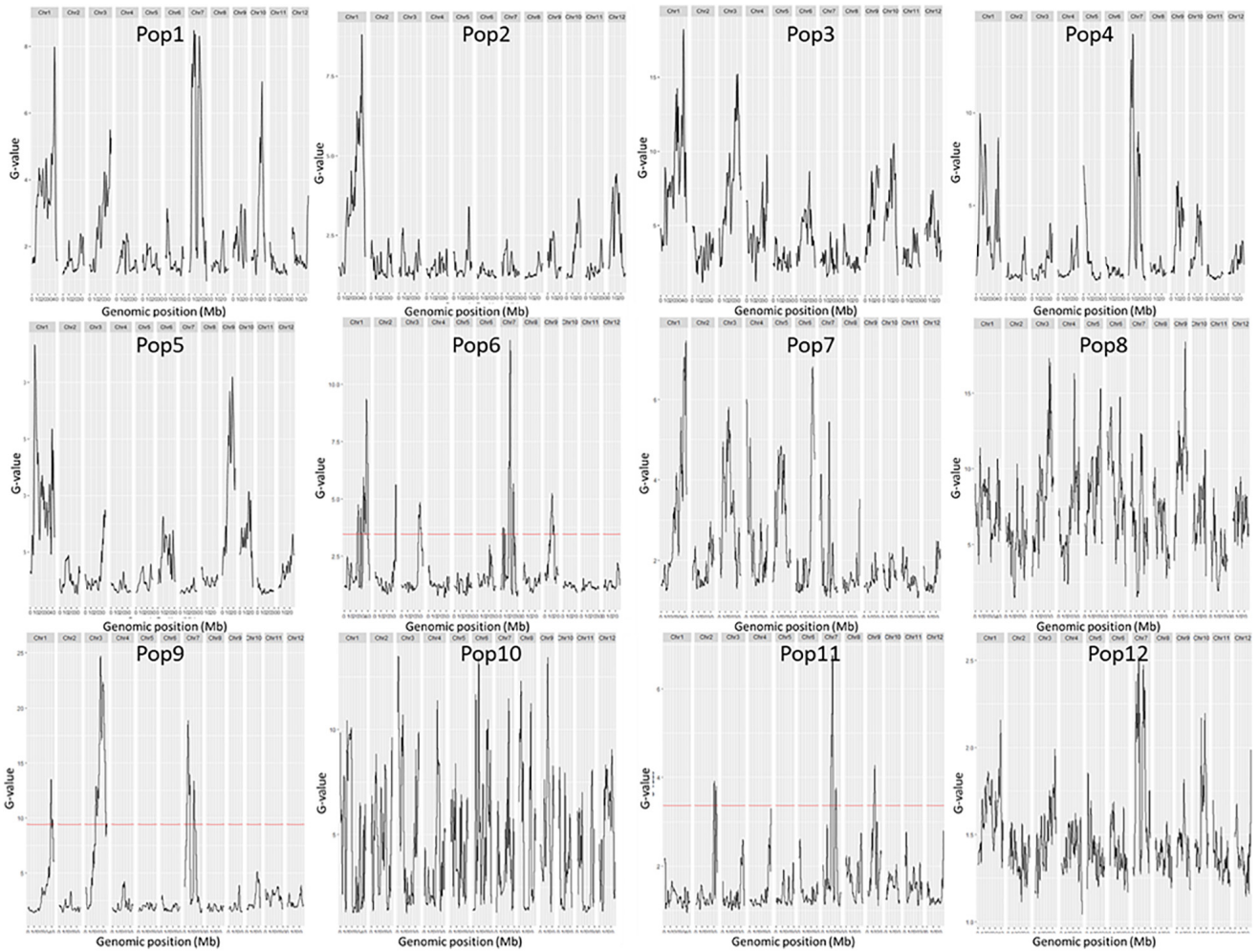


FIGURE 3 | *G*-values of the 12 F_2 populations. The female parents of the 12 F_2 populations were: 79 (Pop1), Bamla Suffaid 32 (Pop2), M 136–20 (Pop3), Cash (Pop4), Balam 2 (Pop5), Kalasu (Pop6), Black 28–573 (Pop7), Bhahuri Kalasu (Pop8), Changai Kalasu (Pop9), Basmati 385 Kalasu (Pop10), IR 64 Kalasu (Pop11), and BR 11 Kalasu (Pop12).

were performed in two independent experiments, each with three repetitions.

Development of Allele-Specific Markers

For InDel markers, the primers were designed by Primer Premier 5 software (Lalitha, 2000), with a pair of primers spanning the InDel region and the amplified fragment size was set to no more than 10 times that of the InDel. A PCR procedure was implemented for the two markers. The PCR procedure was conducted in a volume of 25.0 μl, which includes 12.5 μl of 2× Taq Master Mix (Vazyme, Nanjing, China), 2 μl of template DNA, 1 μl of each primer (10 μM), and 8.5 μl of ddH_2O. The PCR program was set as follows: an initial denaturation at 94°C for 5 min, 35 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 30 s, and a final extension at 72°C for 10 min.

Statistical Analysis

One-way analysis of variance (ANOVA) was performed using SPSS Statistics 17.0[10]. Tukey's multiple test was employed for multiple comparisons ($^*P < 0.05$ and $^{**}P < 0.01$).

RESULTS

ML Exhibited Continuous Variation in the F_2 Populations and GWAS Panels

ML of the common parental line "IR 145" was 0.18 cm, while the other 12 parental lines had MLs ranging from 3.39 to 5.13 cm (**Table 1** and **Figure 1**). The average MLs of 12 F_2 populations were: 0.85 (Pop1, 0–5.03 cm), 0.75 (Pop2, 0–4.61 cm), 0.78 (Pop3, 0–4.98 cm), 0.91 (Pop4, 0–5.71 cm), 0.72 (Pop5, 0–4.31 cm), 0.69 (Pop6, 0–5.32 cm), 0.78 (Pop7, 0–4.98 cm), 0.83

[10]http://www-01.ibm.com/software/analytics/spss/

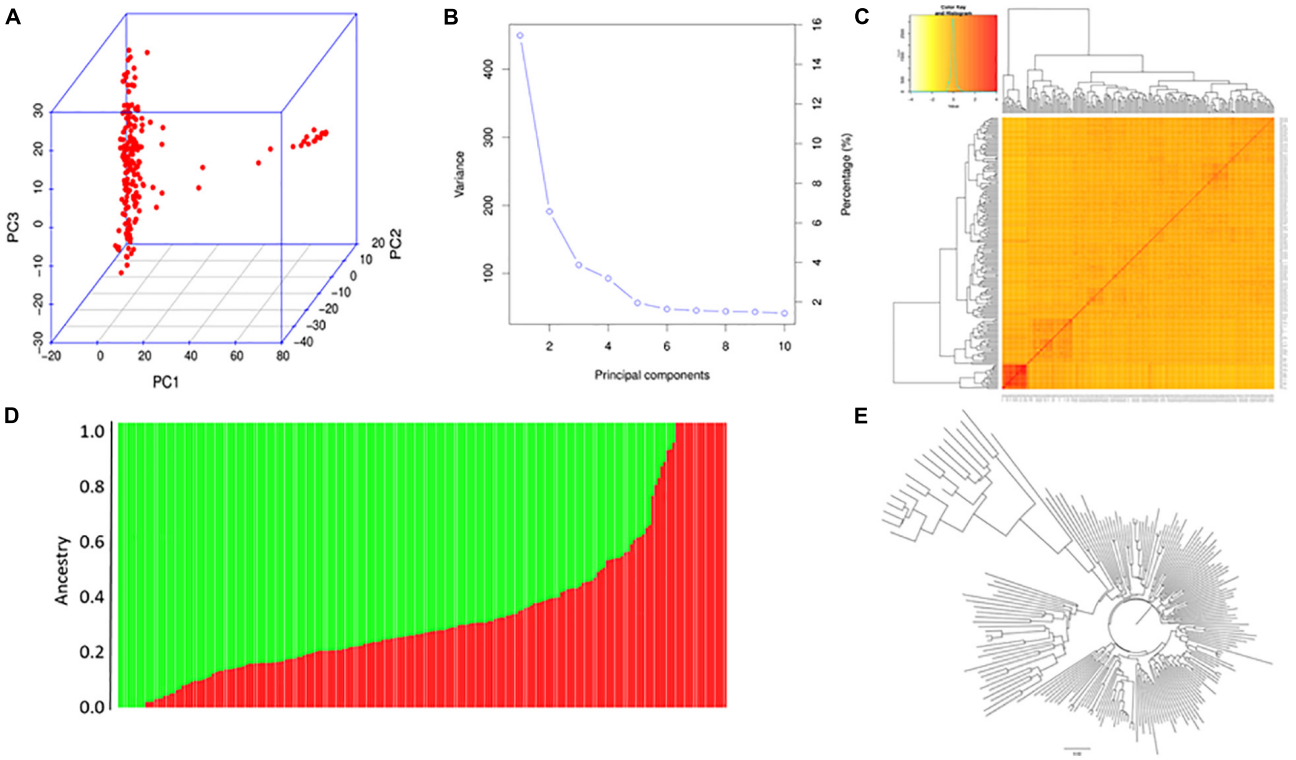


FIGURE 4 | Population analysis for the XI-1A panel. **(A)** Principal component analysis (PCA) plots. **(B)** Total variation explained by the principal components (PCs). **(C)** Kinship relatedness. **(D)** Two subgroups inferred by structure analysis. **(E)** Neighbor-joining (NJ) tree.

(Pop8, 0–5.25 cm), 0.81 (Pop9, 0–4.91 cm), 0.70 (Pop10, 0–5.22 cm), 0.95 (Pop11, 0.2–5.13 cm), and 0.84 (Pop12, 0–5.93 cm) (**Supplementary Figure 1**). The MLs ranged from 0.0 to 3.41 cm, with an average of 0.85 cm in the XI-1A panel (**Supplementary Table 1** and **Supplementary Figure 2**), whereas it ranged from 0.2 to 4.42 cm with an average of 2.41 cm in the AUS panel (**Supplementary Table 2** and **Supplementary Figure 2**). Continuous variation with transgressive segregation on both sides was observed across both populations with approximately normal distributions (**Supplementary Figures 1, 2**).

QTL Identified by QTL-seq

In total, 85–112 Gb clean reads were obtained, and all the Q30 reached 85%. The average sequencing depth was 46.83×, and the average mapped ratio and genome coverage have reached 96.5 and 95.4%, respectively (**Supplementary Table 3**). A total of 5,128,693–6,984,215 SNPs were detected in 12 F_2 populations (**Supplementary Table 4**). The marker density ranged from 14.2 markers/kb (chromosome 3 of Pop11) to 22.8 markers/kb (chromosome 7 of Pop1), with an average of 18.2 markers/kb.

A total of 14 regions were identified by the Δ(SNP-index) value or the *G*-value (**Table 2** and **Figures 2**, **3**). These loci were located on chromosomes 1, 3–7, and 9, and the interval size ranged from 0.43 to 3.54 Mb (**Table 2**). Four unique QTL for ML—*qML1.1* (Pop5, 6.57–8.12 Mb), *qML1.2* (Pop4, 14.21–17.32 Mb), *qML1.3* (Pop2, Pop3, Pop6, and Pop8, 18.59–21.12 Mb), and *qML1.4* (Pop4 and Pop 12, 36.59–39.67 Mb)—were identified on chromosome 1. Three adjacent QTL were isolated on chromosome 3. Of these, *qML3.1* was detected in Pop3 and located at the interval of 25.12–27.50 Mb, whereas *qML3.2* was discovered in Pop8 and Pop12 and fixed at 27.21–32.30 Mb. *qML3.3*, unearthed in Pop9, was set at the interval of 35.72–36.15 Mb. Two loci were found to exist on chromosome 4. Of these, *qML4.1* was recognized at Pop10 and situated at the interval of 19.59–21.92 Mb, whereas *qML4.2*, fixed at 25.50–27.13 Mb, was discovered at Pop8. On chromosome 5, only *qML5.1* was identified from Pop8 and located at the 8.99- to 11.03-Mb region. Another genomic region (2.59–5.12 Mb) identified on chromosome 6 with Δ(SNP-index) plots greater than the statistical confidence intervals ($P < 0.05$) was named as *qML6.1*. Two adjacent QTL (*qML7.1* and *qML1.2*) were identified on chromosome 7. Besides, *qML7.1* (4.35–8.65 Mb) was identified in Pop4, Pop6, and Pop9, whereas *qML7.2* (13.69–18.49 Mb) was identified in Pop4, Pop6, Pop8, Pop9, Pop11, and Pop12. In addition, only one locus, *qML9.1* (11.53–15.28 Mb), was identified on chromosome 9 in Pop5 and Pop12.

QTL Identified by GWAS

In total, 2,338,386 SNPs were left and employed for GWAS. The chromosome size varied from 22.8 Mb for chromosome 9 to

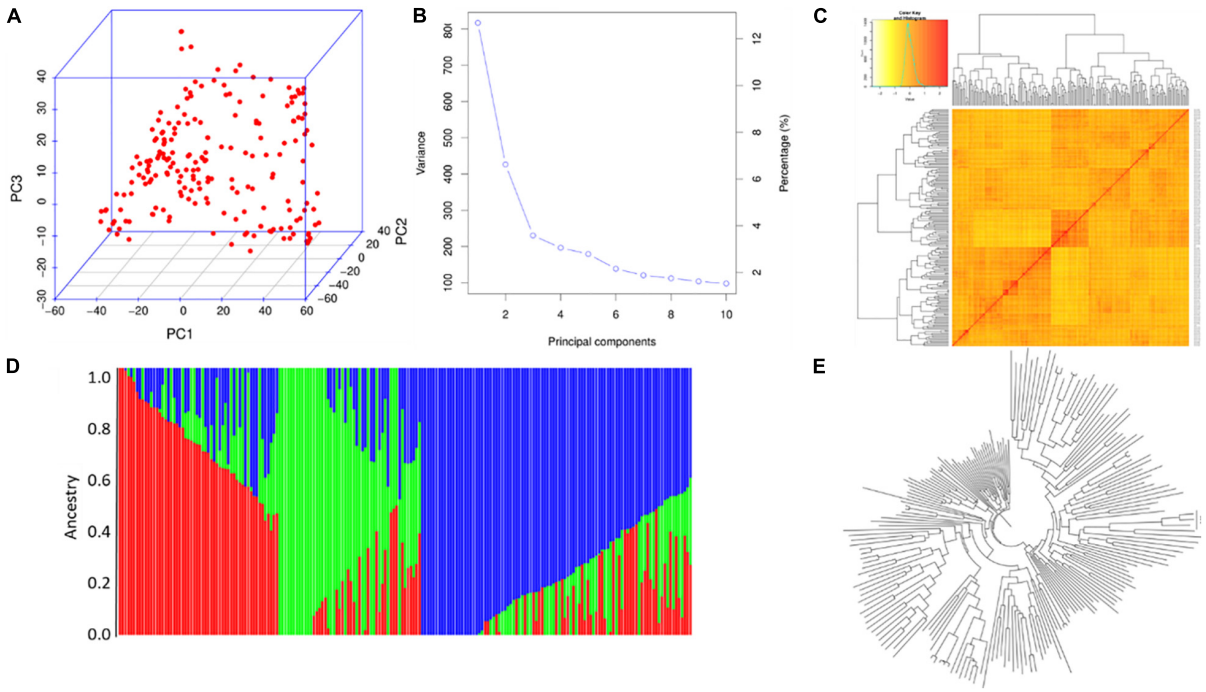


FIGURE 5 | Population analysis for the AUS panel. **(A)** Principal component analysis (PCA) plots. **(B)** Total variation explained by the principal components (PCs). **(C)** Kinship relatedness. **(D)** Three subgroups inferred by structure analysis. **(E)** Neighbor-joining (NJ) tree.

43.2 Mb for chromosome 1. These markers spanned a physical distance of 373 Mb, with an average density of 6.25 markers/kb.

PCA of XI-1A panel indicated that the top three principal components (PCs) could explain 15.8, 6.2, and 4.1% of total variation (**Figures 4A,B**), respectively, and this panel consists of two subgroups (**Figure 4C**). The NJ tree showed that the two clades represented two subpopulations (**Figure 4E**). Structure analysis indicated that XI-1A could be divided into two subgroups, subgroup 1 and subgroup 2 (**Figure 4D**), whose characterizations were largely consistent with their geographic origins. The XI-1A-1 accessions were mainly from South China, whereas those of XI-1A-2 were mainly from the Yangtze River plain, China. Also, admixture accessions were observed in the present study (**Figure 4**).

For the AUS panel, PCA indicated that the top three PCs could explain 13.5, 6.5, and 4.2% of the total variation, respectively (**Figure 5A,B**). Structure analysis indicated that the AUS panel could be divided into three subgroups, AUS-1, AUS-2, and AUS-3 (**Figure 5C,D**), whose characterizations were largely consistent with their geographic origins and the results of PCA (**Figure 5A**) and NJ tree analysis (**Figure 5E**), which showed three clades in this panel. AUS-1, AUS-2, and AUS-3 mainly include accessions from Bangladesh, India, and Pakistan, respectively. The LD decay along the physical distances for the XI-1A and AUS panels are shown in **Supplementary Figure 3**; the corresponding LD decay distance was about 150 and 175 kb, respectively.

Manhattan plots for the markers significantly associated with ML are shown in **Figure 6**. For the XI-1A panel, three unique loci located on chromosomes 8 (*qML-XI-1A-8.1*, 8.98–9.38 Mb), 9 (*qML-XI-1A-9.1*, 14.52–14.58 Mb), and 12 (*qML-XI-1A-12.1*, 5.58–5.76 Mb) were detected, which explained 5.5–11.3% of the ML variance (**Table 3** and **Supplementary Table 5**). Notably, *qML-XI-1A-9.1* overlapped with the *qML9.1* (11.53–15.28 Mb) identified in Pop5 and Pop12 (**Table 2**). In the AUS panel, two unique loci located on chromosomes 1 (*qML-AUS-1.1*, 16.03–17.82 Mb) and 12 (*qML-AUS-12.1*, 18.29–18.50 Mb) were detected, which explained ML variations of 5.3 and 14.6%, respectively (**Table 3** and **Supplementary Table 6**). Among which, the *qML-AUS-1.1* locus overlapped with the *qML1.2* (14.21–17.32 Mb) identified in Pop4 (**Table 2** and **Table 3**).

Candidate Genes in the Important QTL Regions

The genes located in the LD block region around the peak SNP (±150 kb based on previous LD decay analysis) of each important QTL were excavated from the MSU Rice Genome Annotation Project[8]. Then, all available SNPs located inside of these genes were searched. Fifty-seven genes (except for the expressed protein, hypothetical protein, transposon protein, and retrotransposon protein) with SNPs in the coding region that lead to sense mutations were considered as candidate genes (**Supplementary Table 7**). A candidate gene with identical SNPs or InDels in the male parents of the 12 F_2 populations showing overlapping QTL was further selected. In total, 12 genes were screened by this method.

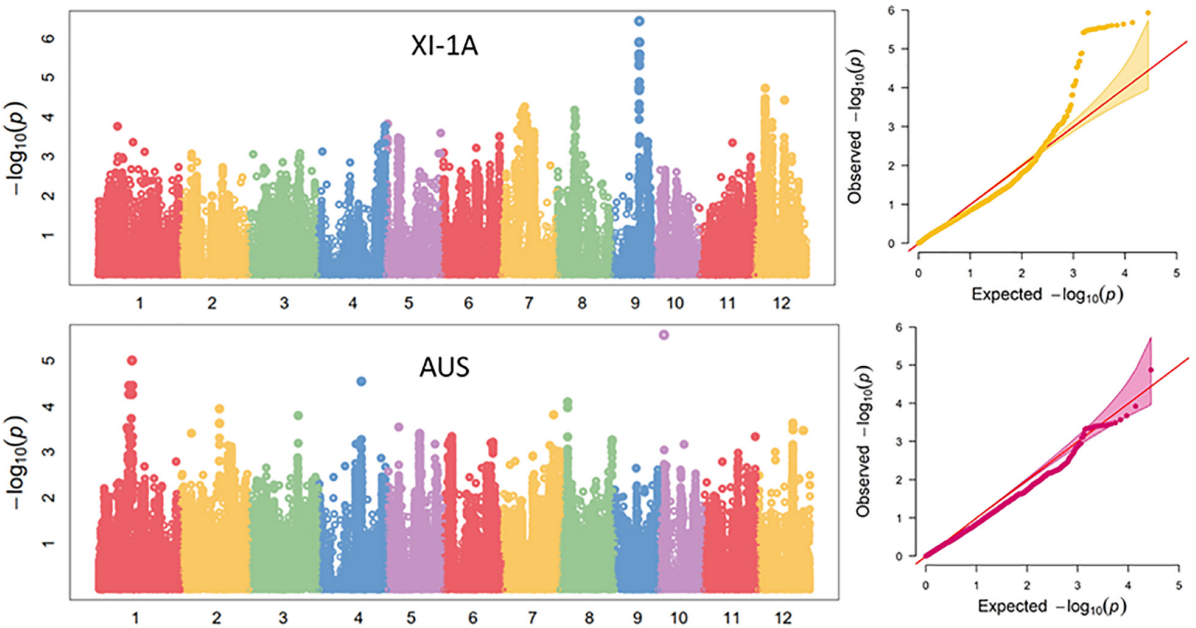


FIGURE 6 | Genome-wide association study (GWAS) for mesocotyl length in the XI-1A and AUS panels.

TABLE 3 | GWAS for mesocotyl length in the XI-1A and AUS panels.

Population	Locus name	Chromosome	Peak	Interval (Mb)	Peak	R^2 (%)	Others
			Position (bp)		*P*-value		
XI-1A	*qML-XI-1A-8.1*	8	8,982,811	8.98–9.38	2.81E-05	6.69	
	qML-XI-1A-9.1	9	14,572,632	14.52–14.58	7.39E-07	11.27	*qML9.1*
	qML-XI-1A-12.1	12	5,716,291	5.58–5.76	0.000025	6.75	
AUS	*qML-AUS-1.1*	1	17,551,160	16.03–17.82	9.82E-06	14.57	Lu et al., 2016; *qML1.2*
	qML-AUS-12.1	12	18,365,206	18.29–18.50	2.28E-04	6.52	Rohilla et al., 2020

The expressions of 12 candidate genes in the parents of the corresponding F_2 populations were detected using qRT-PCR (**Figure 7** and **Supplementary Table 8**). Four genes, *LOC_Os01g36580*, *LOC_Os01g67670*, *LOC_Os04g44240*, and *LOC_Os07g08540*, showed no significant differences between the extreme ML parental accessions. Four genes—*LOC_Os01g13200*, *LOC_Os06g09660*, *LOC_Os07g13634*, and *LOC_Os09g20350*—showed more than 2.0- to 4.9-fold higher expressions in the long mesocotyl accessions compared to the short ones. Four genes, namely, *LOC_Os01g66100*, *LOC_Os03g56060*, *LOC_Os04g33360*, and *LOC_Os07g28060*, showed more than 1.9- to 7.6-fold lower expressions in the long mesocotyl accessions (**Figure 7**).

The eight genes showing obvious differences involved in the biological metabolism of phytohormones, cell elongation, and division were selected as the high-confidence candidate genes for mesocotyl elongation (**Table 4**). *LOC_Os01g13200* for *qML1.1* related to abscisic acid insensitive 8, *LOC_Os01g66100* from *qML1.4* encodes gibberellin oxidase, *LOC_Os03g56060* for *qML3.2* is a member of the cellulose synthase-like family, *LOC_Os04g33360* for *qML4.1* encodes gibberellin 2-beta-dioxygenase, *LOC_Os06g09660* for *qML6.1* is an auxin response factor, *LOC_Os07g13634* for *qML7.1* (7,815,442 bp) encodes cytokinin-*N*-glucosyltransferase 1, *LOC_Os07g28060* for *qML7.2* is an ethylene receptor, and *LOC_Os09g20350* for *qML9.2* is an ethylene-responsive transcription factor.

Marker Development and Validation

For the loci (*qML1.3* and *qML7.2*) identified by both the QTL-seq and GWAS, a 22-bp insertion in the 18,932,318 bp for *qML1.3* at chromosome 1 and a 30-bp insertion in the 15,404,166 bp for *qML7.2* at chromosome 7 were selected for marker development to validate their effects. DNA fragments larger than 10 bp can be easily identified by agarose gel electrophoresis. To facilitate breeders utilizing the polymorphic information, we attempted to transfer all the InDels into PCR gel-based markers. The two molecular markers for *qML1.3* and *qML7.2* were named *Indel-Chr1:18932318* and *Indel-Chr7:15404166*, respectively (**Table 5**, **Supplementary Table 9**, and **Figure 8**).

X1-1B, with no obvious population structure (**Supplementary Figure 4**), was used to validate the effectiveness of the two markers. *Indel-Chr1:18932318* could divide the 140 accessions into two groups: the amplified fragment of group 1 was 275 bp, whereas that of group 2 was 297 bp (**Figure 8A**).

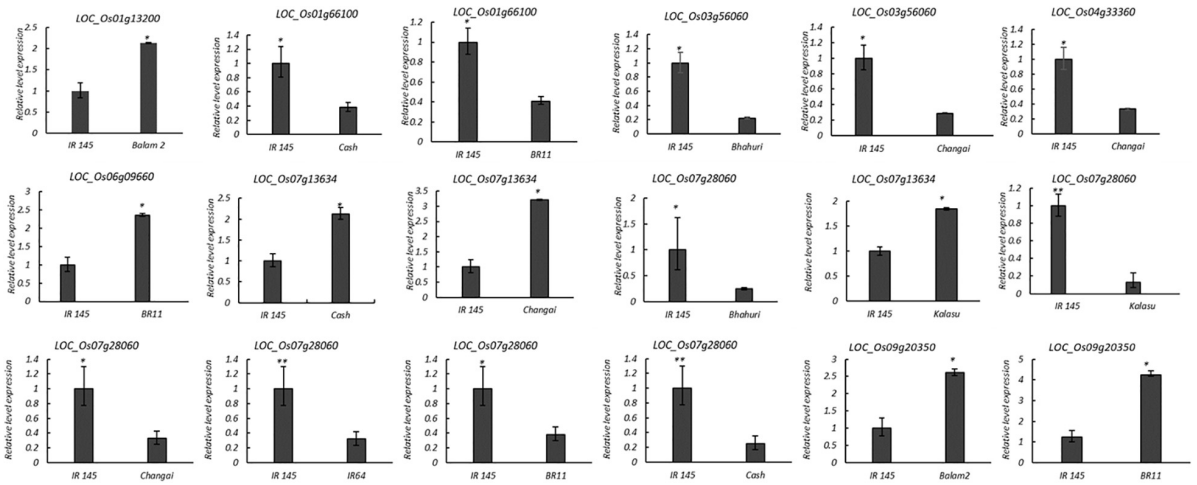


FIGURE 7 | Expression difference of the candidate genes for mesocotyl elongation by qRT-PCR. The *X*-axis represents the two parents of the F_2 population, whereas the *Y*-axis represents the relative expression of each gene to β-*actin* in mesocotyl tissue. Relative transcription levels were calculated by the $2^{-\Delta\Delta CT}$ method. Significance at $^*P \leq 0.05$ and $^{**}P \leq 0.01$.

TABLE 4 | Candidate genes for mesocotyl length by QTL-seq and GWAS.

Loci	Candidate gene	Start (bp)	End (bp)	Functional annotation
qML1.1	*LOC_Os01g13200*	7,352,175	7,357,620	Abscisic acid insensitive 8
qML1.4	*LOC_Os01g66100*	38,382,382	38,385,504	Gibberellin 20 oxidase 2
qML3.2	*LOC_Os03g56060*	31,930,541	31,934,624	CSLC9—cellulose synthase-like family C
qML4.1	*LOC_Os04g33360*	20,200,072	20,201,885	Gibberellin 2-beta-dioxygenase 7
qML6.1	*LOC_Os06g09660*	4,926,492	4,932,177	Auxin response factor
qML7.1	*LOC_Os07g13634*	7,815,442	7,832,311	Cytokinin-*N*-glucosyltransferase 1
qML7.2	*LOC_Os07g28060*	16,368,462	16,368,840	Ethylene receptor
qML9.2	*LOC_Os09g20350*	12,216,432	12,218,346	Ethylene-responsive transcription factor

TABLE 5 | Primer information on the two developed markers for *qML1.3* and *qML7.2*.

Primer names	Primer sequence (5′–3′)	Amplicon size (bp)
InDel-Chr1:18932318-F	AGAACCCTTTTATCCTCATTA	275 (297)
InDel-Chr1:18932318-R	ACAAAGGGACTTGATGATGG	
InDel-Chr7:15404166-F	CACTAGCAAGAGTGCTCCCA	230 (200)
InDel-Chr7: 15404166-R	TTCTCAATACCCATGCCAAC	

The amplicon size of 'IR 145' with the two markers were 275 bp and 230 bp respectively; whereas that of the female parent of Pop6, Kalasu, were 297 bp and 200 bp, respectively.

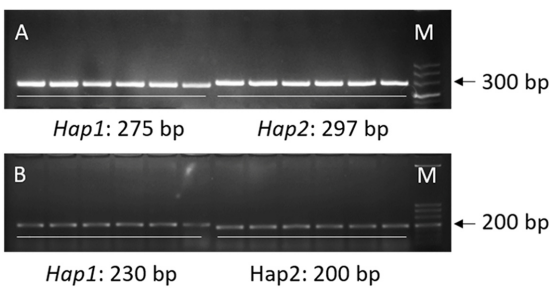


FIGURE 8 | Two PCR gel-based markers based on the 22-bp InDel at chromosome 1 and the 30-bp InDel at chromosome 7. **(A)** Marker *Indel-Chr1:18932318*. **(B)** Marker *Indel-Chr7:15404166*.

Similarly, *Indel-Chr7:15404166* could divide the 140 accessions into two groups, with the amplified fragments being 230 and 200 bp (**Figure 8B**). An association analysis between the genotypes and phenotypes was conducted with 140 accessions from XI-1B of the 3K Resequencing Project, and significant differences in ML were detected. For *Indel-Chr1:18932318*, the mean ML of *Hap2* was significantly higher than that of *Hap1* by 0.42 cm ($P < 0.05$). For *Indel-Chr7:15404166*, the mean ML of *Hap2* was significantly higher than that of *Hap1* by 0.41 cm ($P < 0.05$) (**Figure 9**). ANOVA indicated that *Indel-Chr1:18932318*, *Indel-Chr7:15404166*, and the interaction between the two markers could explain ML variations at 2.68, 2.71, and 1.71%, respectively (**Table 6**).

DISCUSSION

Considerable variations were present in the two panels used for GWAS and the panel for validating the developed markers.

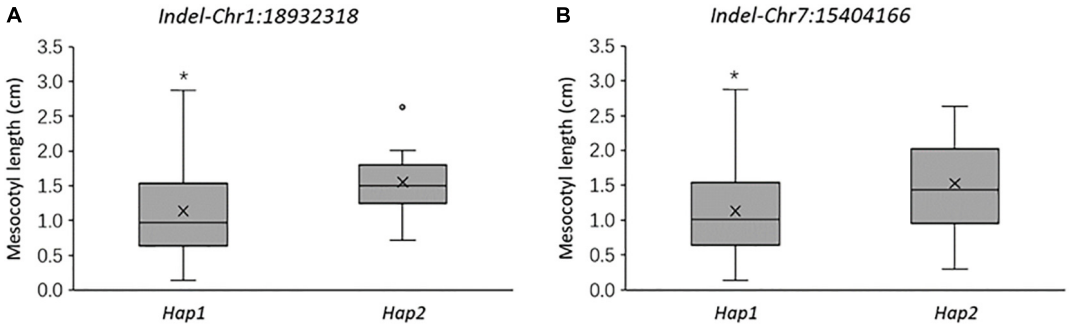


FIGURE 9 | Mesocotyl length comparison of two haplotypes at 18,932,318 bp on chromosome 1 **(A)** and at 15,404,166 bp on chromosome 7 **(B)** in 140 accessions from the XI-1B population. *$P < 0.05$ (significant differences between the alleles).

TABLE 6 | ANOVA for mesocotyl length in the XI-1B panel by two markers.

Marker	*df*	Sum of squares	Mean square	*F*-value	Pr(>*F*)
Indel-Chr1:18932318	1	1.178	1.776	2.984	0.0464
Indel-Chr7: 15404166	1	1.185	1.853	3.003	0.0354
(*Indel-Chr1:18932318*) × (*Indel-Chr7:15404166*)	2	0.737	0.737	1.867	0.174
Residuals	136	40.678	0.395		

The existence of a wide range of genetic variation has also been reported previously by many studies (Wu et al., 2005, 2015; Liu et al., 2020). Thus, identifying new loci and the associated markers is urgent and important for ML improvement in released cultivars, and breeding for long mesocotyl is feasible and has great potential. Although mesocotyl is a crucial rice developmental trait and is imminently required to improve crop adaptability to modern cultivation modes, QTL and genes are still not adequate, and more QTL and long mesocotyl donors need to be identified.

Traditional QTL mapping includes the construction of a mapping population and the phenotypic identification of a large number of individuals in the segregation population, which largely restricted the progress of QTL mapping (Salvi and Tuberosa, 2005). With the reduction of sequencing cost, QTL-seq has become an optimized strategy to rapidly identify the region harboring the genes/QTL of interest (Takagi et al., 2015). Besides, GWAS also provides a more representative gene pool and is an efficient tool that bypasses the time and expand to the developing population (Zhu et al., 2008). In total, 20 QTL for ML were identified in this study and indicated that the combination of GWAS and QTL-seq is an effective and reliable strategy for the rapid identification of major QTL for mesocotyl elongation.

In this study, 12 long mesocotyl accessions were selected as male parents to cross with the same female parent "IR 145," a typical variety with a short mesocotyl. Six of the 14 unique loci identified by QTL-seq were detected in two or more populations, namely, *qML1.3*, *qML1.4*, *qML3.2*, *qML7.1*, *qML7.2*, and *qML9.1*, whereas the other loci were identified in one population. Also, the parents of Pop4, Pop5, Pop6, Pop7, Pop8, Pop9, and Pop12 were originated from the AUS panel, whereas the parents of Pop1, Pop2, Pop3, Pop10, and Pop11 were originated from the XI-1A panel. The loci identified in GWAS of the AUS and XI-1A panel were also detected in QTL-seq. These stable loci could be utilized in the breeding of long mesocotyl rice. Thus, QTL-seq with multiple populations based on the same parent could reduce the sequencing cost effectively while preserving a higher genetic diversity to identify the shared or unique loci and, thus, is a cost-saving and effective approach for genetic analysis of complex traits.

Six of the 14 QTL identified in this study overlapped with those identified previously by linkage mapping or GWAS. The QTL *qML7.2* has been reported at least five times using different biparental mapping populations or GWAS using different panels (Ouyang et al., 2005; Zhao et al., 2018). The QTL on chromosome 1 (*qML1.2*, 14.21–17.32 Mb), identified in Pop4 and the XI-1A panel, coincided with the loci detected by Lu et al. (2016) (marker: *seq-rs303*). *qML1.3* (18.59–21.12 Mb) and *qML1.4* (36.59–39.67 Mb) overlapped with the loci discovered by Wu et al. (2015) and Xiong et al. (2017). The *qML3.2* locus (27.21–32.30 Mb) on chromosome 3 crossed with the *qML3-2* isolated by GWAS based on 621 cultivars originated from the 3K Rice Resequencing Project (Zhao et al., 2018). QTL on chromosomes 8 (8.98–9.38 Mb), identified in the XI-1A panel, located in the same region with the loci identified from the Zhaxima/Hanhui 3 population (Yuldashev et al., 2012). Rohilla et al. (2020) have identified 20 genes significantly associated with anaerobic germination in a diverse panel with 94 deep-water rice genotypes of Assam. Of these, two most relevant genes, *OsXDH1* (*LOC_Os03g31550*) and *SSXT* (*LOC_Os12g31350*), have been identified that explain the higher phenotypic variability ($R^2 > 20\%$). The locus on chromosome 12 identified by Rohilla et al. (2020) overlapped with the *qML-AUS-12.1* identified in the AUS panel.

According to the results of GWAS and QTL-seq, eight genes involved in the biological metabolism of plant hormones, cell elongation, and cell division were selected as high-confidence candidate genes. A candidate gene (*LOC_Os07g13634*) on chromosome 7 encoding cytokinin-*N*-glucosyltransferase 1 was identified. CTKs are a class of plant hormones firstly identified as cell division-promoting factors and were subsequently identified as factors that control various processes in plant growth and development, including mesocotyl elongation (Cao et al., 2005). Hu et al. (2014) reported that there was an antagonistic relationship between SLs and CTKs in regulating mesocotyl elongation. CTKs downregulate the expression of *OsTCP5* and promote cell division, whereas SLs upregulate *OsTCP5* expression and inhibit mesocotyl elongation. Besides, CTKs play an important role in the biosynthesis of BRs, a group of steroid plant hormones essential to plant growth and development (Yuldashev et al., 2012). A candidate gene (*LOC_Os01g13200*) on chromosome 1 related to abscisic acid insensitive 8 was identified. The structure–activity of ABA analogs had a close relationship with the growth of rice mesocotyl and seedlings (Watanabe and Takahashi, 1999). ABA compounds regulate activity of the meristem localized near the coleoptile node (Watanabe et al., 2001). In addition, ABA interacts with GA in regulating rice mesocotyl elongation (Wu et al., 2002). A candidate gene (*LOC_Os06g09660*) on chromosome 6, which is an auxin response factor, was identified. Auxin is a simple small molecule based on the indole ring, which plays an important role in plant growth and development, such as cell differentiation, cell division, and cell elongation (Xu and Xue, 2012). Feng et al. (2017) reported that the exogenous IAA could promote mesocotyl elongation of etiolated rice seedlings for 2 days after germination in darkness.

Two candidate genes (*LOC_Os01g66100* and *LOC_Os04g33360*) on chromosomes 1 and 4 were related with the biosynthesis of GA, which encodes gibberellin 20 oxidase 2 and gibberellin 2-beta-dioxygenase 7, respectively. Liang et al. (2016) reported that GAs regulate mesocotyl cell elongation in rice. The destabilization of cortical microtubules (CMTs) increased the GA level and induced mesocotyl cell elongation, while polymerization of CMT showed the opposite result. The expressions of *GA20ox2*, *GA3ox2*, and *GID1* in GA biosynthesis and transferring pathway were regulated by CMTs (Liang et al., 2016). Two candidate genes (*LOC_Os07g28060* and *LOC_Os09g20350*) on chromosomes 7 and 9 were related with ethylene biosynthesis, which are an ethylene receptor and ethylene-responsive transcription factor, respectively. All of the above plant hormones have significant influence on mesocotyl elongation: SLs, CTKs, ABA, BRs, IAAs, JAs, and ecdysis-triggering hormone (ETH). The above hormones, except for ETH, have direct influence on mesocotyl elongation by affecting either cell division or elongation. Only ETH works as a signal to regulate cell elongation through the JA biosynthesis pathway. The expression of *GY1* was inhibited by ETH signaling in a MHZ7/OsEIN2-dependent way. When OsEIL2 binds to the *GY1* promoter, its activity was repressed directly, which further inhibited JA biosynthesis and promoted cell elongation (Xiong et al., 2017). A candidate gene (*LOC_Os03g56060*) on chromosome 3, a member of the CSLC9-cellulose synthase-like family C, was identified. Cellulose synthase is involved in plant growth and development, including the development of roots and stems and the elongation of root hair (Li et al., 2019; Moon et al., 2019). The expressions of eight candidate genes in different accessions indicated that these genes were all functional in regulating mesocotyl elongation.

For loci identified by GWAS and QTL-seq, based on the variations in *qML1.3* and *qML7.2*, functional markers *Indel-Chr1:18932318* and *Indel-Chr7:15404166* were developed to genotype the XI-1B panel. It is co-dominant, breeder-friendly, and can be easily implemented in the laboratory. Its potential value for the selection of long mesocotyl germplasm was validated by association analysis. *Indel-Chr1:18932318* and *Indel-Chr7:15404166*, identified with the marker, were associated with ML. Although frequencies of favorable alleles of the two loci were higher among XI-1A globally, further increases are still feasible in regions where the alleles occur at higher frequencies. Functional markers related to higher MLs so far reported are available for *OsGSK2* and *OsSMAX1* (Sun et al., 2018; Zheng et al., 2020). Marker-assisted selection (MAS) for longer mesocotyl breeding based on combinations of these favorable alleles will be more effective than selection for single markers. Our study provides two valuable functional markers for longer mesocotyl breeding. Our follow-up studies will focus on validating the effects of these QTL, decoding the mechanism on mesocotyl elongation, and utilizing the markers to assist breeding practice.

CONCLUSION

Identifying QTL associated with mesocotyl elongation could accelerate genetic improvements of ML in rice breeding. QTL-seq of 12 F_2 populations and GWAS of two diverse panels were conducted to identify QTL for ML. In total, 14 QTL were identified by QTL-seq and five loci explaining 5.3–14.6% of the phenotypic variations were identified by GWAS. Among these, six were stable across two or more populations and 10 were potentially novel loci. Besides, eight high-confidence candidate genes were identified involved in the biological metabolism of plant hormones, cell elongation, and cell division. In addition, two PCR gel-based markers (*Indel-Chr1:18932318* for *qML1.3* and *Indel-Chr7:15404166* for *qML7.2*) were validated and could be used as breeder-friendly markers in further breeding. This study provides new insights into the genetic regulation mechanism of rice mesocotyl elongation and could promote the rice breeding process.

DATA AVAILABILITY STATEMENT

The datasets to the Genome Sequence Archive can be found at https://ngdc.cncb.ac.cn/search/?dbId=&q=PRJCA005531, accession number PRJCA005531.

AUTHOR CONTRIBUTIONS

GY conceived and designed the experiments. YW, YM, HL, and CL performed the experiments. JL and YW analyzed the data. YW, JL, and GY wrote and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by China Postdoctoral Science Foundation (2020M682999) and the Agricultural Science and Technology Innovation Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.713446/full#supplementary-material

Supplementary Figure 1 | Distribution of mesocotyl length in 12 F_2 populations.

Supplementary Figure 2 | Distribution of mesocotyl length in XI-1A, AUS and XI-1B panel.

Supplementary Figure 3 | LD decay along the whole genome of the AUS and XI-1A panel.

Supplementary Figure 4 | Population and kinship analysis for the XI-1B panel. **(A)** PCA analysis; **(B)** Kinship analysis.

Supplementary Table 1 | Mesocotyl length of the accessions from XI-1A panel.

Supplementary Table 2 | Mesocotyl length of the accessions from AUS panel.

Supplementary Table 3 | Details of the next-generation sequencing data for 12 F_2 populations.

Supplementary Table 4 | Number of SNPs for the 12 F_2 populations.

Supplementary Table 5 | GWAS for mesocotyl length in the XI-1A panel.

Supplementary Table 6 | GWAS for mesocotyl length in the AUS panel.

Supplementary Table 7 | Candidate genes in the loci identified by QTL-seq and GWAS.

Supplementary Tables 8 | Primer information for qRT-PCR of the candidate genes.

Supplementary Table 9 | Mesocotyl length and allele information for developed markers in XI-1B panel.

REFERENCES

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Arikit, S., Wanchana, S., Khanthong, S., Saensuk, C., Thianthavon, T., Vanavichit, A., et al. (2019). QTL-seq identifies cooked grain elongation QTLs near soluble starch synthase and starch branching enzymes in rice (Oryza sativa L.). *Sci. Rep.* 9, 1–10. doi: 10.1016/0168-9452(88)90135-5

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

Breseghello, F., and Sorrells, M. E. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172, 1165–1177. doi: 10.1534/genetics.105.044586

Cao, L. Y., Yuan, S. J., Zhou, H. P., Zhan, X. D., Wu, W. M., Gao, J. X., et al. (2005). Effect of different hormones on mesocotyl length in Oryza sativa L. *Acta Agronomica Sinica* 31, 1098–1100.

Cao, L. Y., Zhu, J., Yan, Q. C., He, L. B., Wei, X. H., and Cheng, S. H. (2002). Mapping QTLs with epistasis for mesocotyl length in a DH population from indica-japonica cross of rice (Oryza sativa. L). Chin. *J. Rice Sci.* 16, 221–224.

Chen, J., Zou, W., Meng, L., Fan, X., Xu, G., and Ye, G. (2019). Advances in the uptake and transport mechanisms and QTLs mapping of cadmium in rice. *Int. J Mol. Sci.* 20:3417. doi: 10.3390/ijms20143417

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem. Bull.* 19, 11–15.

Feng, F. J., Mei, H. W., Fan, P. Q., Li, Y. N., Xu, X. Y., Wei, H. B., et al. (2017). Dynamic transcriptome and phytohormone profiling along the time of light exposure in the mesocotyl of rice seedling. *Sci. Rep.* 7:11961.

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374.

Hu, Z. Y., Yamauchi, T., Yang, J. H., Jikumaru, Y., Tsuchida-Mayama, T., Ichikawa, H., et al. (2014). Strigolactone and cytokinin act antagonistically in regulating rice mesocotyl elongation in darkness. *Plant Cell Physiol.* 55, 30–41. doi: 10.1093/pcp/pct150

Kato, Y., and Katsura, K. (2014). Rice adaptation to aerobic soils: physiological considerations and implications for agronomy. *Plant Prod. Sci.* 17, 1–12. doi: 10.1626/pps.17.1

Kumar, V., and Ladha, J. K. (2011). Direct seeding of rice: recent developments and future research needs. *Adv. Agron.* 111, 297–302. doi: 10.1016/b978-0-12-387689-8.00001-1

Lalitha, S. (2000). Primer premier 5. biotech software internet report. *Comput. Softw. J. Sci.* 1, 270–272. doi: 10.1089/152791600459894

Lee, H. S., Sasaki, K., Kang, J. W., Sato, T., Song, W. Y., and Ahn, S. N. (2017). Mesocotyl elongation is essential for seedling emergence under deep-seeding condition in rice. *Rice* 10:32.

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. N., Yan, M., Feng, F. J., Wu, J. H., Xu, X. Y., Fan, P. Q., et al. (2017). Identification of chromosomal regions influencing mesocotyl elongation by bulked segregation analysis based on genome resequencing in rice. *Acta Agric. Shanghai* 33, 10–15.

Li, Y., Cheng, X., Fu, Y., Wu, Q., Guo, Y., Peng, J., et al. (2019). A genome-wide analysis of the cellulose synthase-like (Csl) gene family in maize. *Biologia. Plantarum* 63, 721–732. doi: 10.32615/bp.2019.081

Liang, Q., Wang, C., Ma, D. R., Li, L., Cui, Z. B., Wang, X. X., et al. (2016). Cortical microtubule disorganized related to an endogenous gibberellin increase plays an important role in rice mesocotyl elongation. *Plant Biotechnol. J.* 33, 59–69. doi: 10.5511/plantbiotechnology.16.0205a

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Liu, H., Hussain, S., Zheng, M., Peng, S., Huang, J., Cui, K., et al. (2015). Dry direct-seeded rice as an alternative to transplanted-flooded rice in Central China. *Agron. Sustain. Dev.* 35, 285–294. doi: 10.1007/s13593-014-0239-0

Liu, H., Zhan, J., Li, J., Lu, X., Liu, J., Wang, Y., et al. (2020). Genome-wide association study (GWAS) for mesocotyl elongation in rice (*Oryza sativa* L.) under multiple culture conditions. *Genes* 11:49. doi: 10.3390/genes11010049

Liu, J. D., He, Z. H., Rasheed, A., Wen, W. E., Yan, J., Zhang, P. Z., et al. (2017). Genome-wide association mapping of black point reaction in common wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 17:220. doi: 10.1186/s12870-017-1167-3

Liu, S., Zou, W., Lu, X., Bian, J., He, H., Chen, J., et al. (2021). Genome-wide association study using a multiparent advanced generation intercross (MAGIC) population identified QTLs and candidate genes to predict shoot and grain Zinc contents in Rice. *Agriculture* 11:70. doi: 10.3390/agriculture11010070

Lu, Q., Zhang, M. C., Niu, X. J., Wang, C. H., Xu, Q., and Feng, Y. (2016). Uncovering novel loci for mesocotyl elongation and shoot length in indica rice through genome-wide association mapping. *Planta* 243, 645–657. doi: 10.1007/s00425-015-2434-x

Luo, X. D., Liu, J., Zhao, J., Dai, L. F., Chen, Y. L., Zhang, L., et al. (2018). Rapid mapping of candidate genes for cold tolerance in *Oryza rufipogon* Griff. by QTL-seq of seedlings. *J. Integr. Agr.* 17, 265–275. doi: 10.1016/s2095-3119(17)61712-x

Lv, Y., Shao, G., Jiao, G., Sheng, Z., Xie, L., Hu, S., et al. (2021). Targeted mutagenesis of POLYAMINE OXIDASE 5 that negatively regulates mesocotyl elongation enables the generation of direct-seeding rice with improved grain yield. *Mol. Plant* 14, 344–351. doi: 10.1016/j.molp.2020.11.007

Mahender, A., Anandan, A., and Pradhan, S. K. (2015). Early seedling vigour, an imperative trait for direct-seeded rice: an overview on physio-morphological parameters and molecular markers. *Planta* 241, 1027–1050. doi: 10.1007/s00425-015-2273-9

Meng, L., Zhao, X., Ponce, K., Ye, G., and Leung, H. (2016). QTL mapping for agronomic traits using multi-parent advanced generation inter-cross (MAGIC) populations derived from diverse elite indica rice lines. *Field Crop. Res.* 189, 19–42. doi: 10.1016/j.fcr.2016.02.004

Moon, S., Chandran, A. K. N., Kim, Y. J., Gho, Y., Hong, W. J., An, G., et al. (2019). Rice RHC encoding a putative cellulase is essential for normal root hair elongation. *J. Plant Biol.* 62, 82–91. doi: 10.1007/s12374-018-0393-z

Nishida, S., Dissanayaka, D. B., Honda, S., Tateishi, Y., Chuba, M., Maruyama, H., et al. (2018). Identification of genomic regions associated with low phosphorus tolerance in japonica rice (*Oryza sativa* L.) by QTL-Seq. *Soil Sci. Plant Nutr.* 64, 278–281. doi: 10.1080/00380768.2017.1412238

Ohno, H., Banayo, N. P., Bueno, C. S., Kashiwagi, J. I., Nakashima, T., Corales, A. M., et al. (2018). Longer mesocotyl contributes to quick seedling establishment, improved root anchorage, and early vigor of deep-sown rice. *Field Crop. Res.* 228, 84–92. doi: 10.1016/j.fcr.2018.08.015

Ouyang, Y. N., Zhang, Q. Y., Zhang, K. Q., Yu, S. M., Zhuang, J. Y., and Jin, Q. Y. (2005). QTL mapping and interaction analysis of genotype/environment (Fe2+-concentrations) for mesocotyl length in rice (*Oryza sativa* L.). *Acta Genet. Sin.* 32, 712–718.

Ponce, K. S., Meng, L. J., Guo, L. B., Leng, Y. J., and Ye, G. Y. (2021). Advances in sensing, response and regulation mechanism of salt tolerance in rice. *Int. J. Mol. Sci.* 22:2254. doi: 10.3390/ijms22052254

Prodhomme, C., Vos, P. G., Paulo, M. J., Tammes, J. E., Visser, R. G., Vossen, J. H., et al. (2020). Distribution of P1 (D1) wart disease resistance in potato germplasm and GWAS identification of haplotype-specific SNP markers. *Theor. Appl. Genet.* 6:133.

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevedo, C. F., Takahashi, E. K., Silva-Junior, O. B., et al. (2017). Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in Eucalyptus. *New Phytol.* 213, 1287–1300. doi: 10.1111/nph.14266

Rohilla, M., Singh, N., Mazumder, A., Sen, P., Roy, P., Chowdhury, D., et al. (2020). Genome-wide association studies using 50 K rice genic SNP chip unveil genetic architecture for anaerobic germination of deep-water rice population of Assam. *India. Mol Genet Genomics* 295, 1211–1226. doi: 10.1007/s00438-020-01690-w

Salvi, S., and Tuberosa, R. (2005). To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci.* 10, 297–304. doi: 10.1016/j.tplants.2005.04.008

Sun, S. Y., Wang, T., Wang, L. L., Li, X. M., Jia, Y. C., Liu, C., et al. (2018). Natural selection of a GSK3 determines rice mesocotyl domestication by coordinating strigolactone and brassinosteroid signaling. *Nat Commun.* 9:2523.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105

Takagi, H., Tamiru, M., Abe, A., Yoshida, K., Uemura, A., Yaegashi, H., et al. (2015). MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nat. Biotechnol.* 33, 445–449. doi: 10.1038/nbt.3188

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.3923/ajcs.2011.43.48

Watanabe, H., and Takahashi, K. (1999). Effects of abscisic acid and its related compounds on rice seedling growth. *Plant Growth Regul.* 25, 5–8.

Watanabe, H., Takahashi, K., and Saigusa, M. (2001). Morphological and anatomical effects of abscisic acid (ABA) and fluridone (FLU) on the growth of rice mesocotyls. *Plant Growth Regul.* 34, 273–275.

Wu, J. H., Feng, F. J., Lian, X. M., Teng, X. Y., Wei, H. B., Yu, H., et al. (2015). Genome-wide association study (GWAS) of mesocotyl elongation based on re-sequencing approach in rice. *BMC Plant Biol.* 15:218. doi: 10.1186/s12870-015-0608-0

Wu, M., Zhang, G., Lin, J., and Cheng, S. (2005). Screening for rice germplasms with specially elongated mesocotyl. *Rice Sci.* 12, 226–228.

Wu, S. Q., Ding, R., and Li, X. S. (2002). Regulation of mesocotyl growth by gibberellic acid and abscisic acid in etiolated seedlings of black rice. *Amino. Acids. Biotic. Res.* 24, 44–45.

Xiong, Q., Ma, B., Lu, X., Huang, Y. H., He, S. J., Yang, C., et al. (2017). Ethylene inhibited jasmonic acid biosynthesis promotes mesocotyl/coleoptile elongation of etiolated rice seedlings. *Plant Cell* 29, 1053–1072. doi: 10.1105/tpc.16.00981

Xu, Z. H., and Xue, H. W. (2012). *Plant Hormones: Function and Molecular Mechanism*. Shanghai: Shanghai Scientific and Technical Publishers.

Yang, Y., Chai, Y., Zhang, X., Lu, S., Zhao, Z., Wei, D., et al. (2020). Multi-locus GWAS of quality traits in bread wheat: mining more candidate genes and possible regulatory network. *Front. Plant Sci.* 11:1091. doi: 10.3389/fpls.2020.01091

Yuldashev, R., Avalbaev, A., Bezrukova, M., Vysotskaya, L., Khripach, V., and Shakirova, F. (2012). Cytokinin oxidase is involved in the regulation of cytokinin content by 24-epibrassinolide in wheat seedlings. *Plant Physiol. Bioch.* 55, 1–6. doi: 10.1016/j.plaphy.2012.03.004

Zhan, J., Lu, X., Liu, H., Zhao, Q., and Ye, G. (2020). Mesocotyl elongation, an essential trait for dry-seeded rice (*Oryza sativa* L.): a review of physiological and genetic basis. *Planta* 251, 1–14.

Zhang, Y., Ponce, K. S., Meng, L., Chakraborty, P., Zhao, Q., Guo, L., et al. (2020). QTL identification for salt tolerance related traits at the seedling stage in indica rice using a multi-parent advanced generation intercross (MAGIC) population. *Plant Growth Regul.* 92, 365–373. doi: 10.1007/s10725-020-00644-x

Zhao, Y., Zhao, W. P., Jiang, C. H., Wang, X. N., Xiong, H. Y., Todorovska, E. G., et al. (2018). Genetic architecture and candidate genes for deep-sowing tolerance in rice revealed by non-syn GWAS. *Front. Plant Sci.* 9:332. doi: 10.3389/fpls.2018.00332

Zheng, J. S., Hong, K., Zeng, L. J., Wang, L., Kang, S. J., Qu, M. H., et al. (2020). Karrikin signaling acts parallel to and additively with strigolactone signaling to regulate rice mesocotyl elongation in darkness. *Plant Cell* 32, 2780–2805. doi: 10.1105/tpc.20.00123

Zhu, C. S., Gore, M., Buckler, E. S., and Status, Y. J. M. (2008). Prospects of association mapping in plants. *Plant Genome* 1, 5–20.

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 September 2021
doi: 10.3389/fgene.2021.714942

Identification of the Carbohydrate and Organic Acid Metabolism Genes Responsible for Brix in Tomato Fruit by Transcriptome and Metabolome Analysis

Ning Li[1,2†], Juan Wang[1†], Baike Wang[1], Shaoyong Huang[1,2], Jiahui Hu[3], Tao Yang[1], Patiguli Asmutola[1], Haiyan Lan[3] and Yu Qinghui[1]**

[1] Institute of Horticulture Crops, Xinjiang Academy of Agricultural Sciences, Urumqi, China, [2] College of Forestry and Horticulture, Xinjiang Agricultural University, Urumqi, China, [3] Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Urumqi, China

OPEN ACCESS

Edited by:
Ahmed Sallam,
Assiut University, Egypt

Reviewed by:
Maria Ines Zanor,
CONICET Instituto de Biología Molecular y Celular de Rosario (IBR), Argentina
Diaa Abd El Moneim Abd El Moneim,
Arish University, Egypt

***Correspondence:**
Haiyan Lan
lanhaiyan@xju.edu.cn
Yu Qinghui
yuqinghui98@sina.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 26 May 2021
Accepted: 20 July 2021
Published: 03 September 2021

Citation:
Li N, Wang J, Wang B, Huang S, Hu J, Yang T, Asmutola P, Lan H and Qinghui Y (2021) Identification of the Carbohydrate and Organic Acid Metabolism Genes Responsible for Brix in Tomato Fruit by Transcriptome and Metabolome Analysis. Front. Genet. 12:714942. doi: 10.3389/fgene.2021.714942

Background: Sugar and organic acids not only contribute to the formation of soluble solids (Brix) but also are an essential factor affecting the overall flavor intensity. However, the possible metabolic targets and molecular synthesis mechanisms remain to be further clarified.

Methods: UHPLC-HRMS (ultrahigh-performance liquid chromatography and high-resolution mass spectrometry) combined with comparative transcriptome analysis were performed in fruits at green ripe (S1), turning-color (S2), and red ripe (S3) stages of two tomato genotypes TM-1 (*Solanum galapagense* L., LA0436) and TM-38 (*S. lycopersicum* L. cultivar M82, LA3475) that vary in fruit Brix.

Results: The fruit Brix of TM-1 was nearly twice that of TM-38 at S3. Nevertheless, TM-1 accumulated 1.84- and 2.77-fold the L-malic acid and citric acid in red ripe fruit (S3) compared with TM-38, respectively. D-glucose and D-fructose in TM-1 and TM-38 fruits tended to be similar at S3. Concomitantly, the sugar/organic acid ratio of TM-38 fruits were 23. 08-, 4. 38-, and 2.59-fold higher than that of TM-1 fruits at S1, S2, and S3, respectively. Among starch and sucrose (carbohydrate, CHO) metabolism (ko00500) genes, *SUS* (Solyc07g042550.3) and *BAM* (Solyc08g077530.3) were positively (r = 0.885–0.931) correlated with the sugar/organic acid ratio. Besides, *INV* (Solyc09g010080.3 and Solyc09g010090.5.1), *AAM* (Solyc04g082090.3), *4-α-GTase* (Solyc02g020980.2.1), *BGL2* (Solyc06g073750.4, Solyc06g073760.3, and Solyc01g081170.3), *TPS* (Solyc01g005210.2 and Solyc07g006500.3), and *TPP* (Solyc08g079060.4) were negatively (r = −0.823 to −0.918) correlated with the sugar/organic acid ratio. The organic acid (TCA cycle) metabolism (ko00020) gene *ALMT* (Solyc01g096140.3) was also negatively (r = −0.905) correlated with the sugar/organic acid ratio.

Conclusion: Citric acid may play a more dominant role in the sugar/organic acid ratio of the tomato fruit, and the contribution of both L-malic acid and citric acid to the fruit Brix was much greater than that of D-glucose and D-fructose. Genes involved in CHO and TCA metabolism, which have a significant correlation with the sugar/organic acid ratio were considered to be the contributing factors of fruit Brix.

Keywords: tomato, fruit brix, organic acid, carbohydrate, metabolic, transcriptome, TCA cycle

INTRODUCTION

Tomato (*Solanum lycopersicum*) is one of the most popular and valuable fruits with limited caloric supply, and an excellent source of fiber, minerals, phenols, vitamins A, C, E, and lycopene, making it an excellent "functional food" meeting basic nutritional requirements (Dorais et al., 2008; Beckles et al., 2011; Giovannetti et al., 2012). Total soluble solids (TSS) is the most important fruit quality parameter in both fresh market and processed tomatoes, indicating the proportion (%) of dissolved solids in a solution (Schaffer et al., 1999; Xu et al., 2018). It is the sum of sugars (sucrose and hexoses; 65%), acids (citrate and malate; 13%), and other minor components (phenols, amino acids, soluble pectins, ascorbic acid, and minerals) in the tomato fruit pulp (Kader, 2008; Beckles, 2012). Sugars and organic acids not only contribute to the improvement of TSS (Brix), one of the key parameters in tomato processing but also play an essential role in overall flavor intensity (Barickman et al., 2016; Carlos et al., 2018). The aim of this study was to compare the contents of organic acids and carbohydrates in two tomato genotypes with drastically different fruit Brix, and to further elucidate a correlation of carbohydrate and organic acids with gene expression levels, which would supply a reference of molecular-assisted selection of high fruit Brix tomato germplasm resources.

Tomato fruit taste and quality vary with varieties, growing conditions, production methods, harvest, and storage time (Langlois et al., 1996; Giovannetti et al., 2012). Natural biodiversity provides opportunities to investigate the multitude of characteristics that affect plant growth and development. Tomato has been an introgression model for valuable traits from wild species (Rick, 1974). Wild plant species offer a way to understand the genetic basis of past domestication events and polymorphisms, providing a basis for breeding superior varieties in the future (Ofner et al., 2016). *S. lycopersicum* is known to hybridize easily with the wild relatives, which can provide valuable sources for the improvement of important agronomic traits (Koblitz, 1991; Schauer et al., 2005). The fruit quality of *S. lycopersicum* is associated with various parameters, including appearance, color intensity, size, shape, flavor, nutritional value, and texture, which ultimately determine acceptability for the consumer (Zhu et al., 2018; Verma et al., 2020). The intensity of tomato flavor is mainly determined by the amount of sugars and organic acids (Çolaka et al., 2020).

The precursor substance ADP-glucose of starch synthesis is catalyzed by the rate-limiting enzyme of ADP-Glc pyrophosphorylase (AGPase), and starch is mainly degraded into reducing sugars under acidic conditions by the action of starch phosphorylase (SP) (Sweetlove et al., 1999). To be metabolized, starch can be hydrolyzed into sucrose to release energy, and sucrose can also form starch to store energy (Gifford et al., 1984; Roby et al., 2002). Sucrose, as a non-reducing sugar, can be decomposed by invertase (INV; EC 3.2.1.26) to form reducing glucose and fructose, and also be reacted with UDP by sucrose synthase (SUS; EC 2.4.1.13) to form uridine diphosphate glucose (UDPG) and fructose (Ruan, 2014; Wan et al., 2018). Sucrose-phosphate synthase (SPS; EC 2.4.1.14) is a key enzyme for catalyzing the conversion of UDP-glucose and fructose-6-phosphate into sucrose-6-phosphate, and sucrose-phosphatase (SPP) can convert sucrose-6-phosphate into sucrose (Grof et al., 1998; Wind et al., 2010; Hashida et al., 2016). UDPG is both a precursor substance of starch synthesis and is also affected by uridine diphosphate glucose pyrophosphorylase (UGPase) forms glucose-1-phosphate to promote the re-synthesis of sucrose (Menendez et al., 2002; Finlay et al., 2003). Through amylase (AMY) activity, starch can also yield maltose, which is exported to the cytoplasm and cleaved to produce glucose monomers (Weise et al., 2004). Although glucose and fructose are interconverted after phosphorylation, glucose is more preferentially used than fructose in several plant cells (Kandel-Kfir et al., 2006). After phosphorylation, glucose and fructose are used for the growth or synthesis of storage materials: sucrose and starch (Krook et al., 1998). On the other aspect, organic acids are key factors in maintaining pH and changing the sensory quality of fruit, and the evaluation of fruit maturity and the quality of a particular variety depends on the sugar/organic acid ratio (Jie et al., 2018). The genes encoding the aluminum-activated malate transporter have often been reported to be involved in the regulation of the organic acid levels (Jawad et al., 2020; Umer et al., 2020).

Since sugars and organic acids function as signaling molecules in many developmental processes throughout the plant life cycle, uncovering these functions and their interactions with other signaling pathways presents a formidable challenge (Lastdrager et al., 2014; Li et al., 2016; Jawad et al., 2020). A new technology

Abbreviations: *4-α-GTase,* 4-alpha-glucanotransferase; *AAM,* alpha-amylase; *ACS,* ATP-citrate synthase alpha chain protein; *AGP,* ADP-glucose pyrophosphorylase large subunit; AGPase, ADP-Glc pyrophosphorylase; *ALMT,* aluminum-activated malate transporter; *BAM,* beta-amylase; *BGL,* beta-glucosidase; CHO, carbohydrate; *CS,* citrate synthase; DEGs, differentially expressed genes; GO, Gene Ontology; HPLC, high-performance liquid chromatography; *HXK,* hexokinase; *ICDH,* isocitrate dehydrogenase; *INV,* invertase; KEGG, Kyoto Encyclopedia of Genes and Genomes; *PDHB-1,* pyruvate dehydrogenase E1 component subunit beta-1; *SDH,* succinate dehydrogenase (ubiquinone) iron–sulfur subunit 3; *SP,* starch phosphorylase; *SPP,* sucrose-phosphatase; *SPS*, sucrose-phosphate synthase; *TPP,* trehalose-phosphate phosphatase; TPS, trehalose-phosphate synthase; TSS, total soluble solids; UDP-glucose, uridine diphosphate glucose; UGPase, uridine diphosphate glucose pyrophosphorylase.

such as transcriptome has been used to uncover the genes involved in starch and sucrose (carbohydrate, CHO) and organic acid (TCA) metabolisms (Zhang et al., 2015; Umer et al., 2020). Therefore, a potential mechanism for identifying key candidate genes responsible for divergent fruit Brix content is presented in this study. Metabolite profiles evaluated using UHPLC-HRMS (ultrahigh-performance liquid chromatography and high-resolution mass spectrometry) and transcript level of selected genes coding for enzymes metabolism were determined in fruits of TM-1 (*S. galapagense* L., LA0436) and TM-38 (*S. lycopersicum* L. cultivar M82, LA3475) at three developmental stages.

MATERIALS AND METHODS

Plant Materials

Two tomato genotypes, differing in carbohydrate content, TM-1 (*S. galapagense* L., LA0436), and TM-38 (*S. lycopersicum* L. cultivar M82, LA3475) were introduced from the UC Davis/C.M. Tomato Genetics Resource Center (TGRC) and maintained by the Department of Plant Science, University of California, Davis, CA 95616. The TGRC undertook formal identification of the samples, provided details of the specimens deposited and allowed the collection. Both tomato genotypes were grown in the Anningqu experimental station of the Xinjiang Academy of Agricultural Sciences (87°49′63″N, 43°95′16″E; altitude: 680–920 m). Briefly, six fruits from each of the two strains were randomly sampled as a biological repeat at green ripe stage (S1), turning-color period (S2), and red ripe (S3) stage, respectively. There were three biological replicates for each genotype. All fruit samples were immediately frozen in liquid nitrogen and stored at −80°C for further physical, metabolic, and gene expression evaluation.

Determination of Fruit Diameter and Total Soluble Solids

The determination of fruit diameter and TSS was based on the tomato fruit at S3. The transverse and longitudinal diameters were measured by a vernier caliper with an accuracy of 0.01 mm (Mitutoyo CD-15CPX, Japan). TSS was measured by refractometer sugar sweetness meter (Guangzhou Weilai Electronic Technology Co., Ltd.).

Metabolite Extraction for Ultrahigh-Performance Liquid Chromatography and High-Resolution Mass Spectrometry

Lyophilized tomato fruit samples were finely ground, and a certain amount of powdered samples (see Data Sheet 1) were placed into EP tubes. Then an appropriate amount of extraction solution (10% methanol, see Data Sheet 1 for specific volume) was added. After vortex mixing for 30 s, steel balls were added and ground for 4 min at 45 Hz in a Retsch® Mixer Mill MM400 (Retsch, Haan, Germany), then ultrasonicated for three times and incubated in ice water every 5 min. After centrifugation at 13,000 × *g* for 15 min at 4°C, the supernatant was transferred for the UHPLC-HRMS analysis.

Targeted Metabolomics Profiling of Ultrahigh-Performance Liquid Chromatography and High-Resolution Mass Spectrometry

UHPLC-HRMS analyses were performed using Waters ACQUITY UPLC (Waters, Millford, MA, United States) ultrahigh-performance liquid chromatograph equipped with a Waters ACQUITY UPLC BEH C_{18} column (100 mm × 2.1 mm, 1.7 μm, Waters) to separate the target compounds. Full scan mass spectrometry was performed by XEVO G2XS Q-TOF high-resolution mass spectrometer. The ion source parameters are as follows: capillary voltage = 2,000 V, sampling cone = 40 V, source temperature = 115°C, desolvation temperature = 500°C, and desolvation gas = 900 L/h. For each target compound, the parent ions under high-resolution (QTOF) conditions were used for quantitative analysis. The specific parameters [retention time (RT), mass-to-charge ratio (m/z), and polarity] are shown in **Supplementary Table 2**. The calibration curve is shown in Data Sheet 2; y is the peak area of the target compound, and x is the concentration of the target compound (μg/ml). The least square method was used for regression analysis. When the weight was set at 1/x, the calibration solution recovery rate (accuracy) and correlation coefficient (R^2) were the best. If the signal-to-noise ratio (S/N) of a calibration concentration is close to 20, or the recovery rate exceeds the range of 80–120%, the calibration point of the concentration was excluded.

RNA Isolation

Total RNA was isolated from approximately 200 mg of lyophilized tomato fruit samples collected at 45 DAF using an RNAprep Pure Plant Plus Kit (TIANGEN, Beijing, China). Then the quality and quantity of the purified RNA samples were preliminarily characterized by Multiskan Go Full Wavelength Microplate Spectrophotometer (Thermo Fisher Scientific, MA, Waltham, United States).

RNA-seq and Differential Gene Expression Analysis

The six triplicate samples (TM-1 and TM-38 at three developmental stages) yielded 18 non-directional cDNA libraries with a total of 121.92 Gb of clean data (**Table 1**) using illumina HiSeq 2500 platform by signal end read libraries method of the SBS (Sequencing By Synthesis) technology, which was performed at the Biomarker Technologies Co., Ltd. (Beijing, China). The raw reads were cleaned, and the clean reads were aligned onto the tomato reference genome[1]. During the detection process of differentially expressed genes (DEGs), fold change > 2 and false discovery rate (FDR) < 0.01 was used as the screening standard. Gene expression was scaled using values of the fragments per kilobase of exon per million mapped reads (FPKM) ≥ 1.0 as a threshold to identify significant DEGs

[1]ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG4.0_release/)

TABLE 1 | Summary of RNA-Seq data and mapping metrics.

Variety	Replicate	Total reads	Clean reads	Mapped reads	% ≥ Q30
TM-1-S1	1	46,037,248	23,018,624	43,430,222 (94.34%)	93.21%
	2	39,150,626	19,575,313	36,351,508 (92.85%)	93.88%
	3	52,161,828	26,080,914	48,816,501 (93.59%)	93.68%
TM-1-S2	1	42,747,216	21,373,608	37,308,962 (87.28%)	93.67%
	2	46,262,136	23,131,068	40,358,796 (87.24%)	93.94%
	3	45,726,338	22,863,169	38,948,767 (85.18%)	93.21%
TM-1-S3	1	42,891,178	21,445,589	40,301,779 (93.96%)	93.78%
	2	48,810,696	24,405,348	45,831,719 (93.90%)	93.72%
	3	46,020,758	23,010,379	43,113,028 (93.68%)	92.13%
TM-38-S1	1	43,410,496	21,705,248	40,664,031 (93.67%)	93.65%
	2	45,692,424	22,846,212	42,741,586 (93.54%)	93.71%
	3	44,372,464	22,186,232	41,571,244 (93.69%)	93.64%
TM-38-S2	1	42,858,120	21,429,060	39,913,732 (93.13%)	93.86%
	2	45,430,462	22,715,231	42,228,260 (92.95%)	93.71%
	3	49,700,956	24,850,478	46,521,555 (93.60%)	93.99%
TM-38-S3	1	49,868,822	24,934,411	46,489,230 (93.22%)	93.79%
	2	44,414,518	22,207,259	41,241,871 (92.86%)	93.44%
	3	41,359,802	20,679,901	38,359,691 (92.75%)	93.94%

(Kandel-Kfir et al., 2006). The Gene Ontology (GO)[2] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [3]databases were used to assign tomato genes for GO categories and KEGG pathway analyses, respectively.

RNA-seq Results Verification by Using Quantitative Reverse-Transcription PCR

The cDNA synthesis was conducted with total RNA using RNeasy Mini Kit (QIAGEN, GmbH, Hilden, Germany). Primers (**Supplementary Table 2**) were designed and synthesized by Sangon Biotech (Shanghai) Co., Ltd. (Shanghai, China). qRT-PCR assays were performed with Quanti Nova SYBR Green PCR Kit (QIAGEN) according to the instructions. Three biological and three technical replicates for each reaction were analyzed on a LightCycler® 96 SW 1.1 instrument (Roche). All relative expression levels of individual genes were normalized by comparing with TM-38 expression at S1 and calculated using the $2^{-\Delta\Delta C}$T method (Livak and Schmittgen, 2001).

Statistical Analysis

The Pearson correlation coefficient (*r*) was calculated for correlation analysis, and a two-tailed test was carried out. Origin 9.0 software was used to draw line charts and histograms; Heml software was used to generate heat maps.

RESULTS

Characterization of the Ripening Parameters in TM-1 and TM-38

TM-1 is a round fruit with thick skin and orange peel, TM-38 is an oval fruit with red peel and pink flesh (**Figure 1A**). The transverse and longitudinal diameters of the TM-38 fruits at S3 were 2.2- to 3.0-fold higher than that of the TM-1 fruits in 2019 and 2020 (**Figure 1B**). In contrast, the TSS of the TM-1 fruits at S3 were 7.6 and 5.7% Brix in 2019 and 2020, which were 2.0- and 1.5-fold higher than that of the TM-38 fruits, respectively.

Target Metabolite Profiles During Fruit Ripening and Development in TM-1 and TM-38

The overall trend of D-fructose and D-glucose contents in the TM-1 fruits increased from S1 to S3 (**Figure 2**). Nevertheless, L-malic acid and citric acid contents displayed an opposite trend. The contents of D-glucose and L-malic acid in the TM-38 fruits decreased from S1 to S3, whereas the contents of D-fructose and citric acid increased first and then decreased from S1 to S3. D-fructose in the TM-38 fruits was 4.55- and 1.72-fold higher than that in the TM-1 fruits at S1 and S2, respectively. Similarly, D-glucose in the TM-38 fruits was 4.89- and 1.51-fold higher than that in the TM-1 fruits at S1 and S2, respectively. However, the concentrations of D-fructose and D-fructose in the TM-1 and TM-38 fruits at S3 were similar. In contrast, the content of L-malic acid in the TM-1 fruits was 2. 17-, 1. 57-, and 1.84-fold higher than that in the TM-38 fruits at the three evaluated developmental stages, respectively. The content of citric acid in the TM-1 fruits was 7. 52-, 2. 97-, and 2.77-fold higher than that in the TM-38 fruits at the three developmental stages, respectively. Besides, the sugar/organic acid ratio of the TM-1 fruits increased from S1 to S3, whereas in the TM-38 fruits, it decreased. However, the sugar/organic acid ratio of the TM-38 fruits were 23. 08-, 4. 38-, and 2.59-fold higher than that of the TM-1 fruits at S1, S2, and S3, respectively.

Transcriptome Profiling of the TM-1 and TM-38 Fruits

Triplicate sampling of the TM-1 and TM-38 fruits at the three developmental stages yielded 18 RNA samples for transcriptome

[2]http://www.geneontology.org

[3]https://www.genome.jp/kegg/pathway.html

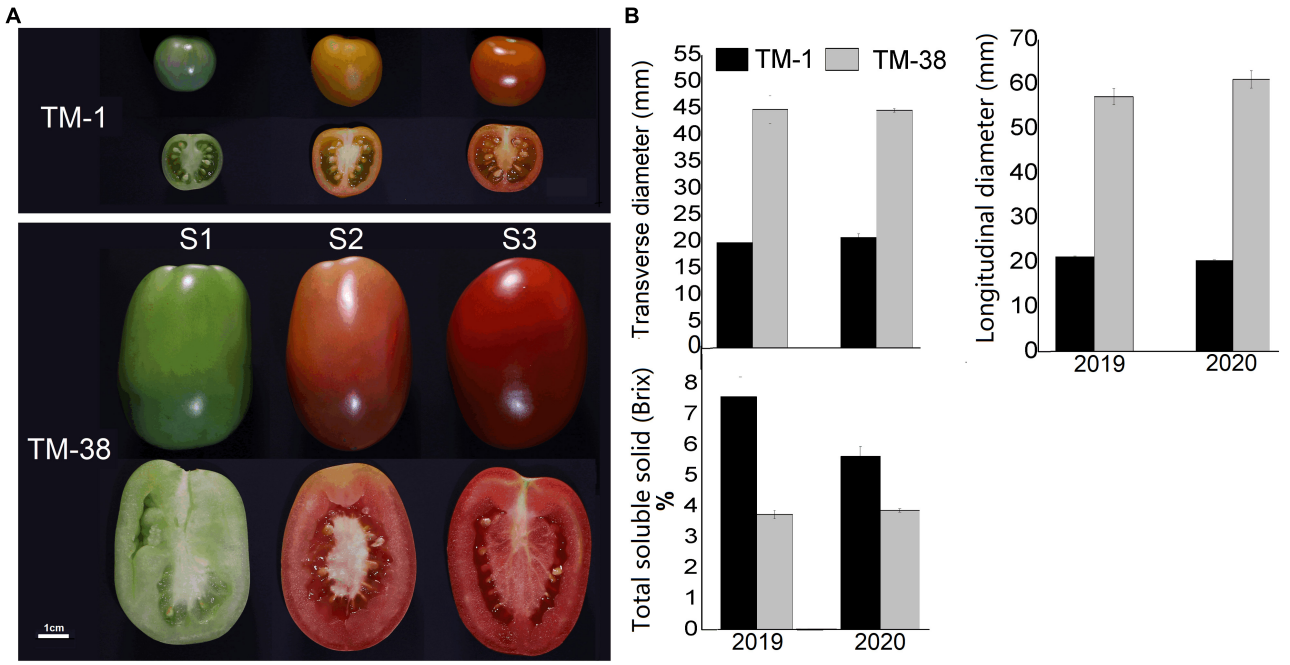


FIGURE 1 | (A) Close-up views of TM-1 (*Solanum galapagense* L., LA0436) and TM-38 (*S. lycopersicum* L. cultivar M82, LA3475) at green ripe (S1), turning-color (S2), and red ripe (S3) stages. **(B)** The changes in fruit diameters and total soluble solids (TSS) at S3. Values are means ± SE.

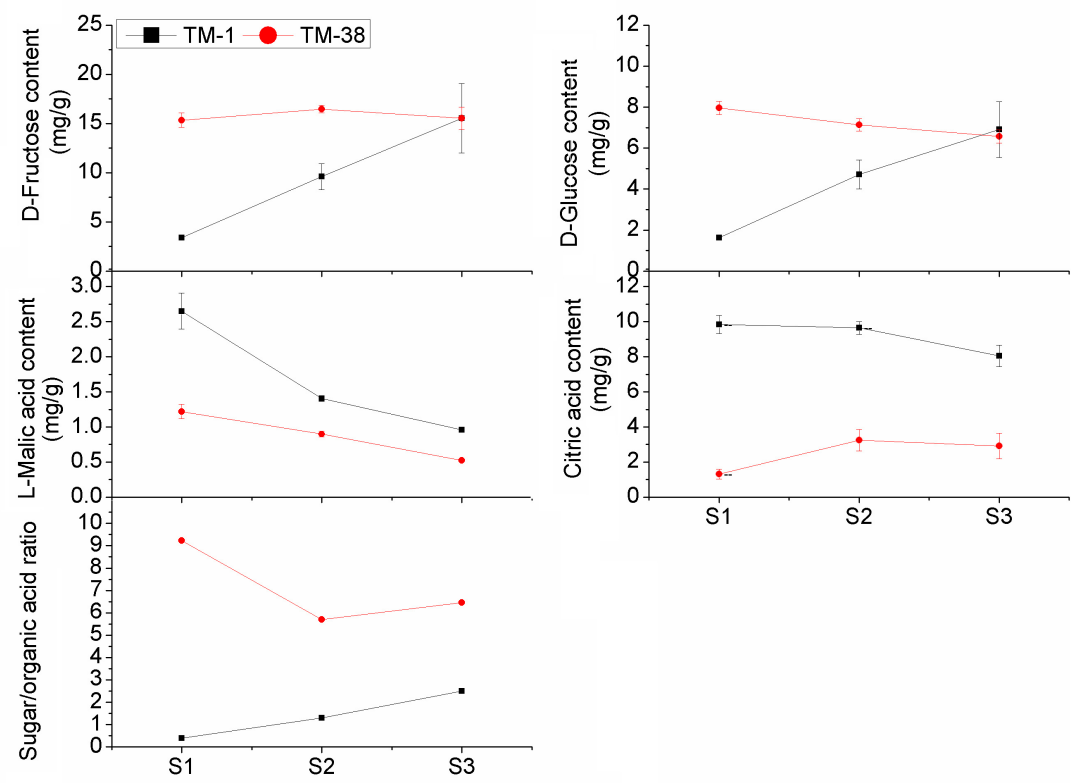


FIGURE 2 | Developmental profiles of D-fructose, D-glucose, L-malic acid, citric acid contents, and the sugar/organic acid ratio in TM-1 and TM-38 at three developmental stages. Values are means ± SE.

analysis; a total of 121.92 Gb of clean data were obtained (**Table 1**). The clean data of each sample reached 5.84 Gb, and the percentage of Q30 base was more than 92.13%. The clean reads of each sample were aligned with the tomato reference genome(see text footnote 1), and the mapping rate ranged from 85.18 to 94.34%.

Digital Analysis of Differentially Expressed Genes Between the TM-1 and TM-38 Fruits at the Three Developmental Stages

Seven pairwise transcriptome comparisons [i.e., TM-1 vs. TM-38 at S1, S2 and S3, TM-1 (S1 vs. S2), TM-1 (S2 vs. S3), TM-38 (S1 vs. S2), TM-38 (S2 vs. S3)] were performed to identify DEGs in the TM-1 and TM-38 fruits at the three developmental stages (**Figure 3** and **Table 2**). The number of DEGs in TM-1 vs. TM-38 was very large in the three developmental stages, accounting for 5,012, 4,881, and 5,245 transcripts, respectively (**Figure 3A**). There were also more DEGs in S1 vs. S2 of TM-1 and TM-38, which were 5,191 and 3,599, respectively. However, S2 vs. S3 of TM-1 and TM-38 had less DEGs (1,267 and 774, respectively). Among them, the number of DEGs shared by TM-1 vs. TM-38 at S1, S2, and S3 were 1,782 (**Figure 3B**), while that of S1 vs. S2 and S2 vs. S3 of TM-1 and TM-38 were 709 and 358, respectively (**Figures 3C,D**). On the whole, there were more downregulated genes than upregulated genes. The number of DEGs annotated to COG, GO, KEGG, KOG, NR, Pfam, Swiss-Prot, and eggNOG databases are also presented in **Supplementary Table 3**.

Identification of Key Processes Responsible for Organic Acid and Carbohydrate Accumulation in the TM-1 and TM-38 Fruits

To understand the main functional categories represented by DEGs, GO functional enrichment analysis was performed with all reference genes as the background. The top eight significantly enriched GO terms in TM-1 vs. TM-38 at the three developmental stages are displayed in three main categories: biological process, cellular component, and molecular function (**Figure 4**). GO terms of photosynthetic electron transport in photosystem II (GO:0009772), protein–chromophore linkage (GO:0018298), ATP synthesis-coupled proton transport (GO:0015986), ATP hydrolysis-coupled proton transport (GO:0015991), response to herbicide

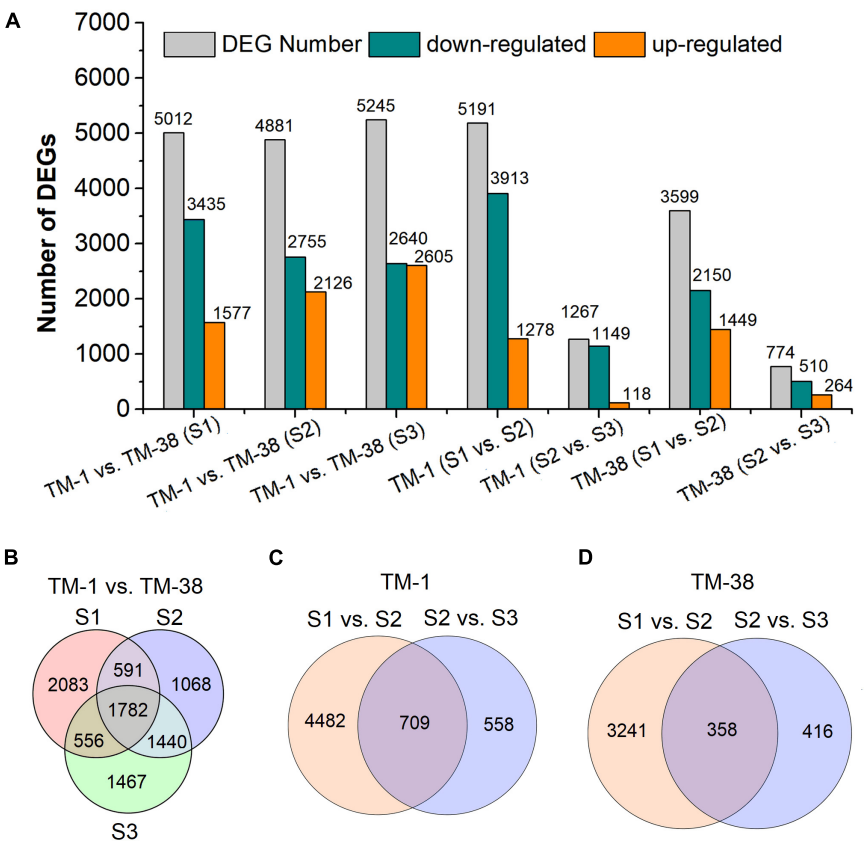


FIGURE 3 | Summary of the number of differentially expressed genes (DEGs) identified by RNA-seq analysis in tomato fruits of TM-1 and TM-38 at three developmental stages. The number of total DEGs, upregulated DEGs, and downregulated DEGs are presented by histograms **(A)**. The Venn diagrams represents the number of common DEGs shared by TM-1 vs. TM-38 at three developmental stages **(B)**, S1 vs. S2 and S2 vs. S3 of TM-1 **(C)**, and TM-38 **(D)**.

TABLE 2 | Correlation analysis of mean FPKM value of the starch and sucrose metabolism (ko00500) genes with D-fructose, D-glucose, L-malic acid, and citric acid contents, respectively.

Gene name	Accession No.	KEGG annotation	D-fructose content	D-glucose content	L-malic acid content	Citric acid content	Sugar/organic acid ratio
SUS	Solyc03g098290.4	sucrose synthase 5-like [EC 2.4.1.13]	0.519	0.454	−0.402	−0.603	0.484
	Solyc07g042550.3	sucrose synthase [EC 2.4.1.13]	0.596	0.595	−0.594	−0.908*	0.885*
SPS	Solyc08g042000.3	sucrose-phosphate synthase A2 [EC 2.4.1.14]	0.416	0.400	−0.483	−0.747	0.728
	Solyc09g092130.3	sucrose-phosphate synthase B [EC 2.4.1.14]	−0.763	−0.652	0.867*	0.618	−0.561
GAA	Solyc02g069670.4	alpha-glucosidase-like [EC 3.2.1.20]	−0.090	0.113	0.318	−0.358	0.479
	Solyc04g009630.3	alpha-glucosidase [EC 3.2.1.20]	0.424	0.379	−0.418	0.260	−0.243
INV	Solyc03g083910.5	acid beta-fructofuranosidase precursor [EC 3.2.1.26]	−0.116	−0.155	−0.020	0.726	−0.674
	Solyc09g010080.3	beta-fructofuranosidase [EC 3.2.1.26]	−0.805	−0.789	0.686	0.940**	−0.907*
	Solyc09g010090.5.1	cell-wall invertase [EC 3.2.1.26]	−0.866*	−0.827*	0.770	0.940**	−0.881*
	Solyc10g083290.4	acid invertase [EC 3.2.1.26]	−0.970**	−0.939**	0.973**	0.617	−0.617
BGL1	Solyc03g119080.4	beta-mannosidase precursor [EC 3.2.1.21]	0.290	0.499	−0.076	−0.389	0.533
	Solyc10g045240.2	vicianin hydrolase [EC 3.2.1.21]	−0.946**	−0.945**	0.946**	0.618	−0.640
	Solyc12g040640.2	beta-glucosidase 44-like isoform X1 [EC 3.2.1.21]	0.142	0.343	0.083	−0.525	0.649
SP	Solyc02g077680.4	glycogen phosphorylase 1-like isoform X1 [EC 2.4.1.1]	0.159	0.203	0.039	−0.722	0.671
	Solyc03g065340.3	alpha-1,4 glucan phosphorylase L-1 isozyme [EC 2.4.1.1]	0.147	0.345	0.083	−0.539	0.659
	Solyc05g012510.3	alpha-1,4 glucan phosphorylase L-2 isozyme [EC 2.4.1.1]	−0.037	0.155	0.276	−0.435	0.542
SSs	Solyc03g083090.4	soluble starch synthase 1 [EC 2.4.1.21]	0.655	0.560	−0.606	−0.445	0.336
AAM	Solyc04g078930.4	alpha-amylase [EC 3.2.1.1]	−0.873*	−0.886*	0.914*	0.573	−0.604
	Solyc04g082090.3	probable alpha-amylase 2 [EC 3.2.1.1]	−0.631	−0.626	0.577	0.958**	−0.918**
BAM	Solyc01g094580.3	beta-amylase 7 isoform X2 [EC 3.2.1.2]	−0.589	−0.510	0.617	0.784	−0.672
	Solyc08g077530.3	beta-amylase 3, chloroplastic [EC 3.2.1.2]	0.780	0.847*	−0.566	−0.922**	0.931**
	Solyc09g091030.3	beta-amylase [EC 3.2.1.2]	0.550	0.450	−0.513	−0.737	0.600
4-α-GTase	Solyc02g020980.2.1	4-alpha-glucanotransferase DPE2-like [EC 2.4.1.25]	−0.973**	−0.963**	0.915*	0.866*	−0.853*
	Solyc04g053120.3	4-alpha-glucanotransferase, chloroplastic/amyloplastic [EC 2.4.1.25]	0.026	0.227	0.201	−0.465	0.583
AGP	Solyc01g109790.3	ADP-glucose pyrophosphorylase large subunit [EC 2.7.7.27]	−0.025	0.172	0.256	−0.428	0.544
	Solyc07g019440.3	ADP-glucose pyrophosphorylase large subunit [EC 2.7.7.27]	0.659	0.617	−0.524	−0.901*	0.809
BGL2	Solyc01g010390.3	beta-glucosidase 40 [EC 2.4.1.15]	−0.848*	−0.850*	0.875*	0.406	−0.439
	Solyc01g081170.3	unnamed protein product, partial [EC 2.4.1.15]	−0.659	−0.656	0.580	0.971**	−0.916*
	Solyc02g080290.3	beta-glucosidase 18-like [EC 2.4.1.15]	0.205	0.331	−0.136	−0.668	0.749
	Solyc04g015560.4	uncharacterized protein LOC101247513 [EC 2.4.1.15]	−0.639	−0.506	0.780	0.540	−0.424
	Solyc06g005970.2	uncharacterized protein LOC101263519 [EC 2.4.1.15]	0.561	0.508	−0.457	−0.789	0.678
	Solyc06g073740.3	beta-glucosidase BoGH3B-like [EC 2.4.1.15]	0.091	0.078	−0.155	0.569	−0.508
	Solyc06g073750.4	uncharacterized protein LOC101266643 [EC 2.4.1.15]	−0.771	−0.748	0.695	0.971**	−0.911*
	Solyc06g076780.3	uncharacterized protein LOC101260057 [EC 2.4.1.15]	−0.178	−0.007	0.413	−0.381	0.460
	Solyc07g063390.3	beta-glucosidase 18-like isoform X2 [EC 2.4.1.15]	−0.351	−0.378	0.335	0.844*	−0.805
	Solyc07g063880.4	putative beta-glucosidase 41 [EC 2.4.1.15]	−0.983**	−0.961**	0.968**	0.690	−0.690

(Continued)

TABLE 2 | Continued

Gene name	Accession No.	KEGG annotation	D-fructose content	D-glucose content	L-malic acid content	Citric acid content	Sugar/organic acid ratio
	Solyc09g075060.3	beta-glucosidase 11-like [EC 2.4.1.15]	0.574	0.524	−0.603	−0.876*	0.799
	Solyc11g008720.3	beta-glucosidase 42 [EC 2.4.1.15]	0.022	−0.065	−0.045	0.535	−0.581
	Solyc11g071640.3	uncharacterized protein LOC101256554 isoform X2 [EC 2.4.1.15]	−0.948**	−0.904*	0.936**	0.838*	−0.796
TPS/TPP	Solyc01g005210.2	alpha, alpha-trehalose-phosphate synthase [EC 3.1.3.12]	−0.299	−0.371	0.230	0.846*	−0.846*
	Solyc02g071590.3	alpha, alpha-trehalose-phosphate synthase [EC 3.1.3.12]	−0.943**	−0.945**	0.947**	0.695	−0.709
	Solyc02g072150.3	probable alpha, alpha-trehalose-phosphate synthase [EC 3.1.3.12]	0.485	0.472	−0.567	−0.777	0.761
	Solyc03g007290.4	probable trehalose-phosphate phosphatase 2 [EC 3.1.3.12]	−0.006	0.123	0.162	−0.614	0.662
	Solyc03g083960.3	trehalose-phosphate phosphatase A [EC 3.1.3.12]	0.260	0.441	−0.048	−0.669	0.772
	Solyc04g054930.3	probable trehalose-phosphate phosphatase J [EC 3.1.3.12]	−0.819*	−0.768	0.914*	0.331	−0.331
	Solyc04g072920.4	probable trehalose-phosphate phosphatase J [EC 3.1.3.12]	0.333	0.510	−0.132	−0.708	0.792
	Solyc06g060600.3	probable trehalose-phosphate phosphatase F [EC 3.1.3.12]	−0.918**	−0.922**	0.937**	0.603	−0.627
	Solyc07g006500.3	trehalose-6-phosphate synthase [EC 3.1.3.12]	−0.895*	−0.946**	0.858*	0.783	−0.823*
	Solyc07g062140.3	trehalose-phosphate synthase 1 [EC 3.1.3.12]	0.005	0.204	0.204	−0.476	0.591
	Solyc08g079060.4	probable trehalose-phosphate phosphatase F [EC 3.1.3.12]	−0.986**	−0.971**	0.919**	0.838*	−0.828*
HXK	Solyc02g091830.3	probable hexokinase-like 2 protein [EC 2.7.1.1]	−0.243	−0.081	0.493	−0.315	0.382
	Solyc04g081400.3	plastidic hexokinase [EC 2.7.1.1]	−0.315	−0.134	0.597	−0.031	0.134
	Solyc11g065220.2	hexokinase-3-like [EC 2.7.1.1]	−0.090	0.100	0.302	−0.430	0.533
	Solyc12g008510.2	Hexokinase [EC 2.7.1.1]	−0.553	−0.471	0.365	0.605	−0.520
ALMT	Solyc01g096140.3	Aluminum-activated malate transporter	−0.870*	−0.848*	0.752	0.943**	−0.905*
	Solyc11g068970.2		−0.537	−0.510	0.409	0.0864*	−0.807
ICDH	Solyc01g005560.3	isocitrate dehydrogenase [EC 2.3.3.8]	−0.058	−0.095	−0.057	0.681	−0.649
CS	Solyc07g055840.3	citrate synthase [EC 2.3.3.1]	−0.464	−0.431	0.304	0.828*	−0.751
ACS	Solyc12g099260.2	ATP-citrate synthase alpha chain protein [EC 1.1.1.42]	−0.832*	−0.772	0.937**	0.382	−0.370
PDHB-1	Solyc05g024160.3	pyruvate dehydrogenase E1 component subunit beta-1 [EC 1.2.4.1]	−0.783	−0.698	0.666	0.776	−0.697
SDH	Solyc04g055020.2 Solyc04g055030.2	succinate dehydrogenase [ubiquinone] iron-sulfur subunit 3 [EC 1.3.5.1]	−0.673	−0.632	0.503	0.845*	−0.782
			−0.788	−0.748	0.629	0.858*	−0.806

**Significantly correlated at the 0.05 level (bilateral). **Significantly correlated at the 0.01 level (bilateral).*

(GO:0009635), photosynthesis, light reaction (GO:0019684), photosynthesis (GO:0015979), ATP synthesis-coupled electron transport (GO:0042773) and transcription, and DNA templated (GO:0006351) in the biological process category were shared in TM-1 vs. TM-38 at the three developmental stages. In the cellular component category, GO terms of chloroplast thylakoid membrane (GO:0009535), photosystem II (GO:0009523), chloroplast (GO:0009507), proton-transporting ATP synthase complex, catalytic core F(1) (GO:0045261), plasma membrane (GO:0005886), photosystem II reaction center (GO:0009539), and mitochondrial small ribosomal subunit (GO:0005763) were observed in TM-1 vs. TM-38 at the three developmental stages. In the molecular function category, electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity (GO:0045156), chlorophyll-binding (GO:0016168), proton-transporting ATP synthase activity, rotational mechanism (GO:0046933), proton-transporting ATPase activity, rotational mechanism (GO:0046961), oxygen evolving activity (GO:0010242), quinone binding (GO:0048038), iron ion binding (GO:0005506), NADH dehydrogenase (ubiquinone) activity (GO:0008137), DNA-directed 5′–3′ RNA polymerase activity (GO:0003899), and

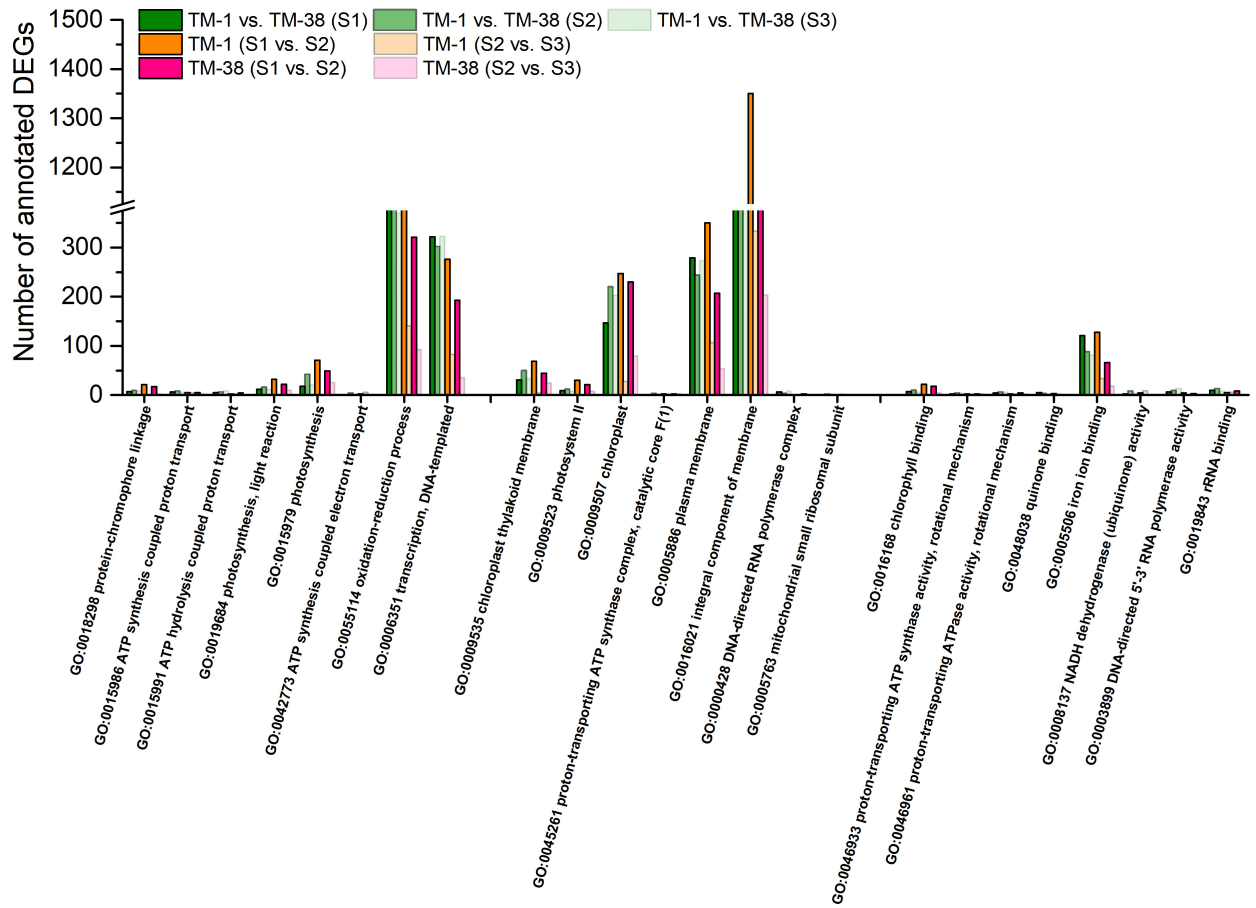


FIGURE 4 | Fisher's exact test for the significant (top eight) enrichment analysis of Gene Ontology (GO) terms of the biological process, cellular components, and molecular function categories of the annotated DEGs in TM-1 and TM-38 at three developmental stages.

rRNA binding (GO:0019843) were also shared in TM-1 vs. TM-38 at the three developmental stages.

Kyoto Encyclopedia of Genes and Genomes analysis was performed to further systematically understand the molecular interactions among the DEGs, the top 12 KEGG pathways with a p-value ≤ 0.01 were found to be significantly enriched (**Figure 5**). The significantly enriched KEGG pathways of CHO metabolism (ko00500) were shared in TM-1 vs. TM-38 at the three developmental stages. In addition, although the TCA cycle metabolism (ko00020) was not significantly enriched, it has been reported to be closely related to organic acid metabolism (Umer et al., 2020). Therefore, the CHO and TCA metabolisms were selected for subsequent analysis.

Transcriptional Expression Analysis of Differentially Expressed Genes Involved in Carbohydrate and TCA Cycle Metabolisms

The transcriptional expression of CHO metabolism genes in the TM-1 and TM-38 fruits at the three developmental stages was investigated by preparing heat maps (**Figure 6**). Two *SUSs* (Solyc03g098290.4 and Solyc07g042550.3), one *SPS* (Solyc08g042000.3), three *SPs* (Solyc02g077680.4, Solyc03g065340.3, and Solyc05g012510.3), one soluble starch synthase (*SSs*; Solyc03g083090.4), two beta-amylase (*BAMs*; Solyc08g077530.3 and Solyc09g091030.3), two ADP-glucose pyrophosphorylase large subunit (*AGPs*; Solyc01g109790.3 and Solyc07g019440.3), three beta-glucosidase (*BGL2s*; Solyc06g005970.2, Solyc06g076780.3, and Solyc09g075060.3), one trehalose-phosphate synthase (*TPS*; Solyc02g072150.3), three trehalose-phosphate phosphatases (*TPPs;* Solyc03g007290.4, Solyc03g083960.3, and Solyc04g072920.4), and two hexokinases [*HXKs*; (Solyc02g091830.3 and Solyc11g065220.2] were significantly upregulated in the TM-38 fruits relative to that in the TM-1 fruits at S1, S2, and S3. Nevertheless, one *SPS* (Solyc09g092130.3), four *INVs* (Solyc03g083910.5, Solyc09g010080.3, Solyc09g010090.5.1, and Solyc10g083290.4), two alpha-amylase (*AAMs*; Solyc04g078930.4 and Solyc04g082090.3), one *BAM* (Solyc01g094580.3), one 4-alpha-glucanotransferase (*4-α-GTase*; Solyc02g020980.2.1), seven *BGL2s* (Solyc01g081170.3, Solyc04g015560.4, Solyc06g073740.3, Solyc06g073750.4, Solyc06g073760.3, Solyc07g063390.3, Solyc07g063880.4, and Solyc11g071640.3), three *TPSs*

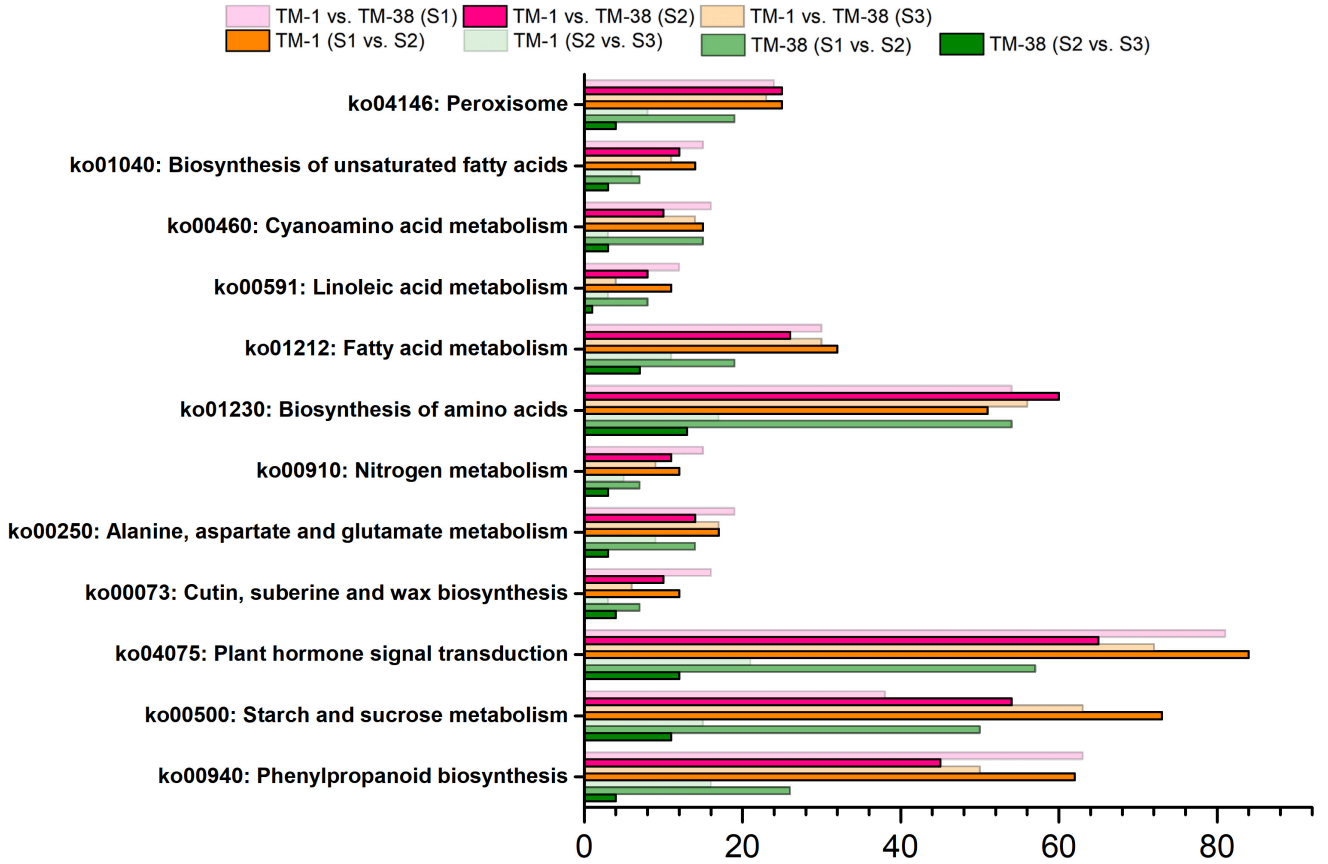


FIGURE 5 | Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the annotated DEGs in TM-1 and TM-38 at three developmental stages.

(Solyc01g005210.2, Solyc02g071590.3, and Solyc07g006500.3), and two *TPPs* (Solyc06g060600.3 and Solyc08g079060.4) were significantly downregulated in the TM-38 fruits relative to that in the TM-1 fruits at S1, S2, and S3. As for the expression levels of TCA metabolism genes, including two aluminum-activated malate transporter (*ALMT*; Solyc01g096140.3 and Solyc11g068970.2), one isocitrate dehydrogenase (*ICDH*; Solyc01g005560.3), two citrate synthase (*CS*; Solyc07g055840.3 and Solyc12g099260.2), one pyruvate dehydrogenase E1 component subunit beta-1 (*PDHB-1*; Solyc05g024160.3), and two *SDH* (Solyc04g055020.2 and Solyc04g055030.2) were significantly downregulated in the TM-38 fruits relative to that in the TM-1 fruits at the three developmental stages (**Figure 7**).

Key Candidate Differentially Expressed Genes Responsible for Organic Acids and Carbohydrate Accumulation in the TM-1 and TM-38 Fruits

To further survey the relationship between genes related to sugar metabolism and the synthesis of organic acids and carbohydrates, correlation analysis was carried out between the transcriptional expression levels of CHO and TCA metabolism genes with the content of organic acids and sugars in the TM-1 and TM-38 fruits at the three developmental stages (**Table 2**). The results showed that *INV* (Solyc09g010090.5.1 and Solyc10g083290.4), *BGL1* (Solyc10g045240.2), *AAM* (Solyc04g078930.4), *4-α-GTase* (Solyc02g020980.2.1), *BGL2* (Solyc01g010390.3, Solyc07g063880.4, and Solyc11g071640.3), *TPS* (Solyc02g071590.3 and Solyc07g006500.3), *TPP* (Solyc06g060600.3, Solyc08g079060.4), *ALMT* (Solyc01g096140.3), and *ACS* (Solyc12g099260.2) were negatively (r = −0.819 to −0.986) correlated with D-fructose and D-glucose contents, whereas they positively (r = 0.858–0.973) correlated with L-malic acid content. In addition, the content of citric acid was positively (r = 0.858–0.973) correlated with *INV* (Solyc09g010080.3 and Solyc09g010090.5.1), *AAM* (Solyc04g082090.3), *4-α-GTase* (Solyc02g020980.2.1), *BGL2* (Solyc01g081170.3, Solyc06g073750.4, Solyc06g073760.3, Solyc07g063390.3, and Solyc11g071640.3), *TPS* (Solyc01g005210.2), *TPP* (Solyc08g079060.4), *ALMT* (Solyc01g096140.3 and Solyc11g068970.2), *CS* (Solyc07g055840.3), and *SDH* (Solyc04g055020.2 and Solyc04g055030.2), whereas negatively (r = 0.858–0.973) correlated with *SUS* (Solyc07g042550.3), *BAM*, (Solyc08g077530.3), *AGP* (Solyc07g019440.3), and *BGL2* (Solyc09g075060.3). The sugar/organic acid ratio was positively (r = 0.885–0.931) correlated with

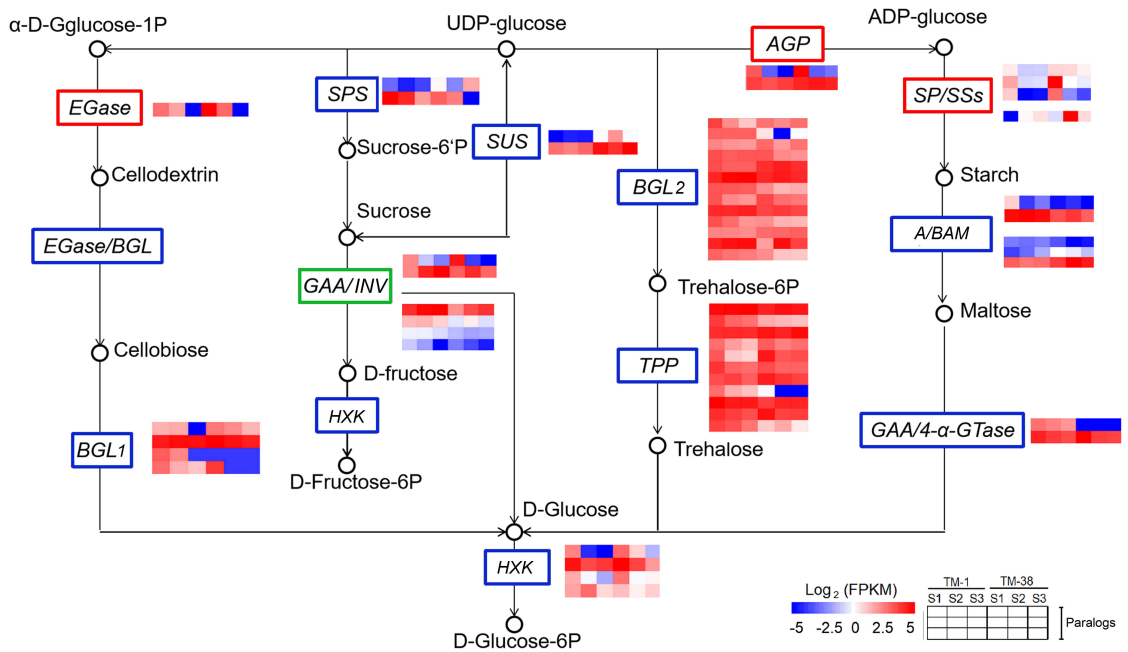


FIGURE 6 | Differential expression of genes involved in starch and sucrose metabolism pathway in TM-1 and TM-38 at three developmental stages. Heat maps depict the normalized gene expression values, which represent the means ± SD of three biological replicates. Expression values of 18 libraries are presented as FPKM normalized $\log_2$-transformed counts.

SUS (Solyc07g042550.3) and *BAM* (Solyc08g077530.3), while negatively (r = −0.823 to −0.918) correlated with *INV* (Solyc09g010080.3 and Solyc09g010090.5.1), *AAM* (Solyc04g082090.3), *4-α-GTase* (Solyc02g020980.2.1), *BGL2* (Solyc06g073750.4, Solyc06g073760.3, and Solyc01g081170.3), *TPS* (Solyc01g005210.2 and Solyc07g006500.3), *TPP* (Solyc08g079060.4), and *ALMT* (Solyc01g096140.3).

To further search for candidate genes with major contributions within the complex regulatory networks, the annotation information of all these genes was extracted from the tomato reference genome database. A total of 12 previously used DEGs, including eight genes linked to CHO metabolism, and four genes linked to TCA metabolism shown in **Figures 6**, **7** were selected for qRT-PCR expression level detection. The general trend of relative expression levels in the three stages was consistent with the in-depth sequencing (**Figure 8**).

DISCUSSION

Sugar metabolism plays a key role in plant development, stress response, and yield, which is closely linked to sugar signaling (Ruan, 2014). This coupling is achieved by the production of sugar signaling molecules, such as sucrose, glucose, fructose, and trehalose-6-phosphate, or possibly through signal transduction in the metabolic process itself (O'Hara et al., 2013). Sweetness is particularly appreciated in industrial tomatoes, usually related to fructose and glucose concentrations, mainly accumulated in the vacuoles of the fruit cells (Bona et al., 2016). Although tomato flavor characteristics are the result of complex interactions among multiple metabolites (including more than 400 volatile compounds), the sugar/acid ratio is a key determinant of the sensory experience, which leads to repeated purchase and consumer loyalty (Casals et al., 2011; Beckles, 2012; Hart et al., 2015; Wang et al., 2015). The major organic acids in tomatoes are

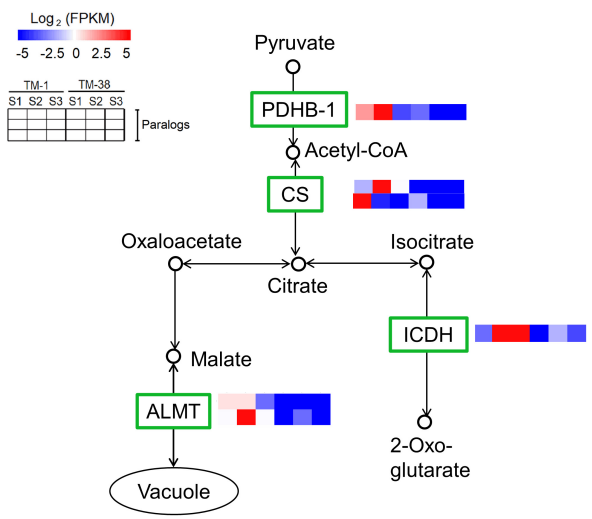


FIGURE 7 | Differential expression of genes involved in citrate cycle (TCA cycle) pathway in TM-1 and TM-38 at three developmental stages. Heat maps depict the normalized gene expression values, which represent the means ± SD of three biological replicates. Expression values of 18 libraries are presented as FPKM normalized $\log_2$-transformed counts.

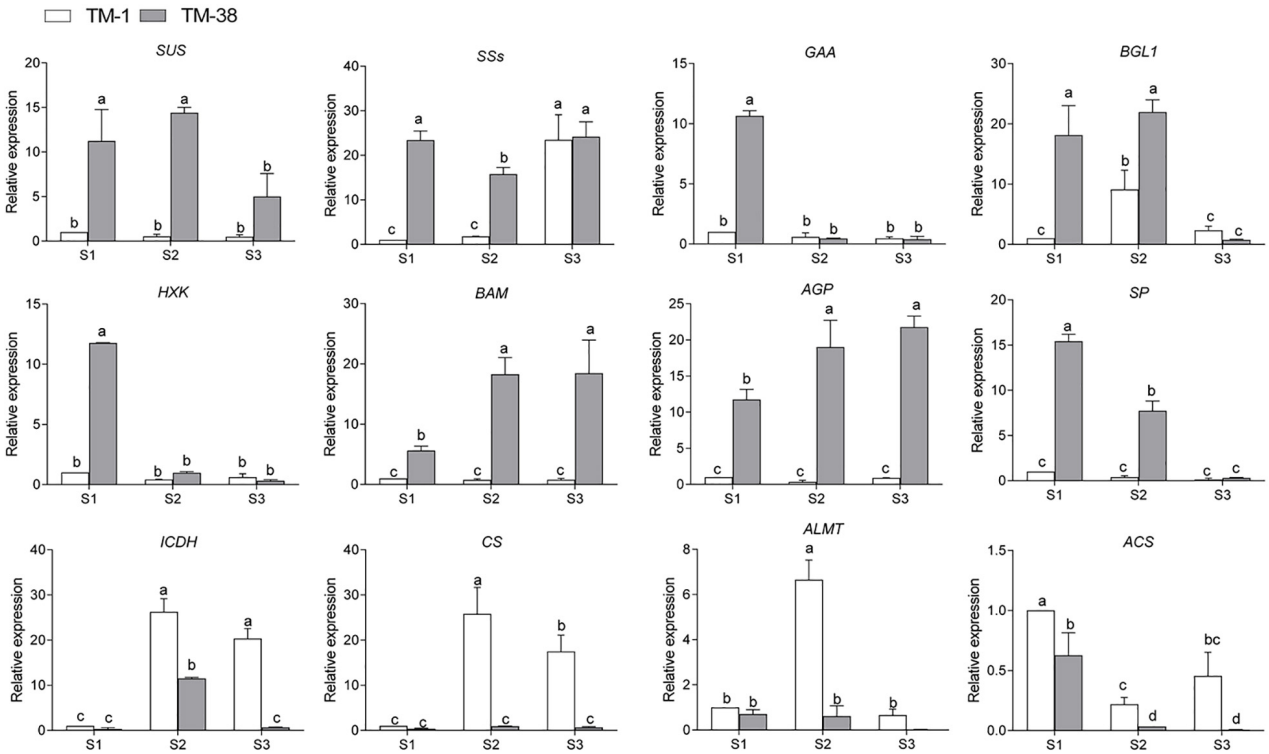


FIGURE 8 | qRT-PCR assay, the mean was calculated from three biological replicates, each with three technical replicates (n = 9). These replicates were then normalized relative to the expression of *Actin*.

citric and malic acids, with citric predominating (Marconi et al., 2007). The purpose of this research is to analyze the sugar and organic acid profiles of the TM-1 and TM-38 tomato fruits at three developmental stages, and to characterize genes involved in sugar metabolism.

As the tomato flavor intensity is largely influenced by the interaction of reducing sugars (glucose, fructose) and organic acids (citric and malic acid), which together represent about 60–65% of the dry matter (Bucheli et al., 1999) and the sugar/acid ratio is a key determinant of the sensory experience (Beckles, 2012). In the present research, the fruit Brix of TM-1 (7.6 and 5.7% in 2019 and 2020) was nearly twice that of TM-38 (3.8 and 3.9% in 2019 and 2020) at S3. However, the contents of D-fructose and D-glucose in the TM-38 fruits were significantly higher than those in the TM-1 fruits at S1 and S2, reaching the same level at S3, while the contents of L-malic and citric acid showed an opposite trend at the three developmental stages (**Figure 2**). The sugar/organic acid ratios of the TM-38 fruits were undoubtedly significantly higher than that of the TM-1 fruits at the three developmental stages. Since TM-38 cultivar is the product of artificial continuous breeding and artificial sweet selection, the sugar/acid ratio of the TM-38 cultivar is higher than that of the wild cultivar TM-1, but it also leads to the loss of flavor substances in TM-1, which eventually leads to a less palatable variety of TM-38. To sum up, these results indicated that the contribution of both L-Malic acid and citric acid to the fruit Brix was much greater than that of glucose and fructose in tomato fruit, or equivalent. Besides, both tomato genotypes had low sucrose content ($<$ 20 μg/g) at all three developmental stages (Data Sheet 1), which may be due to sucrose, which is the photoassimilate transported from the leaves to the fruit of tomato, yet the fruit accumulates predominantly glucose and fructose (Chengappa et al., 1999).

SUS is a highly regulated cytosolic enzyme that catalyzes the reversible conversion of sucrose and UDP into UDP-glucose and fructose (Coleman et al., 2009). Overexpression of *NtSUS3* accelerated the hydrolysis of sucrose and increased fructose content in *Nicotiana tabacum* L. (Daloso et al., 2016). Likewise, two *SUSs* (Solyc03g098290.4 and Solyc07g042550.3) were significantly upregulated in fruits of TM-38 relative to that in TM-1 at the three developmental stages (**Figure 6** and Data Sheet 3). Among which, *SUS* (Solyc07g042550.3) was positively (r = 0.885) correlated with the sugar/organic acid ratio (**Table 2**). Thus, we reply that *SUS* (Solyc07g042550.3) contributes to the accumulation of glucose and fructose.

INV is related to the irreversible hydrolysis of sucrose into glucose and fructose. According to the cell location, INV can be classified into cell-wall invertase (CWIN), vacuolar invertase (VIN), and cytoplasmic invertase (CIN) (Wan et al., 2018). A *CWIN* gene named *LIN5* in tomato (*S. lycopersicum* L.) was identified as positively associated with sugar accumulation (Fridman et al., 2004), and in another report, the higher expression of *PlCWIN1* may promote the increase of soluble

sugar content (Xue et al., 2019). In sugarcane (*Saccharum* spp. hybrid), leaf chemical formulation *vis-à-vis* reduced soluble acid invertase (SAI) activity and increased sucrose content, which may be due to the significant negative correlation between *SAI* gene and non-reducing sugar (sucrose) in low- and high-sugar genotypes (Jain et al., 2017). Based on the present study, four *INVs* (Solyc03g083910.5, Solyc09g010080.3, Solyc09g010090.5.1, and Solyc10g083290.4), were significantly downregulated in fruits of TM-38 relative to that in TM-1 at the three stages (**Figure 6** and Data Sheet 3). Among which, *INVs* (Solyc09g010090.5.1 and Solyc10g083290.4) were negatively (r = −0.827 to −0.970) correlated with D-fructose and D-glucose contents, whereas *INV* (Solyc10g083290.4) was positively (r = 0.973) correlated with L-malic acid content (**Table 2**). *INVs* (Solyc09g010080.3 and Solyc09g010090.5.1) were positively (r = 0.940) correlated with citric acid content and negatively (r = −0.881 to −0.907) correlated with the sugar/organic acid ratio. Therefore, *INV* (Solyc10g083290.4) and *INVs* (Solyc09g010080.3 and Solyc09g010090.5.1) contributed to L-malic acid and citric acid accumulation, respectively. Citric acid may play a more dominant role in the sugar/organic acid ratio.

Plant starch can be synthesized and degraded by several enzyme reactions (Emes et al., 2003; Zeeman et al., 2010). The main source of AGP activity is glucosyl donor ADPG for starch biosynthesis, starch synthase (SS), including SSs and granule-bound SS (GBSS) add glucosyl units at the non-reducing end of linear chains through new α (1→4) linkages (Baroja-Fernandez et al., 2001). BMY is an exoamylase that hydrolyzes a-1,4 glycosidic linkages of polyglucan chains at the non-reducing end to produce maltose during hydrolytic starch degradation (Kaplan et al., 2006; Zanella et al., 2016). Once maltose is exported to the cytosol, it is further metabolized to glucose and/or sucrose and maltodextrins by the activity of cytosolic glucosyltransferases (Tachibana et al., 1997). The high level of gene expression associated with starch degradation (*AAM*, *BAM*, and *SP*) indicated that starch degradation might be a positive process to ensure the sweetness of chestnuts harvest (Zhang et al., 2015). *Arabidopsis* leaves of osmotically stressed *bam1* accumulated more starch and less soluble sugar during the day than wild-type and *bam3* (Zanella et al., 2016). Three *SPs* (Solyc02g077680.4, Solyc03g065340.3, and Solyc05g012510.3), one *SS* (Solyc03g083090.4), two *BAM*s (Solyc08g077530.3 and Solyc09g091030.3), and two *AGP*s (Solyc01g109790.3 and Solyc07g019440.3) were significantly upregulated in fruits of TM-38 relative to that in TM-1 at the three developmental stages. The upregulated expression of these genes was considered to be beneficial to starch degradation and sugar accumulation, while *AAM* (Solyc04g082090.3) was on the contrary, which was negatively (r = −0.918) correlated with sugar/organic acid ratio.

Plant glucose can also be synthesized and phosphorylated by several enzyme reactions. For instance, 4-α-GTase (EC 2.4.1.25) could react each maltooligosaccharides (from maltose to maltoheptaose) as the effective substrate to form glucose and various maltooligosaccharides (Tachibana et al., 1997). Likewise, trehalose 6-phosphate is synthesized from UDPG and glucose 6-phosphate *via* TPS, and then dephosphorylated by TPP to yield trehalose, which can be hydrolyzed by trehalase to glucose (Yadav et al., 2014). HXK (EC 2.7.1.1) catalyzes the phosphorylation of hexoses, such as D-glucose, D-fructose, and D-mannose, to hexose 6-phosphate (D-glucose 6-phosphate, D-fructose 6-phosphate, and D-mannose 6-phosphate, respectively) (Menu et al., 2001; Granot, 2007). Three *BGL2s* (Solyc06g005970.2, Solyc06g076780.3, and Solyc09g075060.3), one TPS (Solyc02g072150.3), and three *TPPs* (Solyc03g007290.4, Solyc03g083960.3, and Solyc04g072920.4) were significantly upregulated in fruits of TM-38 relative to that in TM-1 at the three stages and were believed to play a positive role in the synthesis of glucose. Two *HXKs* (Solyc02g091830.3 and Solyc11g065220.2) were also significantly upregulated in fruits of TM-38 relative to that in TM-1 at the three stages and were believed to play an active role in glucose phosphorylation. Besides, *4-α-GTase* (Solyc02g020980.2.1), *BGL2* (Solyc06g073750.4, Solyc06g073760.3, and Solyc01g081170.3), *TPS* (Solyc01g005210.2 and Solyc07g006500.3), and *TPP* (Solyc08g079060.4), which were negatively (r = −0.823 to −0.918) correlated with the sugar/organic acid ratio, were presumed to have a negative regulatory effect on D-glucose synthesis.

Overall variations in sugar and/or organic acid content are complex metabolic traits that are regulated by gene networks (Famiani et al., 2015). *ALMT*-family genes have been reported to regulate organic acid contents, for example, four *AtALMT9* homologs in grape berries (Terrier et al., 1998), *ALMT II* in apple (Ma et al., 2015), and *ALMT7* in watermelon (Umer et al., 2020). According to these literature data and our findings, the expression level of *ALMT* (Solyc01g096140.3) was positively correlated with citric acid concentration (r = 0.943) and the sugar/organic acid ratio (r = −0.905) (**Table 2**), indicating that it is likely to be a key candidate in the gene network of organic acid biosynthesis that contributes to the maximum trait variation.

The characteristics of these genes will improve our understanding of the molecular mechanism of sugar and organic acid biosynthesis. Finally, to evaluate the relative expression level of key putative genes involved in CHO and TCA metabolism, 12 candidate genes including *SUS*, *SSs*, G*AA*, *BGL1*, *HXK*, *BAM*, *AGP*, *SP*, *ICDH*, *CS*, and *ALMT* were selected for qRT-PCR analysis (**Figure 8**). The results showed the accuracy of transcriptome sequencing. The data collected in this study established a foundation for further investigations to evaluate the structures and functions of the abovementioned genes using molecular biology techniques in the fruit quality of commercially important plants.

CONCLUSION

We have investigated the tomato fruit transcriptome of TM-1 (*S. galapagense* L., LA0436) and TM-38 (*S. lycopersicum* L. cultivar M82, LA3475) at three developmental stages and identified specific processes that lead to the variation of fruit Brix. The results indicated that citric acid may play a more dominant role in the sugar/organic acid ratio of tomato fruit, and the contribution of both L-malic acid and citric acid to the fruit Brix

was much greater than that of glucose and fructose. The CHO metabolism (ko00500) genes, including *SUS* (Solyc07g042550.3) and *BAM* (Solyc08g077530.3), which were positively (r = 0.885–0.931) correlated with the sugar/organic acid ratio, as well as *INV* (Solyc09g010080.3 and Solyc09g010090.5.1), *AAM* (Solyc04g082090.3), *4-α-GTase* (Solyc02g020980.2.1), *BGL2* (Solyc06g073750.4, Solyc06g073760.3, and Solyc01g081170.3), *TPS* (Solyc01g005210.2 and Solyc07g006500.3), *TPP* (Solyc08g079060.4), and TCA cycle metabolism (ko00020) gene of *ALMT* (Solyc01g096140.3), which were negatively (r = −0.823 to −0.918) correlated with the sugar/organic acid ratio were considered to be the contributing factors of fruit Brix.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI Sequence Read Archive (SRA) under the accession number PRJNA744374.

AUTHOR CONTRIBUTIONS

HL and YQ designed the research, and reviewed the manuscript. NL and JW collected the experimental data, drafted the manuscript, and carried out the experiments with the help of BW, SH, JH, TY, and PA. All authors read and approved the final manuscript.

FUNDING

This work was financially supported by Special Incubation Project of Science & Technology Renovation of Xinjiang Academy of Agricultural Sciences (xjkcpy-2021001), the National Natural Science Foundation of China (31860555), the National Natural Science Foundation of China (31760581), and the China Postdoctoral Science Foundation (2019M663862). The funders have no role in the study design, data analysis, and manuscript writing, but just provided the financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.714942/full#supplementary-material

Supplementary Table 1 | MS Parameters-UHPLC-HRMS.

Supplementary Table 2 | Sequence of primers used for qRT-PCR analysis.

Supplementary Table 3 | Statistical table of the number of annotated DEGs in TM-1 vs. TM-38 at three stages.

Supplementary Data Sheet 1 | The contents of organic acids and carbohydrates in fruit samples at three stages.

Supplementary Data Sheet 2 | The calibration curve of L-Malic acid, Citric acid, D-Fructose, D-Glucose, and D-Sucrose.

Supplementary Data Sheet 3 | Gene composition and mean FPKM value of the carbohydrate (ko00500) and citrate cycle (TCA cycle, ko00020) metabolism.

REFERENCES

Barickman, T. C., Kopsell, D. A., and Sams, C. E. (2016). Abscisic acid impacts tomato carotenoids, soluble sugars, and organic acids. *HortScience* 51, 370–376. doi: 10.21273/hortsci.51.4.370

Baroja-Fernandez, E., Munoz, F. J., Akazawa, T., and Pozueta-Romero, J. (2001). Reappraisal of the currently prevailing model of starch biosynthesis in photosynthetic tissues: a proposal involving the cytosolic production of ADP-glucose by sucrose synthase and occurrence of cyclic turnover of starch in the chloroplast. *Plant Cell Physiol.* 42, 1311–1320. doi: 10.1093/pcp/pce175

Beckles, D. M. (2012). Factors affecting the postharvest soluble solids and sugar content of tomato (*Solanum lycopersicum* L.) fruit. *Postharvest. Biol. Tec.* 63, 129–140. doi: 10.1016/j.postharvbio.2011.05.016

Beckles, D. M., Hong, N., Stamova, L., and Luengwilai, K. (2011). Biochemical factors contributing to tomato fruit sugar content: a review. *Fruits* 67, 49–64. doi: 10.1051/fruits/2011066

Bona, E., Cantamessa, S., Massa, N., Manassero, P., Marsano, F., Copetta, A., et al. (2016). Arbuscular mycorrhizal fungi and plant growth-promoting pseudomonads improve yield, quality and nutritional value of tomato: a field study. *Mycorrhiza* 27, 1–11. doi: 10.1007/s00572-016-0727-y

Bucheli, P., Voirol, E., de la Torre, R., López, J., Rytz, A., Tanksley, S. D., et al. (1999). Definition of biochemical and molecular markers (quality trait loci) for tomato flavour as tools in breeding. *Acta Hortic.* 487, 301–306. doi: 10.17660/actahortic.1999.487.46

Carlos, A., Sabine, V. T., Brigitte, P., and Wilfried, R. (2018). Quantification of sugars and organic acids in tomato fruits. *Methods* 5, 537–550. doi: 10.1016/j.mex.2018.05.014

Casals, J., Pascual, L., Cañizares, J., Cebolla-Cornejo, J., Casañas, F., and Nuez, F. (2011). The risks of success in quality vegetable markets: possible genetic erosion in *Marmande tomatoes* (*Solanum lycopersicum* L.) and consumer dissatisfaction. *Sci. Hortic.* 130, 78–84. doi: 10.1016/j.scienta.2011.06.013

Chengappa, S., Guilleroux, M., Phillips, W., and Shields, R. (1999). Transgenic tomato plants with decreased sucrose synthase are unaltered in starch and sugar accumulation in the fruit. *Plant Mol. Biol.* 40, 213–221.

Çolaka, N. G., Ekena, N. T., Ülgerb, M., Frarya, A., and Doğanlara, S. (2020). Exploring wild alleles from *Solanum pimpinellifolium* with the potential to improve tomato flavor compounds. *Plant Sci.* 298:110567. doi: 10.1016/j.plantsci.2020.110567

Coleman, H. D., Yan, J., and Mansfield, S. D. (2009). Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultrastructure. *Proc. Natl. Acad. Sci. U S A.* 106:13118. doi: 10.1073/pnas.0900188106

Daloso, D. M., Williams, T. C. R., Antunes, W. C., Pinheiro, D. P., Mueller, C., Loureiro, M. E., et al. (2016). Guard cell-specific upregulation of sucrose synthase 3 reveals that the role of sucrose in stomatal function is primarily energetic. *New Phytol.* 209, 1470–1483. doi: 10.1111/nph.13704

Dorais, M., Ehret, D. L., and Papadopoulos, A. P. (2008). Tomato (*Solanum lycopersicum*) health components: from the seed to the consumer. *Phytochem. Rev.* 7, 231–250. doi: 10.1007/s11101-007-9085-x

Emes, M. J., Bowsher, C. G., Hedley, C., Burrell, M. M., Scrase-Field, E. S., and Tetlow, I. J. (2003). Starch synthesis and carbon partitioning in developing endosperm. *J. Exp. Bot.* 54, 569–575. doi: 10.1093/jxb/erg089

Famiani, F., Battistelli, A., Moscatello, S., Cruz-Castillo, J. G., and Walker, R. P. (2015). The organic acids that are accumulated in the flesh of fruits: occurrence, metabolism and factors affecting their contents-a review. *Rev. Chapingo Ser. Horticultura* 21, 97–128. doi: 10.5154/r.rchsh.2015.01.004

Finlay, M., Dale, B., and Bradshaw, J. E. (2003). Progress in improving processing attributes in potato. *Trends Plant Sci.* 8, 310–312. doi: 10.1016/s1360-1385(03)00130-4

Fridman, E., Carrari, F., Liu, Y. S., Fernie, A. R., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* 305, 1786–1789. doi: 10.1126/science.1101666

Gifford, R. M., Thorne, J. H., Hitz, W. D., and Giaquinta, R. T. (1984). Crop productivity and photoassimilate partitioning. *Science* 225, 801–808. doi: 10.1126/science.225.4664.801

Giovannetti, M., Avio, L., Barale, R., Ceccarelli, N., Cristofani, R., Iezzi, A., et al. (2012). Nutraceutical value and safety of tomato fruits produced by mycorrhizal plants. *Br. J. Nutr.* 107, 242–251. doi: 10.1017/s000711451100290x

Granot, D. (2007). Role of tomato hexose kinases. *Funct. Plant Biol.* 34, 564–570. doi: 10.1071/fp06207

Grof, C. P. L., Knight, D. P., McNeil, S. D., Lunn, J. E., and Campbell, J. A. (1998). A modified assay method shows leaf sucrose-phosphate synthase activity is correlated with leaf sucrose content across a range of sugarcane varieties. *Aust. J. Plant Physiol.* 25, 499–502. doi: 10.1071/pp97169

Hart, M., Ehret, D. L., Krumbein, A., Leung, C., Murch, S., Turi, C., et al. (2015). Inoculation with arbuscular mycorrhizal fungi improves the nutritional value of tomatoes. *Mycorrhiza* 25, 359–376. doi: 10.1007/s00572-014-0617-0

Hashida, Y., Hirose, T., Okamura, M., Hibara, K., and Ohsugi, R. N. A. (2016). A reduction of sucrose phosphate synthase (SPS) activity affects sucrose/starch ratio in leaves but does not inhibit normal plant growth in rice. *Plant Sci.* 253, 40–49. doi: 10.1016/j.plantsci.2016.08.017

Jain, R., Singh, S. P., Singh, A., Singh, S., Kishor, R., Singh, R. K., et al. (2017). Soluble acid invertase (SAI) activity and gene expression controlling sugar composition in sugarcane. *Sugar Tech.* 19, 669–674. doi: 10.1007/s12355-017-0511-0

Jawad, U. M., Gao, L., Gebremeskel, H., Safdar, L. B., Yuana, P., Zhao, S., et al. (2020). Expression pattern of sugars and organic acids regulatory genes during watermelon fruit development. *Sci. Hortic.* 265:109102. doi: 10.1016/j.scienta.2019.109102

Jie, Z., Huang, C., Yang, B., Heikki, K., Liu, P., and Ou, S. (2018). Regulation of phytochemicals in fruits and berries by environmental variation-sugars and organic acids. *J. Food Biochem.* 43:e12642. doi: 10.1111/jfbc.12642

Kader, A. A. (2008). Flavor quality of fruits and vegetables. *J. Sci. Food Agric.* 88, 1863–1868.

Kandel-Kfir, M., Damari-Weissler, H., German, M. A., Gidoni, D., Mett, A., Belausov, E., et al. (2006). Two newly identified membrane-associated and plastidic tomato HXKs: characteristics, predicted structure and intracellular localization. *Planta* 224, 1341–1352. doi: 10.1007/s00425-006-0318-9

Kaplan, F., Sung, D. Y., and Guy, C. L. (2006). Roles of β-amylase and starch breakdown during temperatures stress. *Physiol. Plantarum* 126, 120–128. doi: 10.1111/j.1399-3054.2006.00604.x

Koblitz, H. (1991). "Protoplast culture and somatic hybridization in *Lycopersicon*," in *Genetic Improvement of Tomato*, ed. G. Kalloo (Berlin: Springer).

Krook, J., Vreugdenhil, D., Dijkema, C., and van der Plas, L. H. W. (1998). Sucrose and starch metabolism in carrot (*Daucus carota* L.) cell suspensions analysed by 13C-labelling: indications for acytosol and a plastid-localized oxidative pentose phosphate pathway. *J. Exp. Bot.* 49, 1917–1924. doi: 10.1093/jxb/49.329.1917

Langlois, D., Etievant, P. X., Pierron, P., and Jorrot, A. (1996). Sensory and instrumental characterization of commercial tomato varieties. *Z Lebensm-Unters Forsch.* 203, 534–540. doi: 10.1007/bf01193159

Lastdrager, J., Hanson, J., and Smeekens, S. (2014). Sugar signals and the control of plant growth and development. *J. Exp. Bot.* 65, 799–807. doi: 10.1093/jxb/ert474

Li, D., Mou, W., Wang, Y., Li, L., Mao, L., Ying, T., et al. (2016). Exogenous sucrose treatment accelerates postharvest tomato fruit ripening through the influence on its metabolism and enhancing ethylene biosynthesis and signaling. *Acta Physiol. Plant* 38:225.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Ma, B., Liao, L., Zheng, H., Chen, J., Wu, B., Ogutu, C., et al. (2015). Genes encoding aluminum-activated malate transporter II and their association with fruit acidity in apple. *Plant Genome* 8, 1–14.

Marconi, O., Floridi, S., and Montanari, L. (2007). Organic acids profile in tomato juice by HPLC with UV detection. *J. Food Qual.* 30, 43–56. doi: 10.1111/j.1745-4557.2007.00105.x

Menendez, C. M., Ritter, E., Schafer-Pregl, R., Walkemeier, B., Kalde, A., Salamini, F., et al. (2002). Cold sweetening in diploid potato: mapping quantitative trait loci and candidate genes. *Genetics* 162, 1423–1434. doi: 10.1093/genetics/162.3.1423

Menu, T., Rothan, C., Dai, N., Petreikov, M., Etienne, C., Destrac-Irvine, A., et al. (2001). Cloning and characterization of a cDNA encoding hexokinase from tomato. *Plant Sci.* 160, 209–218. doi: 10.1016/s0168-9452(00)00332-0

Ofner, I., Lashbrooke, J., Pleban, T., Aharoni, A., and Zamir, D. (2016). Solanum pennellii backcross inbred lines (BILs) link small genomic bins with tomato traits. *Plant J.* 87, 151–160. doi: 10.1111/tpj.13194

O'Hara, L. E., Paul, M. J., and Wingler, A. (2013). How do sugars regulate plant growth and development? new insight into the role of trehalose-6-phosphate. *Mol. Plant* 6, 261–274. doi: 10.1093/mp/sss120

Rick, C. M. (1974). High soluble-solids content in large-fruited tomato lines derived from a wild green-fruited species. *Hilgardia* 42, 493–510. doi: 10.3733/hilg.v42n15p493

Roby, C., Cortes, S., Gromova, M., LeBail, J. L., and Roberts, J. K. M. (2002). Sucrose cycling in heterotrophic plant cell metabolism: first step towards an experimental model. *Mol. Biol. Rep.* 29, 145–149.

Ruan, Y. L. (2014). Sucrose metabolism: gateway to diverse carbon use and sugar signaling. *Annu. Rev. Plant Biol.* 65, 33–67. doi: 10.1146/annurev-arplant-050213-040251

Schaffer, A. A., Petreikov, M., Miron, D., Fogelman, M., Spiegelman, M., Bnei-Moshe, Z., et al. (1999). Modification of carbohydrate content in developing tomato fruit. *HortScience* 34, 1024–1027. doi: 10.21273/hortsci.34.6.1024

Schauer, N., Zamir, D., and Fernie, A. R. (2005). Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. *J. Exp. Bot.* 56, 297–307. doi: 10.1093/jxb/eri057

Sweetlove, L., Müller-Röber, B., Willmitzer, L., and Hill, S. A. (1999). The contribution of adenosine 5-diphosphoglucose pyrophosphorylase to the content of starch synthesis in potato tubers. *Planta* 209, 330–337. doi: 10.1007/s004250050640

Tachibana, Y., Fujiwara, S., Takagi, M., and Imanaka, T. (1997). Cloning and expression of the 4-α-glucanotransferase gene from the hyperthermophilic archaeon *Pyrococcus sp*. KOD1, and characterization of the enzyme. *J. Fermentat. Bioeng.* 83, 540–548. doi: 10.1016/s0922-338x(97)81134-8

Terrier, N., Deguilloux, C., Sauvage, F. X., Martinoia, E., and Romieu, C. (1998). Proton pumps and anion transport in *Vitis vinifera*: the inorganic pyrophosphatase plays a predominant role in the energization of the tonoplast. *Plant Physiol. Biochem.* 36, 367–377. doi: 10.1016/s0981-9428(98)80078-8

Umer, M. J., Safdar, L. B., Gebremeskel, H., Zhao, S., Yuan, P., Zhu, H., et al. (2020). Identification of key gene networks controlling organic acid and sugar metabolism during watermelon fruit development by integrating metabolic phenotypes and gene expression profiles. *Hortic. Res.* 7:193.

Verma, S., Sharma, V., and Kumari, N. (2020). Microwave pretreatment of tomato seeds and fruit to enhance plant photosynthesis, nutritive quality and shelf life of fruit. *Postharvest Biol. Tec.* 159:111015. doi: 10.1016/j.postharvbio.2019.111015

Wan, H., Wu, L., Yang, Y., Zhou, G., and Ruan, Y. L. (2018). Evolution of sucrose metabolism: the dichotomy of invertases and beyond. *Trends Plant Sci.* 23, 163–177. doi: 10.1016/j.tplants.2017.11.001

Wang, L., Baldwin, E. A., Zhao, W., Plotto, A., Sun, X., Wang, Z., et al. (2015). Suppression of volatile production in tomato fruit exposed to chilling temperature and alleviation of chilling injury by a pre-chilling heat treatment. *LWT-Food Sci. Technol.* 62, 115–121. doi: 10.1016/j.lwt.2014.12.062

Weise, S. E., Weber, A. P., and Sharkey, T. D. (2004). Maltose is the major form of carbon exported from the chloroplast at night. *Planta* 218, 474–482. doi: 10.1007/s00425-003-1128-y

Wind, J., Smeekens, S., and Hanson, J. (2010). Sucrose: metabolite and signaling molecule. *Phytochemistry* 71, 1610–1614. doi: 10.1016/j.phytochem.2010.07.007

Xu, S., Sun, X., Lu, H., Yang, H., Ruan, Q., Huang, H., et al. (2018). Detecting and monitoring the flavor of tomato (*Solanum lycopersicum*) under the impact

of postharvest handlings by physicochemical parameters and electronic nose. *Sensors* 18:1847. doi: 10.3390/s18061847

Xue, J., Tang, Y., Wang, S., Xue, Y., Liu, X., and Zhang, X. (2019). Evaluation of dry and wet storage on vase quality of cut peony based on the regulation of starch and sucrose metabolism. *Postharvest Biol. Tec.* 155, 11–19. doi: 10.1016/j.postharvbio.2019.05.007

Yadav, U. P., Ivakov, A., Feil, R., Duan, G. Y., Walther, D., Giavalisco, P., et al. (2014). The sucrose–trehalose 6-phosphate (Tre6P) nexus: specificity and mechanisms of sucrose signalling by Tre6P. *J. Exp. Bot.* 65, 1051–1068. doi: 10.1093/jxb/ert457

Zanella, M., Borghi, G. L., Pirone, C., Thalmann, M., Pazmino, D., Costa, A., et al. (2016). β-amylase 1 (BAM1) degrades transitory starch to sustain proline biosynthesis during drought stress. *J. Exp. Bot.* 67, 1819–1826. doi: 10.1093/jxb/erv572

Zeeman, S. C., Kossmann, J., and Smith, A. M. (2010). Starch: its metabolism, evolution, and biotechnological modification in plants. *Annu. Rev. Plant Biol.* 61, 209–234. doi: 10.1146/annurev-arplant-042809-112301

Zhang, L., Lin, Q., Feng, Y., Fan, X., Zou, F., Yuan, D. Y., et al. (2015). Transcriptomic identification and expression of starch and sucrose metabolism genes in the seeds of chinese chestnut (*Castanea mollissima*). *J. Agric. Food Chem.* 63, 929–942. doi: 10.1021/jf505247d

Zhu, Z., Zhang, Y., Liu, J., Chen, Y., and Zhang, X. (2018). Exploring the effects of selenium treatment on the nutritional quality of tomato fruit. *Food Chem.* 252, 9–15. doi: 10.1016/j.foodchem.2018.01.064

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 01 October 2021
doi: 10.3389/fgene.2021.749675

Identification and Validation of High LD Hotspot Genomic Regions Harboring Stem Rust Resistant Genes on 1B, 2A (*Sr38*), and 7B Chromosomes in Wheat

*Shamseldeen Eltaher[1†], Amira M. I. Mourad[2†], P. Stephen Baenziger[3], Stephen Wegulo[4], Vikas Belamkar[3] and Ahmed Sallam[5]**

[1]Department of Plant Biotechnology, Genetic Engineering and Biotechnology Research Institute (GEBRI), University of Sadat City (USC), Sadat, Egypt, [2]Department of Agronomy, Faculty of Agriculture, Assiut University, Assiut, Egypt, [3]Department of Agronomy and Horticulture, University of Nebraska–Lincoln, Lincoln, NE, United States, [4]Department of Plant Pathology, University of Nebraska–Lincoln, Lincoln, NE, United States, [5]Department of Genetics, Faculty of Agriculture, Assiut University, Assiut, Egypt

OPEN ACCESS

Edited by:
Deepmala Sehgal,
International Maize and Wheat Improvement Center, Mexico

Reviewed by:
Vijay Gahlaut,
Institute of Himalayan Bioresource Technology (CSIR), India
Kumar Paritosh,
University of Delhi, India

***Correspondence:**
Ahmed Sallam
amsallam@aun.edu.eg

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 29 July 2021
Accepted: 13 September 2021
Published: 01 October 2021

Citation:
Eltaher S, Mourad AMI, Baenziger PS, Wegulo S, Belamkar V and Sallam A (2021) Identification and Validation of High LD Hotspot Genomic Regions Harboring Stem Rust Resistant Genes on 1B, 2A (Sr38), and 7B Chromosomes in Wheat. Front. Genet. 12:749675. doi: 10.3389/fgene.2021.749675

Stem rust caused by *Puccinia graminis f.* sp. *tritic*i Eriks. is an important disease of common wheat globally. The production and cultivation of genetically resistant cultivars are one of the most successful and environmentally friendly ways to protect wheat against fungal pathogens. Seedling screening and genome-wide association study (GWAS) were used to determine the genetic diversity of wheat genotypes obtained on stem rust resistance loci. At the seedling stage, the reaction of the common stem rust race QFCSC in Nebraska was measured in a set of 212 genotypes from $F_{3:6}$ lines. The results indicated that 184 genotypes (86.8%) had different degrees of resistance to this common race. While 28 genotypes (13.2%) were susceptible to stem rust. A set of 11,911 single-nucleotide polymorphism (SNP) markers was used to perform GWAS which detected 84 significant marker-trait associations (MTAs) with SNPs located on chromosomes 1B, 2A, 2B, 7B and an unknown chromosome. Promising high linkage disequilibrium (LD) genomic regions were found in all chromosomes except 2B which suggested they include candidate genes controlling stem rust resistance. Highly significant LD was found among these 59 significant SNPs on chromosome 2A and 12 significant SNPs with an unknown chromosomal position. The LD analysis between SNPs located on 2A and Sr38 gene reveal high significant LD genomic regions which was previously reported. To select the most promising stem rust resistant genotypes, a new approach was suggested based on four criteria including, phenotypic selection, number of resistant allele(s), the genetic distance among the selected parents, and number of the different resistant allele(s) in the candidate crosses. As a result, 23 genotypes were considered as the most suitable parents for crossing to produce highly resistant stem rust genotypes against the QFCSC.

Keywords: Triticum aestivum L, stem rust, LD, association mapping, GBS, gene annotation, genomic region

INTRODUCTION

Stem rust caused by *Puccinia graminis f.* sp. *tritici (Pgt)* Erikss and Henning, has been devastating to wheat (*Triticum aestivum* L.) through many decades of production especially during the 1950s in the United States(Leonard 2001; Leonard and Szabo 2005). In recent years, stem rust losses have been minor in the U.S. due to the successful national barberry (Berberis vulgaris L.) eradication program (the alternate host for *P. graminis*), identifying stable sources of resistance, reducing potential new races in the population, and monitoring potential new races of the pathogen through a global network (Kolmer 1996; Hartman et al., 2016). To date, more than 80 stem rust (Sr) genes have been described in tetraploid and hexaploid wheat, and their wild relatives, (online Sr gene catalog, Singh, 2017). In Nebraska and the United States, QFCSC has been reported as the predominant stem rust race (Jin 2005). Although there have been many research efforts to understand the genetic control of this race, further studies are needed to reveal major and minor genes controlling the resistance against this race (Mourad AMI. et al., 2018).

The winter wheat breeding program in the University of Nebraska-Lincoln aims to select and produce wheat cultivars having high yield attributes, winter survival, disease resistance (including stem rust) and end-use quality (Baenziger et al., 2001). In this program, thousands of crosses are made followed by phenotyping and genotyping to select the most promising genotypes for future breeding program. Although phenotyping for stem rust resistance is routinely performed for all populations each year, but the emergence of new races and weather conditions could affect the progress of selection. Therefore, a fruitful selection should be done at the phenotypic and genotypic levels. Identifying new genes controlling stem rust resistance is one of the main targets to release promising resistant winter wheat cultivars. Phenotypic selection for resistance to stem rust at seedling stage is very important as it provides more understanding about the genetic control of stem rust in the evaluated genotypes and allows pyramiding of many resistant genes by crossing the selected genotypes. This type of selection which used in our study is based on a visual score using various scales such as (Stakman et al., 1962). The problem with the visual score is that it is dependent on the precision of human scores (Sallam et al., 2015), which could lead to errors in the evaluation if many individuals involved in this process. Consequently, phenotypic selection for genotypes may be ineffective. Genotyping with known stem rust genes (such as *Sr24, Sr25, Sr26, Sr32, Sr33, Sr36, Sr37, Sr38, Sr39, Sr40, Sr43, Sr44, Sr45, Sr47, Sr50, Sr51, Sr52* and *Sr53*) is useful to confirm the phenotypic selection (Bariana and McIntosh 1993; Dundas et al., 2007; Anugrahwati et al., 2008; Qi et al., 2011; Liu et al., 2011, 2013; Niu et al., 2011, 2014; Klindworth et al., 2012; Periyannan et al., 2013; Mago et al., 2013; Mcintosh et al., 2013; Yu et al., 2017). Moreover, genotyping-by-sequencing method became one of the common methods that is involved in crop breeding and improvement because it generates hundred thousand to millions of SNPs that can be used for genome-wide association study (GWAS). The GWAS is used to detect genes associated with target traits such as stem rust resistance (Amira M. I. Mourad A. M. I. et al., 2018; Juliana et al., 2018; Kumar et al., 2020). Identification of SNP markers associated with stem rust by GWAS can lead to converting these SNPs to KASP markers which have advantages over other DNA molecular markers (Sallam et al., 2017; Kaur et al., 2020)in marker-assisted selection. Using phenotypic selection combined with GWAS results, genotyping with known stem rust genes and genetic diversity will help to accelerate breeding program by selecting the true-promising genotypes as it was suggested in wheat, rice (*Oryza sativa* L.) and barley (*Hordeum vulgare* L.) by (Abou-Zeid and Mourad 2021; Ghazy et al., 2021).

In a recent study, the plant materials (hereafter referred to as DUP 2017) represent part of the preliminary yield trial of Nebraska winter wheat breeding program that is used for releasing new wheat cultivars. The DUP2017 genotypes were evaluated for grain yield in nine U.S.A. locations and high yielding genotypes were found (Eltaher et al., 2021). Moreover, the genetic diversity and population structure were extensively studied in this population and three subpopulations were detected (Eltaher et al., 2018). In addition to the previous studies carried on these DUP2017 genotypes, breeding efforts included a precise selection for stem rust resistance in the recent study. This selection was done based on phenotyping combined with extensive and detailed genetic analyses (GWAS, genotyping with expected stem rust genes, linkage disequilibrium, genetic diversity, and population structure). The DUP2017 genotypes were derived from different crosses among parents which some possessed well known stem rust resistance genes, Sr*38* and *Sr24* genes.

The objectives of this study were to: 1) screen a nursery of 212 Nebraska winter wheat genotypes for their resistance to stem rust race QFCSC, the common race in the United States, 2) identify SNP markers associated with stem rust resistance using GWAS, 3) screen the presence of *Sr38* and *Sr24* genes in the population, 4) select the most promising stem rust resistance genotypes to be used in future breeding programs.

MATERIALS AND METHODS

Plant Materials

A collection of 212 randomly selected genotypes from 270 $F_{3:6}$ lines (Nebraska Duplicate Nursery, syn. DUP 2017) were selected for this study. As mentioned previously, DUP2017 is the preliminary yield trial and the lines are developed from of 800–1,000 crosses among elite Nebraska adapted and cultivars from Great Plains states (Eltaher et al., 2018). The pedigree of all 212 genotypes is presented in **Supplementary Table S1**.

Stem Rust Experiment at the Seedling Stage

The reaction to stem rust race QFCSC was evaluated at the seedling stage using 212 $F_{3:6}$ lines. In addition, four check cultivars; Morocco and "Cheyenne" as susceptible and "Jagger and Arapahoe" as stem rust resistance checks were included. Stem rust spores (race QFCSC) were previously collected from naturally infected field-grown wheat, then increased in the

greenhouse using a highly susceptible cultivar (McNair 701). Two hundred and twelve genotypes and four check cultivars were evaluated in a randomized complete block design with four replicates: two at the plant pathology greenhouses, University of Nebraska Lincoln, UNL and the other two at the USDA-ARS at Kansas State University (KSU).

Inoculation and Incubation

The evaluation was done at the seedling stage by inoculating three leaves of all the genotypes using a pressurized atomizer to uniformly spray an aqueous suspension of freshly harvested urediniospores of race QFCSC (1 mg ml^{-1} in 3 ml Soltrol 170 mineral oil) (Sigma-Aldrich Corp.) according to (Rowell and Olien 1957). Inoculated plants were kept in a dark moist chamber at 21°C with 100% relative humidity for 16 h after inoculation and then moved to a growth chamber set at 20°C/18°C and a 16/8-h light/dark cycle. The stem rust symptoms were scored on the 14th day after inoculation when the rust pustules fully erupted on the inoculated leaves (Jin et al., 2007).

Infection Types Scored

Rust ratings of three young leaves were averaged to reflect an overall rust score of the infected plants. Genotype reaction to stem rust was determined based on infection types (ITs) using a 0–4 scale (Stakman et al., 1962; Roelfs and Martens 1987). Categorical Stakman infection types on the 0–4 scale (Stakman et al., 1962) were converted to a linearized 0–9 scale removing "+," "–," and "; " notations used in the Stakman scale. The 0–4 Stakman scale corresponds to distinct categories of infection types as follows: "0" = no visible uredinia or hypersensitive flecking, "; " = hypersensitive flecking, "1" = small, round uredinia with necrosis or chlorosis, "2" = small-to medium-sized uredinia with green islands surrounded by chlorosis, "3" = medium-sized uredinia with or without chlorosis, "4" = large uredinia without chlorosis. For plants with heterogeneous infection types, all infection types were recorded. For each infection type, '+' or '– 'was used to indicate size variation compared to typical infection types. Stakman ITs "0," "; ," "1–", "1," "1+," "2–," "2," "2+," "3–," "3," "3+," and "4" were converted to linear values 0, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 9, respectively. Genotypes with ITs from 0 to three were rated as resistant, 4 to 5 as moderately resistant, and 6–9 as susceptible as described in (Kumssa et al., 2015)

Genotyping-By-Sequencing and SNP Calling

Genomic DNA was extracted from the wheat leaves of two to three young two-week-old seedlings using BioSprint 96 DNA Plant Kits (Qiagen Valencia, California, United States) as described in (Eltaher et al., 2018). The extracted DNA were sent to USDA-ARS lab, Manhattan, KS, for genotyping-by-sequencing (GBS), simple sequence-repeat (SSR), or sequence-tagged site (STS) markers that link to known rust resistance genes. Some stem rust resistance genes were predicted in some genotypes based on their pedigrees, such as *Sr24, Sr38, Sr31*, and $Sr1RS^{Amigo}$, so SSR and STS markers for these genes were screened for those genotypes (Mourad et al., 2019). Also, genotyping-by-sequencing (GBS) was done as Poland *et al.* (2012), described previously. The SNPs is called with default parameters using the TASSEL v5.2.40 GBS analytics pipeline (Bradbury et al., 2007). The GBS-tags were aligned to the reference genome using Burrows-Wheeler Aligner (Li et al., 2009). The reference genome v1.0 of the "Chinese Spring" genome assembly from the International Wheat Genome Sequencing Consortium (IWGSC) was used in SNP calling. The raw sequence data of the DUP2017 genotypes of the current study along with 6,791 other genotypes previously genotyped in our program were combined for SNP calling to increase the coverage of the genome and read depth at SNP sites (Zhang et al., 2015; Hussain et al., 2017; Belamkar et al., 2018). SNPs were removed from the dataset if they were either monomorphic, showed more than 20% missing values, had conflicting calls from SNPs, or exhibited minor allele frequencies (MAF) of less than 5% (Zhang et al., 2015; Hussain et al., 2017).

Statistical Analysis

The Analysis of Variance and Genetic Variations

The analysis of variance for stem rust resistance was performed with PLABSTAT software (Utz 1997). The data was analyzed using the following equation.

$$Y_{ij} = \mu + g_i + r_j + ge_{ij}$$

where Yij is observation of genotype i in replicate j; μ is the general mean; g_i, e_j are the main effects of genotypes and replications, respectively; ge_{ij} is genotype × replications interaction of genotype i with replicate j. The random effects were assigned to genotypes whereas the fixed effects were assigned to replications. Broad-sense heritability was calculated as

$$H^2 = V_g/V_P$$

where V_g is the genotypic variation and V_p is phenotypic variation.

Population Structure, Kinship Matrix Estimation

The population structure for the $F_{3:6}$ Nebraska winter wheat was performed using the criteria described in (Eltaher et al., 2018). The analysis was done by STRUCTURE 3.4.0 (Pritchard et al., 2000) and the kinship matrix (K) was estimated using TASSEL v5.2.40 (Bradbury et al., 2007).

Single Marker Analysis (SMA) and Genome-wide Association Study

Single marker analysis was performed using converted phenotypic data (0–9 scale) and genotypic data (SMA) that link to known stem rust resistance gene especially *Sr38* using STS marker VENTRIUP-LN2 and *Sr24* using STS marker Sr24#12. The SMA analysis was done using PowerMarker software V. 3.25 (Liu and Muse 2005).

GWAS for stem rust resistance was performed using 11,991 SNPs markers after filtration of minor allele frequencies (MAF<0.05) were removed and excluding all heterozygous loci which were calculated as missing values. The phenotypic means for both traits and SNPs were subjected to association analysis using a mixed linear model (MLM) in TASSEL v5.2.40 software (Bradbury et al., 2007).

Each SNP marker was then fitted into the regression equation to generate a *p*-value. Marker–trait associations were considered significant at false discovery rate (FDR) at 5% significance level. For each regression model, the SNP markers were ranked from smallest to largest *p*-values. A conservative, close approach to previous studies (Pasam et al., 2012; Gao et al., 2016; Visioni et al., 2018; Kumar et al., 2020) was considered to minimize the risk of neglecting any significantly related marker annotating the resistance of the stem rust. The phenotypic variance explained by significant makers (R^2) was determined using TASSEL v5.2.40. Manhattan plots for stem rust were visualized using FarmCPU package (Liu et al., 2016). Linkage disequilibrium (r^2) was estimated using TASSEL 5.0 between each pair of SNPs located on the same chromosome. The LD heatmap was visualized using 'LDheatmap' R package (Shin et al., 2006)

Candidate Genes Linked With Stem Rust

Important markers detected in the SMA were subjected to in silico annotation. The flanking sequence of these markers was obtained from the 1 kb upstream and downstream of the SNP position2 *EnsemblPlants* database. The flanking sequence was used to make a query against IWGSC RefSeq v1.0 and v1.1 to obtain the reference physical map positions of these markers (Appels et al., 2018). Each significant SNP was selected according to its falling inside the gene models. Functional annotation of the genes harboring significant SNPs was retrieved from the genome annotations provided by IWGSC and examined for their association with stem disease resistance.

RESULTS

Genetic Variation for stem Rust Resistance Trait

The analysis of variance revealed highly significant ($p < 0.01$) means squares for genotypes (G), indicating the presence of considerable differences among genotypes for resistance to stem rust (data not shown) as would be expected in a breeding program that selects for stem rust resistance (**Figure 1**). In the DUP 2017, the broad-sense heritability (H^2) for stem rust scores was 0.78. Out of the 212 genotypes, 184 (86.8%) had different degrees of resistance to this common race with a range extending from very high resistance scored as 0 (9 genotypes) to moderate resistance scored as 5 (23 genotypes). Only 28 genotypes (scored as 6 to 9, 13.2%) were susceptible to stem rust with a score ranging from susceptible with a score ranging from 5–6 (six genotypes) to highly susceptible (9) scored as 8 (two genotypes). All genotypes with a score of 0 or one were considered highly resistant to stem rust. A set of 65 highly resistant genotypes (IT < 1) were selected for further genetic investigation. The stem rust resistance of each genotype is presented in **Supplementary Table S1**.

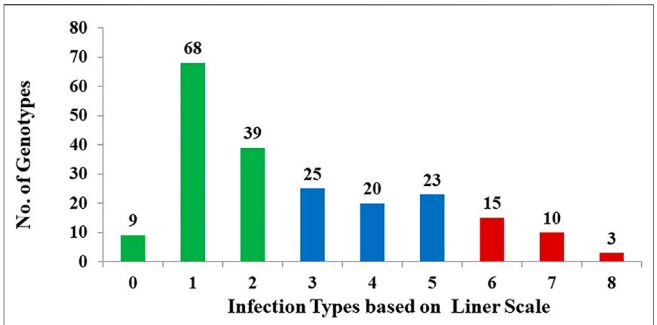


FIGURE 1 | Frequency distribution of stem rust severity scores For 212 randomly selected F 3:6 genotypes to infection with the common stem rust race in Nebraska, QFCSC using the linear scale. Green bars refer to the highly resistant genotypes, blue bars refer to moderate resistance, and red bars refer to susceptible genotypes.

Genetic Variation in Stem Rust Resistance Based on Population Structure

Population structure (PS) analysis of DUP2017 was previously analyzed by (Eltaher et al., 2018). The results of the PS divided the genotypes into three Subgroups. The genetic variation in stem rust resistance to race QFCSC was studied in each subpopulation and presented in **Figure 2**. Sub-population (SP2) had the highest number of tested genotypes (130) followed by subpopulation one (SP1; 55 tested genotypes) then subpopulation three (SP3; 27 tested genotypes). In SP1, 76.4% (42 genotypes) were resistant. While 46.9% (61 genotypes) of the SP2 demonstrated resistance to stem rust. Finally, 48.1% (13 genotypes) of the SP3 displayed high degree of resistance. The mean of IT scores for genotypes in SP1 was 1.72 IT which was higher than that of those in SP2 (2.73 IT) and SP3 (2.25 IT) respectively. Single factor analysis was performed to test if there were significant differences among the three groups for stem rust resistance. The results revealed no significant differences among the three groups for the respective trait (data not shown).

Genome-wide Association Study for Stem Rust Resistance

The mixed linear model (MLM) was used to test the genetic association between the 11,911 SNPs and the stem rust-resistance scores of all tested genotypes. The results of GWAS revealed 84 SNPs located on four chromosomes: 1B (4 SNPs), 2A (59 SNPs), 2B (1 SNP), 7B (8 SNPs) and 12 SNPs that were not mapped to a known chromosome (Unknown chromosome). The summary of GWAS results is presented in **Table 1** (detailed results are presented in **Supplementary Table S2**). The distribution of the 84 significant SNPs across the chromosomes is illustrated *via* Manhattan plot in **Figure 3A**, Quantile–quantile plots of

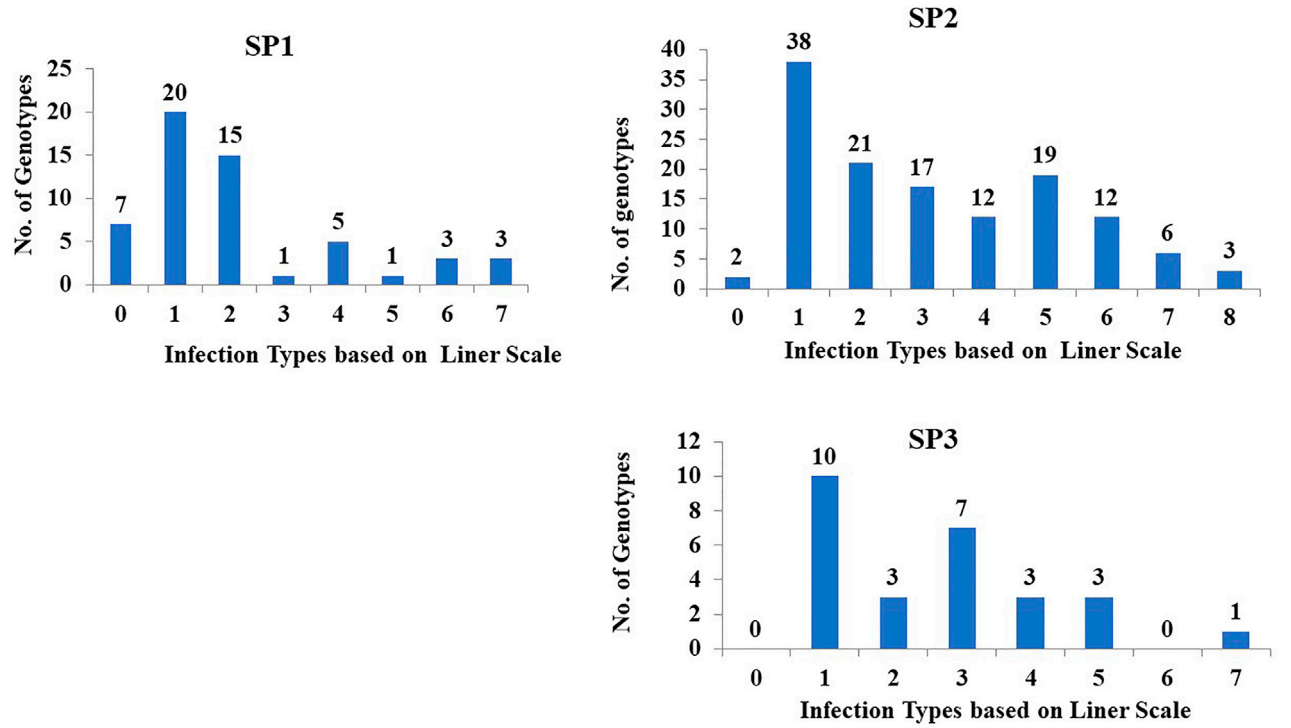


FIGURE 2 | The distribution of 212 tested genotypes with the most common stem rust race in Nebraska QFCSC over the three sub-populations. SP refer to subpopulation based on population structure analysis.

TABLE 1 | Summary of GWAS analysis for stem rust resistance including number of SNPs, range of *p* value, range of R^2, and range of allele effect.

Chromosome	Total no. of SNPs	−10log *p* value		PhenotypicVariation (R^2)		Allele effect range	
		Min	Max	Min	Max	Min	Max
1B	4	5.5642E-06	0.00038525	6.04	12.75	−2.15 (A and C)	−1.42 (T)
2A	59	3.086E-09	2.357E-06	13.66	25.61	−2.84 (G)	−2.09 (A)
2B	1	2.357E-06		2.357E-06		2.357E-06	
7B	8	3.175E-05	0.0003562	11.41	12.74	−1.90 (C)	−1.79 (T)
UN	12	8.5299E-09	0.000065646	9.55	22.74	−2.76 (T)	−1.95 (T)

p-values comparing the uniform distribution of the expected–log10 *p*-value to the observed–log10 *p*-value for stem rust resistance trait showed that the MLM fitted the data well (**Figure 3B**).

The phenotypic variation explained by each SNP marker (R^2) ranged from 6.04% (S1B_561712520) to 25.61% (in both markers S2A_13028312 and S2A_13028321). The effects of alleles associated with decreased susceptibility to stem rust ranged from -2.83 (G) in both markers, S2A_13028312 and S2A_13028321, to -1.42 (T), S1B_547524267. Notably, the highest allele effects which decreased the stem rust symptoms were accounted for alleles found on 2A chromosome and those that belonged to the unknown chromosomal position.

The linkage disequilibrium (r^2) was estimated between each pair of SNPs located on the same chromosome (**Supplementary Figure S1**). If a group of SNPs was in significant LD, this group was named an LD genomic region (GR). A highly significant LD was found between the two SNPs located on 1B (S1B_561712520 and S1B_561712544) (GR1). Complete significant LD was found among all the SNPs located on 2A (GR2) with r^2 of 1. Also, the eight SNPs located on the chromosome 7B were found in a highly significant LD (GR 3). All the 12 SNPs markers located on unknown chromosomal position had significant LD indicating that they are all linked (GR 4). For unknown positional SNP markers, LD was tested between this group and SNPs located on each of the known chromosome GR to determine the most likely chromosome which the SNPs on an unknown chromosome

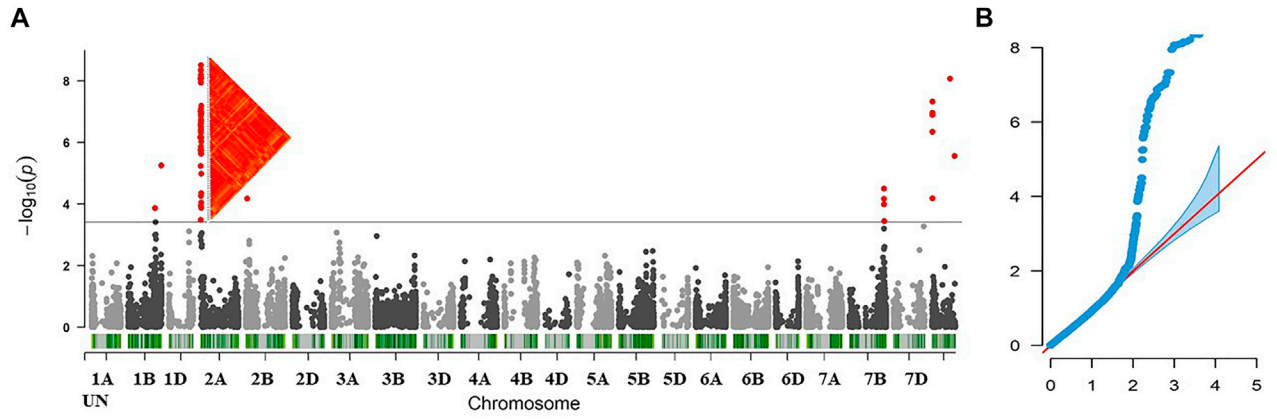


FIGURE 3 | (A) Manhattan plot displaying SNP marker-trait association identified for stem rust resistance in GWAS using 212 wheat lines **(B)** Quantile-Quantile (QQ) plot used to evaluate the performance of the mixed linear model used for of GWAS for stem rust resistance using mixed linear model (MLM).

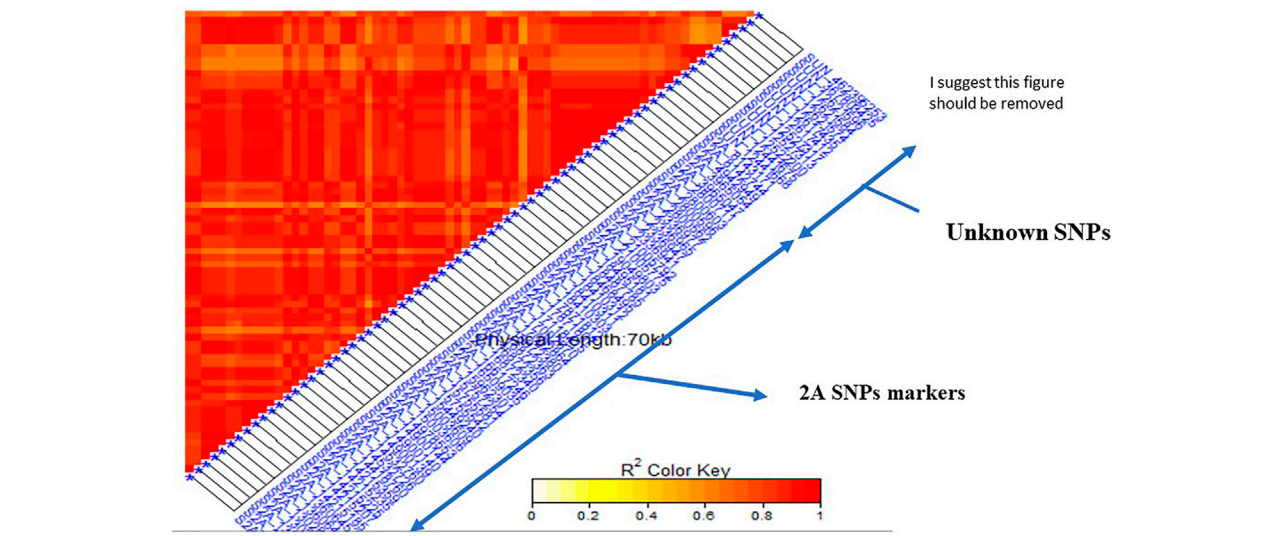


FIGURE 4 | Linkage disequilibrium (LD) analysis in the tested genotypes: heatmap of LD between the 59 SNPs located on 2A and 12 Unknown SNPs markers showed highly significant LD.

belong to. The 12 unknown SNPs were in a high LD with the 59 SNPs located on 2A chromosome (**Figure 4**). According to this finding, the genomic region of the 12 SNP markers was combined with the GR2 and three major GR were identified in this study.

Our next step was to identify candidate genes for resistance to stem rust and to determine if the GR include known major genes for stem rust. We inspected the putative function of gene sequences corresponding to the SNPs associated with the resistant phenotype. The gene annotation analysis of the 84 SNPs markers revealed a large number of candidate genes in each genomic region (**Supplementary Table S2**). Many of these genes were found to be disease-related genes, particularly the hotspot (GR2) which located on the chromosome 2A. The genomic region located on chromosome 7B (GR 3) was found to be had four candidate genes (**Figure 5**). These gene models encoded cytochrome P450 enzymes, cytochrome P450, E-class, group I, and P-loop containing nucleoside triphosphate hydrolase.

Single Marker Analysis for *Sr38* and *Sr24* Genes

Based on the pedigrees of the tested genotypes, some stem rust resistance genes were strongly expected such as *Sr38* (2A) and *Sr24* (3D). Based on the STS marker of the two genes (Sr24#12 and VENTRIUP-LN2), a percentage of 55.7% of the genotypes contained *Sr38* marker in their genome (118 genotypes) while only 52 genotypes with a percentage of 24.5% contained *Sr24*

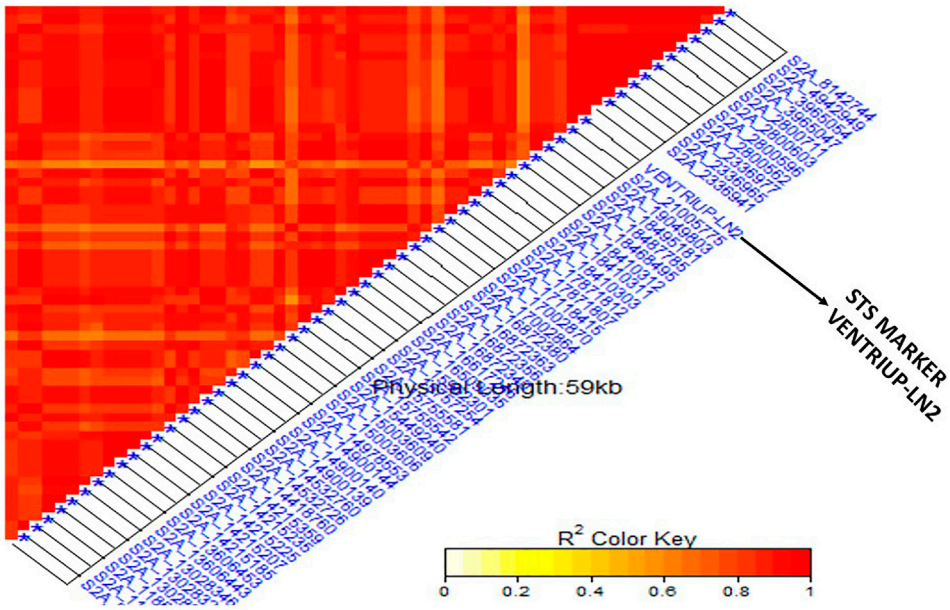


FIGURE 5 | Linkage disequilibrium (LD) analysis in the tested genotypes: heatmap of LD between the specific STS marker "VENTRIUP-LN2" for Sr38 resistance gene and the 59 SNPs on 2A showed highly significant LD with the marker.

TABLE 2 | Single marker analysis for stem rust resistance in the tested genotypes explained by marker VENTRIUP-LN2 for Sr38 gene and STS marker Sr24#12 for Sr24.

	VENTRIUP-LN2	Sr24#12
Gene ID	*Sr38*	*Sr24*
p-value	2.45E-33	0.133472484
F-Statical	226.6559	2.269179534
F-Critical	3.895458	3.88612144
Phenotypic Variation	56.57%	1.06%
MS between group	331.3538	8.784369054
MS within group	1.461924	3.871165292
Group 1(with gene)	118[b] (with IT average 1.15)	52 (with IT average 2.07)
Group 2(Non-gene)	58 (with IT average (4.07)	160 (with IT average 2.54)
Allele effects[a]	−1.46	−0.235

[a]*The effect of the presence band (gene).*
[b]*There are 36 genotypes were heterozygous bands for this gene and excluded from single-marker analysis.*

resistance gene (**Supplementary Table S1**). The results of SMA between the phenotypic data and marker data of the two genes are presented in **Table 2**. The SMA analysis between *Sr24* gene marker and the phenotypic data found non-significant differences between the two groups. On the other hand, a highly significant association between stem rust resistance gene *Sr38* in tested genotypes with a *p*-value of 2.45203E-33 and R^2 of 57%. The average of stem rust resistant for the group possessing the *Sr38* (118 genotypes) was 1.15 while it was 4.07 for the group not possessing the *Sr38* gene (58 genotypes). Thirty-six genotypes were heterozygous/heterogenous for the presence of *Sr38* gene. It was observed that some resistant genotypes (low stem rust score) did not have the markers associated with *Sr38* gene indicating the presence of other gene(s). The linkage disequilibrium was performed again between the STS marker VENTRIUP-LN2 (*Sr38* gene marker) and all SNPs located in GR 3 (71 SNPs) (**Figure 5**). The results of LD revealed a high significant LD (r^2 ~ 0.90) between the SNP markers and STS marker.

Selection for the Most Promising Stem Rust Resistance Genotypes for the Upcoming Breeding Program

Three criteria were considered to determine the most promising stem rust-resistant genotypes as candidate parents for a future cross to increase stem rust resistance. These criteria were based upon the following:

First, phenotypic selection in which all highly resistant genotypes with stem rust score of 0 or one were selected. As a result, 65 genotypes were identified and included in the second stage.

Second, the presence of resistant alleles in each genotype was determined (**Supplementary Table S3**). The number of resistant alleles (84 marker alleles) were tested in each selected genotype to identify the genotypes which possessed the highest number of resistant alleles (**Supplementary Table S3**). The number of resistant alleles in the selected genotypes ranged from 11 in (NE17670) to 84 in 22 genotypes. It was noted all highly resistant genotypes possessed the GR3 and *Sr38* gene except NE17670. Therefore, it was necessary to examine the genetic distance between the NE17670 and the 22 genotypes. Consequently, the 23 genotypes were included in the third stage Third, population structure and the genetic distance matrix (GD) among the selected genotypes.

The genetic distance among the 23 genotypes is illustrated *via* dendrogram cluster analysis in **Figure 7**. All the 23 genotypes were from the three subpopulations (SP) according to our population structure analysis described in Eltaher et al. (2018) with 13, eight, and two for SP1, SP2, and SP3, respectively. The genetic distance extended from 0.130 (NE17627and NE17624) to 0.619 (NE17535 and NE17571). Remarkably, all the 65 genotypes identified by the first criteria had *Sr38* gene except NE17670.

DISCUSSION

Genetic Variation in Stem Rust Resistance

Crop scientists face a serious challenge of increasing productivity by controlling stresses caused by biotic and abiotic effects. Stem rust of wheat, among biotic factors, poses a continuous threat through the rapid evolution of new races. Resistant cultivars are developed and considered to be the most economical and environmentally friendly tool for disease control. The primary gene pool, including indigenous collections comprising landraces, old cultivars, and breeding lines, is considered a valuable genetic resource to provide new and sustainable resistance that can be utilized for the production of today's high yielding cultivars (Mujeeb-Kazi et al., 2013; Kumar et al., 2020; Bhavani et al., 2021). A better understanding of the genetic structure of genetic resistance is the first step towards improving and enhancing the disease resistance of this important crop. Several field and controlled greenhouse studies reported that stem rust resistance is likely under oligenic or polygenic additive regulation, due to the combined effect of multiple loci (major and minor) beneficial alleles with variable effect (Laidò et al., 2015; Saccomanno et al., 2018). Most importantly, identification of promising stem rust-resistant genotypes is the key point of the successful breeding program to truly produce high yielding cultivars with excellent resistance to stem rust.

Significant variation was observed among a collection of 212 selected genotypes from $F_{3:6}$ lines (DUP 2017) as indicated by the analysis of variance (ANOVA). The high broad-sense heritability observed in this study indicated the reliability of data for GWAS and that selection for stem rust-resistant genotypes would be successful. The phenotypic distribution of disease response was not normally distributed which was reported also in previous studies on stem rust resistance (Gao et al., 2016; Edae et al., 2018; Saccomanno et al., 2018; Alqudaha et al., 2019; Kumar et al., 2020). In this study, more than 87% of DUP2017 genotypes were resistant to the most common stem rust race in Nebraska (QFCSC). The result was expected as selection based on resistance to stem rust is one of the main objectives for Nebraska wheat breeding program (El-basyoni et al., 2013). The genotypes studied were derived from many crosses which were known to have parents with excellent stem rust resistance. All segregating generations after crossing were subject to selection for stem rust resistance (though many plants in the field escape the disease), grain yield, and agronomic performance. Although selection for stem rust resistance was attempted in each generation, about 13% of genotypes were susceptible to QFCSC at the seedling stage. Therefore, phenotypic selection at the seedling stage alone can be misleading due to plant development and genotype × environment interaction. Phenotypic selection along with molecular genetic tools will lead to genetic improvement and a better understanding of stem rust resistance in wheat (Mourad et al., 2019; Dawood et al., 2020; Moursi et al., 2020; Ghazy et al., 2021). Moreover, selection at the seedling stage for stem rust resistance is very important as it is an efficient assay for advancing lines to the next generation.

The analysis of genetic diversity and population structure were extensively described in this population by (Eltaher et al., 2018). The genotypes were divided into three subpopulations (SP) (**Supplementary Table S1**). There were no significant differences in stem rust resistance among the three subpopulations (data not shown). This result indicated that the three subpopulations though genetically different were similar in their selection history and stem rust resistance. Resistant genotypes (n = 65) from the three subpopulations can be selected for genetic diversity and stem rust resistance for future wheat breeding (Sallam et al., 2016; Mourad et al., 2019). The three subpopulations had highly stem rust-resistant genotypes (SR scores of <1) with 25, 31 and nine genotypes from subpopulations 1, 2, and 3, respectively.

Genome-Association Study for Stem Rust Resistance

GWAS is effective for identifying novel genes associated with stem rust resistance (Elasyoni et al., 2017; Kankwatsa et al., 2017; Edae et al., 2018; Mourad et al., 2019). In this study, we identified 84 significant MTAs distributed on different chromosomes 1B, 2A, 2B, 7B and UN (later determined to be linked to markers on 2A). A set of 78 SNPs were considered major QTLs R^2 greater than 10%. While six SNPs (S1B_547524267, S1B_561712520, S1B_561712544, S2B_28097761, SUN_12527313 and SUN_12527317) were considered a minor QTL with R^2 less than 10%. Many earlier studies reported that large-effect QTLs controlling target trait have R^2 of >10% (Hussain et al., 2017; Mourad A. M. I. et al., 2018; Alqudaha et al., 2019; Mourad et al., 2019; Niu et al., 2020). Kumar et al.)2020) detected 349 SNPs associated with stem rust resistance at seedling stage with R^2 ranging from 3.04 to 7.47% which was lower range than reported in this study. The analysis of LD between each pair of SNPs located on the same chromosome divided the 84 significant SNPs into three genomic regions. The analysis of LD provides an important information on the markers which tend to be co-inherited together from generation to generation (Sallam et al., 2016). Moreover, the analysis of LD allowed us identifying the chromosomal position of the significant SNPs with unknown chromosomal position. In the current study, we identify and validate genomic regions association with stem rust resistance. The candidate genes within each genomic region were extensively identified and described.

Validation of a Hot Spot Genomic Region Associated With Sr38 Gene

In our study we found a set of 59 significant SNPs located on chromosome 2A and 12 on unknown chromosome were in a

highly significant LD with specific STS marker for *Sr38* stem rust gene in the VENTRIUP-LN2 translocation. The high LD found among the 12 SNPs indicated that those unknown chromosomal positions were part of the translocation that was not well mapped to the reference sequence especially in the region where the 59 SNPs were located. The LD was very useful for identifying the possible chromosomal positions of some of unknown markers. Mourad *et al.* (2019) identified a set of 17 SNPs associated with increased resistance to the same race were located on 2A and linked to *Sr38* in the DUP2015 Nebraska winter wheat. Ten of the 17 previously identified significant SNPs were common between the two studies (**Supplementary Table S2**) and associated with QFCSC stem rust race. The two populations (DUP2015 and DUP 2017) are genetically different and were produced from different crosses, but often had similar parents. Therefore, the 10 significant markers can be considered for marker-assisted selection. As expected, the LD between the SNP markers and *Sr38* gene confirmed the localization of *Sr38* gene in 2A chromosome. The results of the two studies recommended to use the 10 SNP markers as a strong signal for the presence of *Sr38* gene. The 10 SNPs can be converted to KASP (Kompetitive allele specific PCR) markers for further validation studies as KASP markers have more advantages over the other DNA molecular markers. While 10 markers were in common between the two studies, the other seven SNP makers in Mourad et al. (2019) were found in the raw sequence data of our study but they were excluding after marker filtration. Consequently, we can say that in the Nebraska wheat breeding program, the *Sr38* stem rust resistance gene remains a broadly used and effective resistance gene to the QFCSC local strain (Alabushev et al., 2019; Mourad et al., 2019). The other remaining SNPs (42) located on the 2A chromosome in this study were far from those detected by Mourad et al. (2019). Therefore, they could be considered novel SNPs associated with stem rust resistance.

The candidate genes in GR3 were detected and 15 gene models were identified (**Supplementary Table S2**). Due to the presence of many gene models in this region, we expect that this hotspot region may contains many resistance genes in addition *Sr38* gene. By looking on the stem rust genes map https://globalrust.org/knowledge-center/gene, we found that chromosome 2A containing different rust gene such as *Sr21*, *Sr32* and *Sr38/Lr37/Yr17* (Friebe et al., 1996; Nisha et al., 2015; Newcomb et al., 2016; Mourad et al., 2019). However, previous studies concluded that *Sr38*, *Sr21*, and *Sr31* are three different resistant genes which have been transferred to hexaploid wheat from different translocations (The 1973; Roelfs and McVey 1979; Bariana and McIntosh 1993; Friebe et al., 1996). Due to the presence of one GR in our results, we can conclude that *Sr21* and *Sr32* could not be the other genes expected in this genomic region and this region may carry new or unknown resistant genes. The functional annotation of the identified 15 gene models was discussed in the following paragraphs.

New and *VP* Putative Genomic Regions Associated With Resistance to QFCSC Stem Rust Race

Four SNPs were found to be associated with stem rust resistance on 1B chromosome. The four SNPs were found in a complete LD. S1B_547524267, which has T and C alleles, marker was found to be within *TraesCS1B02G322500* gene model which encodes pectin lyase fold/virulence factor. It was reported that pectin is among the plant cell wall components and considered an essential target for different pathogens at the early stages of infection (Wu et al., 2019). Pectin lyases are virulence factors that target the pectic components of the plant cell wall to degrade them. Therefore, this gene model is associated with increased the susceptibility to QFCSC stem rust. The allele T of S1B_547524267 marker decreased the symptoms of stem rust, while the allele C increased the stem rust symptoms. Therefore, the allele C indicate the presence of this gene. A set of QTLs at the 1B chromosome were identified using mapping populations and GWAS panel (Pozniak et al., 2008; Bhavani et al., 2011; Njau et al., 2013; Bajgain et al., 2015). Kumar et al. (2020), reported important SNP markers associated with stem rust resistance in wheat. They found 91 SNPs located on 1B chromosome. The four SNPs found in this study were far from the positions of the 92 SNPs.

For 2A chromosome 71 SNPs were found associated with stem rust resistant (59 SNPs on 2A and 12 SNPs on unknown chromosome) and around 15 gene models were recognized (**Supplementary Table S2**). The gene model *TraesCS2A02G003700.1* (2,336,941–2,336,977 bp) encoded to receptor like kinases (RLKs). RLKs have been discovered to play a role in both broad-spectrum, elicitor-initiated defense responses and race-specific pathogen defense as dominant resistance (R) genes. The majority of defense-related RLKs are of the leucine-rich repeat (LRR) subclass (Kruijt et al., 2004). *TraesCS2A02G010200.2* (3965047–3965054 bp) which encoded to steroidogenic acute regulatory based transfer (StART)-like domain superfamily. In insect, humans, and plants, StART proteins play a variety of roles in the transport of lipid molecules (Tang et al., 2005). These proteins consist of a modular StART domain of approximately 200 amino acids which binds and transfers the lipids. The StART domain is found in many signaling proteins and is believed to have important roles in lipid transport, lipid metabolism and cellular signaling (Soccio and Breslow 2003; Tang et al., 2005). The StART proteins plays role in the plant defense against powdery mildew in Arabidopsis by EDR2, a PH (Pleckstrin homology) and START (lipid/sterol-binding StAR-related lipid transfer) domain-containing protein (Vogel et al., 2002, 2004; Tang et al., 2005). The gene model *TraesCS2A02G028800* (13,028,312–13,028,346 bp) which encoded to F- Box domain proteins family. As one of the largest and most diverse plant gene families, F-box proteins are involved in many cellular processes, including cell cycle, circadian rhythms, embryogenesis, floral organ development, stress responses, and various signal transduction pathways. F-box proteins are reported to be related with the plant response against bacterial, viral and fungal pathogens (van den Burg et al., 2008; Piisilä et al., 2015; Li et al., 2020). Also, many of F-box proteins in higher plants have been characterized by genetic analysis and are involved in various abiotic stresses (Calderón-Villalobos et al., 2007; Xu et al., 2008; Zhang et al., 2008; Bu et al., 2014). The gene model *TraesCS2A02G028800* (13,028,312–13,028,346 bp) which

FIGURE 6 | The schematic representation of genomic regions map on chromosome 7B.

encoded the ABC transporter domain superfamily. Plant ABC transporters are classified into several sub-families (ABCA - ABCH) and play diverse roles (Campbell et al., 2003). Although approximately 131 ABC transporters have been identified in Arabidopsis, via sequence similarity to known ABC transporters in other organisms, very little is known about the functions or the substrate specificities of most of these genes (Campbell et al., 2003; Jasinski et al., 2003). ABC transporters have been associated with various host-pathogen interactions. In plant pathogenic fungi, members of this transporter group play a role in providing resistance to phytoalexins (Nakaune et al., 1998; Urban et al., 1999; Schoonbeek et al., 2001; Fleissner et al., 2002; Campbell et al., 2003; Jasinski et al., 2003), and to antifungal compounds (Hayashi et al., 2002), or act as novel pathogenicity factors (Fleissner et al., 2002; Campbell et al., 2003). In addition, several gene models on chromosome 2A have been discovered to have relationships with plant protection. For example, *TraesCS2A02G042800.1*, *TraesCS2A02G040600.1* and *TraesCS2A02G036900.1* which encoded to Chloramphenicol acetyltransferase-like domain superfamily and Isopenicillin N synthase-like these types of proteins had antibacterial effects. The functional annotation of these gene models confirmed the presence of many stem rust resistance genes.

On 2B chromosome, one SNP (S2B_28,097,761) was found to be associated with stem rust resistance. This SNPs located very near *TraesCS2B02G057600* (28,093,392–28,097,727) which encodes to MFS transporter superfamily. The MFS transporter is a member of plant defense-related proteins that could be involved in exporting the antimicrobial compounds produced by plant pathogens, the plant-generated antimicrobial compounds; and potassium which is important during plant defense reactions (Friebe et al., 1996; Simmons et al., 2003). Antimicrobial compounds were found to provide resistance against fungal in different plants in different ways such as avoidance, enzymatic degradation, and non-degradative mechanisms(Osbourn 1999; Seybold et al., 2020). According to McIntosh atlas (McIntosh et al., 1995), many stem rust resistance genes were mapped on this chromosome such as *Sr10*, *Sr16*, *Sr9*, *Sr12*, *Sr19*, *Sr20*, *Sr23*, *Sr28*, *Sr32*, *Sr36*, *Sr39*, *Sr40*, *Sr47*, and *SrWeb*. However, more studies are definitely needed to provide more information about the significant SNP marker and resistance genes on chromosome 2B. In the study of Kumar et al. (2020), five SNPs were associated with stem rust at seedling stage with positions different from what was found in our study.

A GWAS analysis showed the presence of highly significant SNPs located on 7B chromosome. This set of eight SNPs located on 7B was highly LD and considered as a genomic region (**Figure 6**). Previous studies on bi-parental mapping populations and GWAS panel suggested the presence of *Sr17* on 7B chromosome (Bansal et al., 2008; Yu et al., 2011, 2017; Leonova et al., 2020). However, *Sr17* is a virulent gene against QFCSC race. Therefore, the significant SNPs could not be associated with *Sr17* gene. All the four SNPs were considered major QTLs which tend to co-inherited together. The gene annotation of the candidate significant SNPs supported the results of the marker-trait association. Chromosome 7B seemed to have a very interesting genomic region including four gene models (**Figure 6**). Three candidate genes *TraesCS7B02G439700.1*, TraesCS7B02G439800, and *TraesCS7B02G439400.1* (704,826,623–704,827,404 pb) encoded cytochrome P450 Enzymes which play an important role in enhancing the resistance of several plant-fungal interactions including stem rust, powdery mildew (*Blumeria graminis* f. sp. *tritici*) and *Fusarium* head blight (*Fusarium graminearum)*

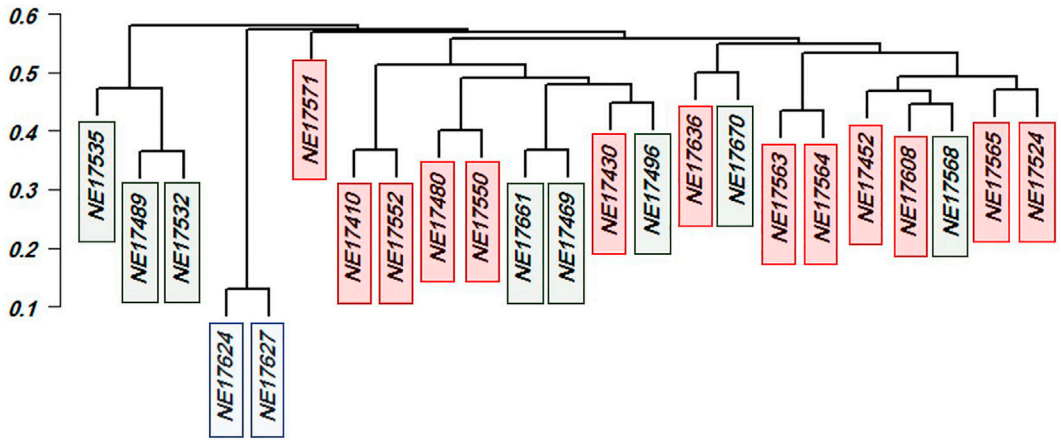


FIGURE 7 | Dendrogram cluster and principal component analysis for the selecting 23 selected genotypes with the most forms of resistance infections types. Red color refer to SPI, Green color refer to SP2 and Blue color refer to SP3.

(Becher and Wirsel 2012). S7B_704827432 was located within *TraesCS7B02G439900* gene model which encoded the protein P-loop containing nucleoside triphosphate hydrolase. This protein is encoded by many wheat disease resistance genes that are distributed across the wheat genome. The domains of this protein were found to be associated with the receptors that can detect pathogenic effectors. Interestingly, a set of 132 genes encoding P-loop containing nucleoside triphosphate hydrolase protein was found in chromosome 7B. By comparing the position of genomic region of this unidentified gene (704,826,623–705,256,765, **Supplementary Table S2**) with the same position of genomic region reported by Becher and Wirsel, (2012), we found that the two genes *TraesCS7B02G437400.1* (703719566) and *TraesCS7B02G441700.1* (706811897) are near the gene detected in this study. These results indicated that 7B chromosome may include very important genomic regions associated with QFCSC resistance which can be used for marker-assisted selection after validating the SNPs in a different genetic background. A set of four SNPs were associated with stem rust and located on 7B detected by Kumar et al. (2020). In this study, All SNPs located in a genomic region starting from 704,826,623 to 705,256,765 (GR3). In the study of Kumar et al. (2020), the four SNPs were located on four different positions 619883523, 638510367, 716966341 and 730906118.

Therefore, by comparing the position of SNPs and candidate genes detected in the study with the positions of SNPs and candidate genes reported by earlier studies, it can be concluded the most genomic regions detected in this study were novel. Bearing in mind that, all detected candidate genes found in this study have a strong relation with plant disease resistance in wheat.

Utilize the Aspects of GWAS and Genetic Diversity Analysis to Identify the Most Stem Rust-Resistant Genotypes

We also expect that many stem rust resistance genes are present in our materials due to, wide range of infection types. For instance, we have 58 genotypes with non *Sr38* gene (**Table 2**) the average infection types ranged from (0.42–7.50). Also, the genotype (NE17670) which marked as resistance in the selected 23 genotypes with IT (0.42) and did not have SNP markers indicative of Sr38 gene, this indicated the presence of other resistance genes in our DUP2017 genotypes.

Phenotypic selection is widely used in traditional plant breeding. However, phenotypic selection could be misleading due to epistasis and the environmental or human errors which could reduce heritability and lead to ineffective selection. Selection based on genotypic and phenotypic values together can address this challenge by truly select the most promising stem rust-resistant genotypes.

To address this challenge, genotypes were selected based on three criteria as described by (Eltaher et al., 2021). Firstly, phenotypic selection for the highest resistant stem rust genotypes. Out of the 212 genotypes, 65 highly stem rust-resistant genotypes were selected to be advanced to the next phase of selection. Secondly, the number of resistant alleles and their genomic regions, detected by GWAS, were counted in each genotype. It was very useful to identify the number of resistant alleles which each selected genotype carried as it shed the light on the number of putative genes controlling the resistance of stem rust in this population. Most of the GWAS literature overlooks the number of resistant alleles in the target genotypes. Here, the number of resistant alleles confirmed the results of phenotypic selection. For example, resistant genotypes should contain at least some of the resistant alleles detected by GWAS and the number of resistant alleles should be more than susceptible alleles. It was noted that all the selected genotypes had the SNPS for GR3 except one genotype, NE17670 which included only 11 resistant allele with an IT of 0.41. By looking to the data, we selected 22 genotypes which had the maximum number of resistant allele (84) and it was interesting to include the NE17670 genotypes as it was thought to possess other resistant genes which were not detected by GWAS. Thirdly, genetic distance and population structure among the remaining 23 genotypes was considered. Genetic distance was very useful in providing information on how each two genotypes

are genetically dissimilar, hence maintain genetic diversity in the breeding program. It also provided a way to screen parents for the number of different GR. We discovered that number of GR in each genotype was not enough to select a parent. For example, NE17624 and NE17627 had 84 resistant alleles (3 GRs) against stem rust. The genetic distance based on resistant alleles was 0.13 which indicated that both genotypes are highly genetically similar. On the other hand, it was noted that NE17670 had lowest number of resistant alleles. This genotype possessed (11) resistant alleles. The genetic distance between this genotype and NE17469 was 0.62 with lines being different for 73 stem-rust related alleles. Bearing in mind that the analysis of population structure assigned NE17670 and NE17469 in SP2 (**Figure 7**). Therefore, NE17670 as a candidate parent should be included in the future crosses to produce cultivars having more resistance to stem rust race QFCSC. The fact that extremely genetically distant genotypes are best in the crossbreeding phase was previously described in (Eltaher et al., 2018). Consequently, hybridization between NE17670, which belongs to SP2, and any of the two genotypes (NE17624 and NE17627), which belong to SP3, should be considered, especially for pyramiding stem rust resistance genes. Hybridization of NE17670 for all 13 SP1 genotypes can also be helpful. Therefore, integration of NE17670 as a main parent in the crosses with the other genotypes in SP1, SP3 genotypes will be fruitful in producing cultivars having more resistance to the common stem rust race QFCSC on one hand, to maintain the genetic diversity among the lines on the other hand.

CONCLUSION

Four LD genomic regions controlling important *Sr* genes associated with stem rust resistance were identified. In particular, the genomic region harboring Sr38, one of the most important resistant genes to stem rust, was found. The validated SNPs in this region can be converted to KASP markers which can be used for marker-assisted selection for Sr38. Moreover, important new SNPs especially those located on 7B chromosome were identified. These markers will need to be validated before using them in MAS, but the first step is completed. The gene annotation analysis revealed putative genes associated with fungal disease resistance. These results further support that GWAS was a powerful method to identify target alleles. Moreover, most of significant SNP detected by GWAS were considered with major effects in stem rust resistance except six SNPs with minor effects.

Finally, information from genetic diversity, population structure, and GWAS results were combined to identify the most promising wheat resistant genotypes as potential parents for future a breeding program. As a result, 23 genotypes were identified, and based on our results, we recommend NE17670 be used as a parent in future crosses as it may have other resistant genes which were not identified by GWAS.

KEY MESSAGE

Important genomic regions associated with resistance to QFCSC stem rust race were identified and validated. Moreover, the most promising stem rust genotypes were identified for future breeding and genetics programs.

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**. The GBS data analyzed during the current study are available in the NBCI repository, http://www.ncbi.nlm.nih.gov/bioproject/680548

AUTHOR CONTRIBUTIONS

Conceptualization: AS, PB. Data Curation: SE, SW, Formal Analysis: SE, AM, AS, VB. Methodology: SE, AM. Supervision: AS, PB. Visualization: SE, AS. Writing–Review and Editing: SE, AM, SW, PB, AS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.749675/full#supplementary-material

REFERENCES

Abou-Zeid, M. A., and Mourad, A. M. I. (2021). Genomic Regions Associated with Stripe Rust Resistance against the Egyptian Race Revealed by Genome-wide Association Study. *BMC Plant Biol.* 21, 42–14. doi:10.1186/s12870-020-02813-6

Alabushev, A. V., Vozhzhova, N. N., Kupreyshvili, N. T., Shishkin, N. V., Marchenko, D. M., and Ionova, E. V. (2019). Identification of Stem Rust Resistance Genes in the winter Wheat Collection from Southern Russia. *Plants* 8, 559. doi:10.3390/plants8120559

Alqudaha, A. M., Sallamb, A., Baenzigerc, P. S., and Börnera, A. (2019). GWAS: Fast-Forwarding Gene Identification in Temperate Cereals: Barley as a Case Study-A Review *J. Adv. Res.* 22, 119–135. doi:10.1016/j.jare.2019.10.013

Anugrahwati, D. R., Shepherd, K. W., Verlin, D. C., Zhang, P., Mirzaghaderi, G., Walker, E., et al. (2008). Isolation of Wheat-rye 1RS Recombinants that Break the Linkage between the Stem Rust Resistance Gene SrR and Secalin. *Genome* 51, 341–349. doi:10.1139/g08-019

Appels, R., Eversole, K., Appels, R., Eversole, K., Feuillet, C., Keller, B., et al. (2018). Shifting the Limits in Wheat Research and Breeding Using a Fully Annotated Reference Genome. *Science* 361, 361. doi:10.1126/science.aar7191

Baenziger, P. S., Shelton, D. R., Shipman, M. J., and Graybosch, R. A. (2001). Breeding for End-Use Quality: Reflections on the Nebraska Experience. *Euphytica* 119, 255–262. doi:10.1023/A:1017583514424

Bajgain, P., Rouse, M., Bulli, P., Bhavani, S., Gordon, T., Wanyera, R., et al. (2015). Association Mapping of North American spring Wheat Breeding Germplasm Reveals Loci Conferring Resistance to Ug99 and Other African Stem Rust Races. *BMC Plant Biol.* 15, 249. doi:10.1186/s12870-015-0628-9

Bansal, U. K., Bossolini, E., Miah, H., Keller, B., Park, R. F., and Bariana, H. S. (2008). Genetic Mapping of Seedling and Adult Plant Stem Rust Resistance in Two European winter Wheat Cultivars. *Euphytica* 164, 821–828. doi:10.1007/s10681-008-9736-z

Bariana, H. S., and McIntosh, R. A. (1993). Cytogenetic Studies in Wheat. XV. Location of Rust Resistance Genes in VPM1 and Their Genetic Linkage with Other Disease Resistance Genes in Chromosome 2A. *Genome* 36, 476–482. doi:10.1139/g93-065

Becher, R., and Wirsel, S. G. R. (2012). Fungal Cytochrome P450 Sterol 14α-Demethylase (CYP51) and Azole Resistance in Plant and Human Pathogens. *Appl. Microbiol. Biotechnol.* 95, 825–840. doi:10.1007/s00253-012-4195-9

Bhavani, S., Singh, R. P., Argillier, O., et al. (2011). "Mapping Durable Adult Plant Stem Rust Resistance to the Race Ug99 Group in Six CIMMYT Wheats," in Borlaug Global Rust Initiative 2011 Technical Workshop June 13-16 Saint Paul, Minnesota, USA, 42.

Bhavani, S., Singh, P. K., Qureshi, N., He, X., Biswal, A. K., Qureshi, N., Juliana, P., et al. (2021). Globally Important Wheat Diseases: Status, Challenges, Breeding and Genomic Tools to Enhance Resistance Durability. *Genomic Des. Biot. Stress Resist. Cereal Crop.*, 59–128. doi:10.1007/978-3-030-75879-0_2

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

Bu, Q., Lv, T., Shen, H., Luong, P., Wang, J., Wang, Z., et al. (2014). Regulation of Drought Tolerance by the F-Box Protein MAX2 in Arabidopsis. *Plant Physiol.* 164, 424–439. doi:10.1104/pp.113.226837

Calderón-Villalobos, L. I. A., Nill, C., Marrocco, K., Kretsch, T., and Schwechheimer, C. (2007). The Evolutionarily Conserved *Arabidopsis thaliana* F-Box Protein AtFBP7 Is Required for Efficient Translation during Temperature Stress. *Gene* 392, 106–116. doi:10.1016/j.gene.2006.11.016

Campbell, E. J., Schenk, P. M., Kazan, K., Penninckx, I. A. M. A., Anderson, J. P., Maclean, D. J., et al. (2003). Pathogen-responsive Expression of a Putative ATP-Binding Cassette Transporter Gene Conferring Resistance to the Diterpenoid Sclareol Is Regulated by Multiple Defense Signaling Pathways in Arabidopsis. *Plant Physiol.* 133, 1272–1284. doi:10.1104/pp.103.024182

Dawood, M. F. A., Moursi, Y. S., Amro, A., Baenziger, P. S., and Sallam, A. (2020). Investigation of Heat-Induced Changes in the Grain Yield and Grains Metabolites, with Molecular Insights on the Candidate Genes in Barley. *Agronomy* 10, 1730. doi:10.3390/agronomy10111730

Dundas, I. S., Anugrahwati, D. R., Verlin, D. C., Park, R. F., Bariana, H. S., Mago, R., et al. (2007). New Sources of Rust Resistance from Alien Species: Meliorating Linked Defects and Discovery. *Aust. J. Agric. Res.* 58, 545–549. doi:10.1071/ar07056

Edae, E. A., Pumphrey, M. O., and Rouse, M. N. (2018). A Genome-wide Association Study of Field and Seedling Response to Individual Stem Rust Pathogen Races Reveals Combinations of Race-specific Genes in North American spring Wheat. *Front. Plant Sci.* 9, 52. doi:10.3389/fpls.2018.00052

El-basyoni, I., Baenziger, P. S., Dweikat, I., Wang, D., Eskridge, K., and Saadalla, M. (2013). Using DArT Markers to Monitor Genetic Diversity throughout Selection: a Case Study in Nebraska's winter Wheat Breeding Nurseries. *Crop Sci.* 53, 2363–2373. doi:10.2135/cropsci2013.01.0051

Elasyoni, I. S., El-Orabey, W. M., Baenziger, P. S., and Eskridge, K. M. (2017). Association Mapping for Leaf and Stem Rust Resistance Using Worldwide spring Wheat Collection. *Asian J. Biol.* 4, 1–25. doi:10.9734/AJOB/2017/38120

Eltaher, S., Baenziger, P. S., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., et al. (2021). GWAS Revealed Effect of Genotype × Environment Interactions for Grain Yield of Nebraska winter Wheat. *BMC Genomics* 22, 2–14. doi:10.1186/s12864-020-07308-0

Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., et al. (2018). Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. *Front. Genet.* 9, 76. doi:10.3389/fgene.2018.00076

Fleissner, A., Sopalla, C., and Weltring, K-M. (2002). An ATP-Binding Cassette Multidrug-Resistance Transporter Is Necessary for Tolerance of Gibberella Pulicaris to Phytoalexins and Virulence on Potato Tubers. *Mol. Plant-microbe Interact* 15, 102–108. doi:10.1094/MPMI.2002.15.2.102

Friebe, B., Jiang, J., Raupp, W. J., McIntosh, R. A., and Gill, B. S. (1996). Characterization of Wheat-Alien Translocations Conferring Resistance to Diseases and Pests: Current Status. *Euphytica* 91, 59–87. doi:10.1007/bf00035277

Gao, L., Turner, M. K., Chao, S., Kolmer, J., and Anderson, J. A. (2016). Genome Wide Association Study of Seedling and Adult Plant Leaf Rust Resistance in Elite spring Wheat Breeding Lines. *PLoS One* 11, e0148671–25. doi:10.1371/journal.pone.0148671

Ghazy, M. I., Salem, K. F. M., and Sallam, A. (2021). Utilization of Genetic Diversity and Marker-Trait to Improve Drought Tolerance in rice (Oryza Sativa L.). *Mol. Biol. Rep.* 48, 157–170. doi:10.1007/s11033-020-06029-7

Hartman, G. L., Pawlowski, M. L., Chang, H.-X., and Hill, C. B. (2016). "Successful Technologies and Approaches Used to Develop and Manage Resistance against Crop Diseases and Pests," in *Emerging Technologies for Promoting Food Security*. (Amsterdam: Elsevier), 43–66. doi:10.1016/b978-1-78242-335-5.00003-2

Hayashi, K., Schoonbeek, H.-j., and De Waard, M. A. (2002). Bcmfs1 , a Novel Major Facilitator Superfamily Transporter from Botrytis Cinerea , Provides Tolerance towards the Natural Toxic Compounds Camptothecin and Cercosporin and towards Fungicides. *Appl. Environ. Microbiol.* 68, 4996–5004. doi:10.1128/aem.68.10.4996-5004.2002

Hussain, W., Baenziger, P. S., Belamkar, V., Guttieri, M. J., Venegas, J. P., Easterly, A., et al. (2017). Genotyping-by-Sequencing Derived High-Density Linkage Map and its Application to QTL Mapping of Flag Leaf Traits in Bread Wheat. *Sci. Rep.* 7, 16394. doi:10.1038/s41598-017-16006-z

Jasinski, M., Ducos, E., Martinoia, E., and Boutry, M. (2003). The ATP-Binding Cassette Transporters: Structure, Function, and Gene Family Comparison between rice and Arabidopsis. *Plant Physiol.* 131, 1169–1177. doi:10.1104/pp.102.014720

Jin, Y. (2005). Races of Puccinia Graminis Identified in the United States during 2003. *Plant Dis.* 89, 1125–1127. doi:10.1094/pd-89-1125

Jin, Y., Singh, R. P., Ward, R. W., Wanyera, R., Kinyua, M., Njau, P., et al. (2007). Characterization of Seedling Infection Types and Adult Plant Infection Responses of Monogenic Sr Gene Lines to Race TTKS of Puccinia Graminis F. Sp. Tritici. *Plant Dis.* 91, 1096–1099. doi:10.1094/pdis-91-9-1096

Juliana, P., Singh, R. P., Singh, P. K., Poland, J. A., Bergstrom, G. C., Huerta-Espino, J., et al. (2018). Genome-wide Association Mapping for Resistance to Leaf Rust, Stripe Rust and Tan Spot in Wheat Reveals Potential Candidate Genes. *Theor. Appl. Genet.* 131, 1405–1422. doi:10.1007/s00122-018-3086-6

Kankwatsa, P., Singh, D., Thomson, P. C., Babiker, E. M., Bonman, J. M., Newcomb, M., et al. (2017). Characterization and Genome-wide Association Mapping of Resistance to Leaf Rust, Stem Rust and Stripe Rust in a Geographically Diverse Collection of spring Wheat Landraces. *Mol. Breed.* 37, 113. doi:10.1007/s11032-017-0707-8

Kaur, B., Mavi, G. S., Gill, M. S., and Saini, D. K. (2020). Utilization of KASP Technology for Wheat Improvement. *Cereal Res. Commun.* 48, 409–421. doi:10.1007/s42976-020-00057-6

Klindworth, D. L., Niu, Z., Chao, S., Friesen, T. L., Jin, Y., Faris, J. D., et al. (2012). Introgression and Characterization of a Goatgrass Gene for a High Level of Resistance to Ug99 Stem Rust in Tetraploid Wheat. *G3 Genes, Genomes, Genet.* 2, 665–673. doi:10.1534/g3.112.002386

Kolmer, J. A. (1996). Genetics of Resistance to Wheat Leaf Rust. *Annu. Rev. Phytopathol.* 34, 435–455. doi:10.1146/annurev.phyto.34.1.435

Kruijt, M. (2004). Molecular Evolution of Cladosporium Fulvum Disease Resistance Genes in Wild Tomato. *Wageningen University*. Available at: http://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=https%3A%2F%2Fedepot.wur.nl%2F121591

Kumar, D., Kumar, A., Chhokar, V., Gangwar, O. P., Bhardwaj, S. C., Sivasamy, M., et al. (2020). Genome-Wide Association Studies in Diverse Spring Wheat Panel for Stripe, Stem, and Leaf Rust Resistance. *Front. Plant Sci.* 11, 748. doi:10.3389/fpls.2020.00748

Kumssa, T. T., Baenziger, P. S., Rouse, M. N., Guttieri, M., Dweikat, I., Brown-Guedira, G., et al. (2015). Characterization of Stem Rust Resistance in Wheat Cultivar Gage. *Crop Sci.* 55, 229–239. doi:10.2135/cropsci2014.05.0348

Laidò, G., Panio, G., Marone, D., Russo, M. A., Ficco, D. B. M., Giovanniello, V., et al. (2015). Identification of New Resistance Loci to African Stem Rust Race TTKSK in Tetraploid Wheats Based on Linkage and Genome-wide Association Mapping. *Front. Plant Sci.* 6, 1033. doi:10.3389/fpls.2015.01033

Leonard, K. J. (2001). "Stem Rust-Future Enemy?," in *Stem Rust Wheat, from Anc Enemy to Mod Foe*, Editor P. D. Peterson (St. Paul: APS Press), 119–146.

Leonard, K. J., and Szabo, L. J. (2005). Stem Rust of Small Grains and Grasses Caused byPuccinia Graminis. *Mol. Plant Pathol.* 6, 99–111. doi:10.1111/j.1364-3703.2005.00273.x

Leonova, I. N., Skolotneva, E. S., Orlova, E. A., Orlovskaya, O. A., and Salina, E. A. (2020). Detection of Genomic Regions Associated with Resistance to Stem Rust in Russian spring Wheat Varieties and Breeding Germplasm. *Int. J. Mol. Sci.* 21, 4706–4713. doi:10.3390/ijms21134706

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/map Format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Li, H., Wei, C., Meng, Y., Fan, R., Zhao, W., Wang, X., et al. (2020). Identification and Expression Analysis of Some Wheat F-Box Subfamilies during Plant Development and Infection by Puccinia Triticina. *Plant Physiol. Biochem.* 155, 535–548. doi:10.1016/j.plaphy.2020.06.040

Liu, K., and Muse, S. V. (2005). PowerMarker: an Integrated Analysis Environment for Genetic Marker Analysis. *Bioinformatics* 21, 2128–2129. doi:10.1093/bioinformatics/bti282

Liu, W., Danilova, T. V., Rouse, M. N., Bowden, R. L., Friebe, B., Gill, B. S., et al. (2013). Development and Characterization of a Compensating Wheat-Thinopyrum Intermedium Robertsonian Translocation with Sr44 Resistance to Stem Rust (Ug99). *Theor. Appl. Genet.* 126, 1167–1177. doi:10.1007/s00122-013-2044-6

Liu, W., Jin, Y., Rouse, M., Friebe, B., Gill, B., and Pumphrey, M. O. (2011). Development and Characterization of Wheat-Ae. Searsii Robertsonian Translocations and a Recombinant Chromosome Conferring Resistance to Stem Rust. *Theor. Appl. Genet.* 122, 1537–1545. doi:10.1007/s00122-011-1553-4

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-wide Association Studies. *Plos Genet.* 12, e1005767. doi:10.1371/journal.pgen.1005767

Mago, R., Verlin, D., Zhang, P., Bansal, U., Bariana, H., Jin, Y., et al. (2013). Development of Wheat-Aegilops Speltoides Recombinants and Simple PCR-Based Markers for Sr32 and a New Stem Rust Resistance Gene on the 2S#1 Chromosome. *Theor. Appl. Genet.* 126, 2943–2955. doi:10.1007/s00122-013-2184-8

Mcintosh, R. A., Dubcovsky, J., Rogers, W. J., Morris, C., Appels, R., and Xia, X. C. (2013) Catalogue of Gene Symbols for Wheat: 2013-2014 Supplement. 12th Int Wheat Genet Symp 1–31.

McIntosh, R. A., Wellings, C. R., and Park, R. F. (1995). *Wheat Rusts: An Atlas of Resistance Genes*. Clayton, Australia: Csiro Publishing.

Mourad, A. M. I., Sallam, A., Belamkar, V., Mahdy, E., Bakheit, B., Abo El-Wafaa, A., et al. (2018b). Genetic Architecture of Common Bunt Resistance in winter Wheat Using Genome-wide Association Study. *BMC Plant Biol.* 18, 280. doi:10.1186/s12870-018-1435-x

Mourad, A. M. I., Sallam, A., Belamkar, V., Wegulo, S., Bai, G., Mahdy, E., et al. (2019). Molecular Marker Dissection of Stem Rust Resistance in Nebraska Bread Wheat Germplasm. *Sci. Rep.* 9, 11694. doi:10.1038/s41598-019-47986-9

Mourad, A. M. I., Sallam, A., Belamkar, V., Wegulo, S., Bowden, R., Jin, Y., et al. (2018a). Genome-Wide Association Study for Identification and Validation of Novel SNP Markers for Sr6 Stem Rust Resistance Gene in Bread Wheat. *Front. Plant Sci.* 9, 1–12. doi:10.3389/fpls.2018.00380

Moursi, Y. S., Thabet, S. G., Amro, A., Dawood, M. F. A., Baenziger, P. S., and Sallam, A. (2020). Detailed Genetic Analysis for Identifying QTLs Associated with Drought Tolerance at Seed Germination and Seedling Stages in Barley. *Plants* 9, 1425. doi:10.3390/plants9111425

Mujeeb-Kazi, A., Kazi, A. G., Dundas, I., Rasheed, A., Ogbonnaya, F., Kishii, M., et al. (2013). Genetic Diversity for Wheat Improvement as a Conduit to Food Security. *Adv. Agron.* 122, 179–257. doi:10.1016/b978-0-12-417187-9.00004-8

Nakaune, R., Adachi, K., Nawata, O., Tomiyama, M., Akutsu, K., and Hibi, T. (1998). A Novel ATP-Binding Cassette Transporter Involved in Multidrug Resistance in the Phytopathogenic Fungus Penicillium digitatum. *Appl. Environ. Microbiol.* 64, 3983–3988. doi:10.1128/aem.64.10.3983-3988.1998

Newcomb, M., Olivera, P. D., Rouse, M. N., Szabo, L. J., Johnson, J., Gale, S., et al. (2016). Kenyan Isolates of Puccinia Graminis F. Sp. Tritici from 2008 to 2014: Virulence to SrTmp in the Ug99 Race Group and Implications for Breeding Programs. *Phytopathology* 106, 729–736. doi:10.1094/phyto-12-15-0337-r

Nisha, R., Sivasamy, M., Gajalakshmi, K., Vikas, V. K., Jayaprakash, P., Sajitha, P., et al. (2015). Pyramiding of Stem Rust Resistance Genes Sr2, Sr36/Pm6 and Sr24/Lr24 to Develop Durable and Multiple Disease Resistant Wheat Varieties through Marker Aided Selection. *Int. J. Ext Res.* 5, 1–9.

Niu, X., Zhu, Y., Sun, Z., Yu, S., Zhuang, J., and Fan, Y. (2020). Identification and Validation of Quantitative Trait Loci for Grain Number in Rice (Oryza Sativa L.). *Agronomy* 10, 180. doi:10.3390/agronomy10020180

Niu, Z., Klindworth, D. L., Friesen, T. L., Chao, S., Jin, Y., Cai, X., et al. (2011). Targeted Introgression of a Wheat Stem Rust Resistance Gene by DNA Marker-Assisted Chromosome Engineering. *Genetics* 187, 1011–1021. doi:10.1534/genetics.110.123588

Niu, Z., Klindworth, D. L., Yu, G., L Friesen, T., Chao, S., Jin, Y., et al. (2014). Development and Characterization of Wheat Lines Carrying Stem Rust Resistance Gene Sr43 Derived from Thinopyrum Ponticum. *Theor. Appl. Genet.* 127, 969–980. doi:10.1007/s00122-014-2272-4

Njau, P. N., Bhavani, S., Huerta-Espino, J., Keller, B., and Singh, R. P. (2013). Identification of QTL Associated with Durable Adult Plant Resistance to Stem Rust Race Ug99 in Wheat Cultivar 'Pavon 76'. *Euphytica* 190, 33–44. doi:10.1007/s10681-012-0763-4

Osbourn, A. E. (1999). Antimicrobial Phytoprotectants and Fungal Pathogens: a Commentary. *Fungal Genet. Biol.* 26, 163–168. doi:10.1006/fgbi.1999.1133

Pasam, R. K., Sharma, R., Malosetti, M., van Eeuwijk, F. A., Haseneyer, G., Kilian, B., et al. (2012). Genome-wide Association Studies for Agronomical Traits in a World Wide spring Barley Collection. *BMC Plant Biol.* 12, 16. doi:10.1186/1471-2229-12-16

Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., et al. (2013). The Gene Sr33, an Ortholog of Barley Mla Genes, Encodes Resistance to Wheat Stem Rust Race Ug99. *Science* 341 (6147), 786–788. doi:10.1126/science.1239028

Piisilä, M., Keceli, M. A., Brader, G., Jakobson, L., Jõesaar, I., Sipari, N., et al. (2015). The F-Box Protein MAX2 Contributes to Resistance to Bacterial Phytopathogens in *Arabidopsis thaliana*. *BMC Plant Biol.* 15, 53–17. doi:10.1186/s12870-015-0434-4

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J.-L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-By-Sequencing Approach. *PLoS One* 7, e32253. doi:10.1371/journal.pone.0032253

Pozniak, C. J., Reimer, S., Fetch, T., Clarke, J. M., Clarke, F. R., Somers, D., et al. (2008). *Association Mapping of UG99 Resistance in a Diverse Durum Wheat Population*. Sydney: Sydney University Press.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945–959. doi:10.1093/genetics/155.2.945

Qi, L. L., Pumphrey, M. O., Friebe, B., Zhang, P., Qian, C., Bowden, R. L., et al. (2011). A Novel Robertsonian Translocation Event Leads to Transfer of a Stem Rust Resistance Gene (Sr52) Effective against Race Ug99 from Dasypyrum Villosum into Bread Wheat. *Theor. Appl. Genet.* 123, 159–167. doi:10.1007/s00122-011-1574-z

Roelfs, A. P., and Martens, J. W. (1987). An International System of Nomenclature for Puccinia Graminis F. Sp. Tritici. *Phytopathology* 78, 526–553.

Roelfs, A. P., and McVey, D. V. (1979). Low Infection Types Produced byPuccinia Graminisf. sp.Triticiand Wheat Lines with Designated Genes for Resistance. *Phytopathology* 69, 722–730. doi:10.1094/phyto-69-722

Rowell, J. B., and Olien, C. R. (1957). Controlled Inoculation of Wheat Seedlings with Uredospores of Puccinia-Graminis Var. Tritici. *Phytopathology* 47, 650–655.

Saccomanno, A., Matny, O., Marone, D., Laidò, G., Petruzzino, G., Mazzucotelli, E., et al. (2018). Genetic Mapping of Loci for Resistance to Stem Rust in a Tetraploid Wheat Collection. *Int. J. Mol. Sci.* 19, 3907–3915. doi:10.3390/ijms19123907

Sallam, A., Sidiqi, J., and Baenziger, S. (2017) Screening Winter Wheat Lines in Nebraska for the Fhb1 Gene Using Kompetitive Allele Specific PCR (KASP). *J. Plant Genet. Breed.* 1:e104

Sallam, A., Arbaoui, M., El-Esawi, M., Abshire, N., and Martsch, R. (2016). Identification and Verification of QTL Associated with Frost Tolerance Using Linkage Mapping and GWAS in winter Faba Bean. *Front. Plant Sci.* 7, 1098. doi:10.3389/fpls.2016.01098

Sallam, A., Martsch, R., and Moursi, Y. S. (2015). Genetic Variation in Morpho-Physiological Traits Associated with Frost Tolerance in Faba Bean (Vicia faba L.). *Euphytica* 205, 395–408. doi:10.1007/s10681-015-1395-2

Schoonbeek, H., Del Sorbo, G., and De Waard, M. A. (2001). The ABC Transporter BcatrB Affects the Sensitivity of Botrytis Cinerea to the Phytoalexin Resveratrol

and the Fungicide Fenpiclonil. *Mol. Plant Microbe Interact* 14, 562–571. doi:10.1094/mpmi.2001.14.4.562

Seybold, H., Demetrowitsch, T. J., Hassani, M. A., Szymczak, S., Reim, E., Haueisen, J., et al. (2020). A Fungal Pathogen Induces Systemic Susceptibility and Systemic Shifts in Wheat Metabolome and Microbiome Composition. *Nat. Commun.* 11, 1910–1912. doi:10.1038/s41467-020-15633-x

Shin, J-H., Blay, S., McNeney, B., and Graham, J. (2006). LDheatmap: an R Function for Graphical Display of Pairwise Linkage Disequilibria between Single Nucleotide Polymorphisms. *J. Stat. Softw.* 16, 1–10. doi:10.18637/jss.v016.c03

Simmons, C. R., Fridlender, M., Navarro, P. A., and Yalpani, N. (2003). A maize Defense-Inducible Gene Is a Major Facilitator Superfamily Member Related to Bacterial Multidrug Resistance Efflux Antiporters. *Plant Mol. Biol.* 52, 433–446. doi:10.1023/a:1023982704901

Singh, D. P. (2017). *Management of Wheat and Barley Diseases*. Boca Raton: CRC Press.

Soccio, R. E., and Breslow, J. L. (2003). StAR-related Lipid Transfer (START) Proteins: Mediators of Intracellular Lipid Metabolism. *J. Biol. Chem.* 278, 22183–22186. doi:10.1074/jbc.r300003200

Stakman, E. C., Stewart, D. M., and Loegering, W. Q. (1962). *Identification of Physiologic Races of Puccinia Graminis Var. Tritici*. Washington: USDA Washington.

Tang, D., Ade, J., Frye, C. A., and Innes, R. W. (2005). Regulation of Plant Defense Responses in Arabidopsis by EDR2, a PH and START Domain-Containing Protein. *Plant J.* 44, 245–257. doi:10.1111/j.1365-313x.2005.02523.x

The, T. T. (1973). Chromosome Location of Genes Conditioning Stem Rust Resistance Transferred from Diploid to Hexaploid Wheat. *Nat. New Biol.* 241, 256. doi:10.1038/newbio241256a0

Urban, M., Bhargava, T., and Hamer, J. E. (1999). An ATP-Driven Efflux Pump Is a Novel Pathogenicity Factor in rice Blast Disease. *EMBO J.* 18, 512–521. doi:10.1093/emboj/18.3.512

Utz, H. F. (1997). A Computer Program for Statistical Analysis of Plant Breeding Experiments. Version 2N. Stuttgart: University of Hohenheim.

van den Burg, H. A., Tsitsigiannis, D. I., Rowland, O., Lo, J., Rallapalli, G., MacLean, D., et al. (2008). The F-Box Protein ACRE189/ACIF1 Regulates Cell Death and Defense Responses Activated during Pathogen Recognition in Tobacco and Tomato. *Plant Cell* 20, 697–719. doi:10.1105/tpc.107.056978

Visioni, A., Gyawali, S., Selvakumar, R., Gangwar, O. P., Shekhawat, P. S., Bhardwaj, S. C., et al. (2018). Genome Wide Association Mapping of Seedling and Adult Plant Resistance to Barley Stripe Rust (Puccinia Striiformis F. Sp. Hordei) in India. *Front. Plant Sci.* 9, 520. doi:10.3389/fpls.2018.00520

Vogel, J. P., Raab, T. K., Schiff, C., and Somerville, S. C. (2002). PMR6, a Pectate Lyase-like Gene Required for Powdery Mildew Susceptibility in Arabidopsis. *Plant Cell* 14, 2095–2106. doi:10.1105/tpc.003509

Vogel, J. P., Raab, T. K., Somerville, C. R., and Somerville, S. C. (2004). Mutations in PMR5 Result in Powdery Mildew Resistance and Altered Cell wall Composition. *Plant J.* 40, 968–978. doi:10.1111/j.1365-313x.2004.02264.x

Wu, W., Nemri, A., Blackman, L. M., Catanzariti, A.-M., Sperschneider, J., Lawrence, G. J., et al. (2019). Flax Rust Infection Transcriptomics Reveals a Transcriptional Profile that May Be Indicative for Rust Avr Genes. *PLoS One* 14, e0226106. doi:10.1371/journal.pone.0226106

Xu, W., Li, Z., Deng, X. W., Wu, W., and Xue, Y. (2008). F-box Protein DOR Functions as a Novel Inhibitory Factor for ABA-Induced Stomatal Closure under Drought Stress in *Arabidopsis thaliana*. *Plant Physiol.* 148, 2121. doi:10.1104/pp.108.126912

Yu, L.-X., Chao, S., Singh, R. P., and Sorrells, M. E. (2017). Identification and Validation of Single Nucleotide Polymorphic Markers Linked to Ug99 Stem Rust Resistance in spring Wheat. *PLoS One* 12, e0171963. doi:10.1371/journal.pone.0171963

Yu, L.-X., Lorenz, A., Rutkoski, J., Singh, R. P., Bhavani, S., Huerta-Espino, J., et al. (2011). Association Mapping and Gene-Gene Interaction for Stem Rust Resistance in CIMMYT spring Wheat Germplasm. *Theor. Appl. Genet.* 123, 1257–1268. doi:10.1007/s00122-011-1664-y

Zhang, J., Xu, Y., Chen, W., Dell, B., Vergauwen, R., Biddulph, B., et al. (2015). A Wheat 1- FEH W3 Variant Underlies Enzyme Activity for Stem WSC Remobilization to Grain under Drought. *New Phytol.* 205, 293–305. doi:10.1111/nph.13030

Zhang, Y. e., Xu, W., Li, Z., Deng, X. W., Wu, W., and Xue, Y. (2008). F-box Protein DOR Functions as a Novel Inhibitory Factor for Abscisic Acid-Induced Stomatal Closure under Drought Stress in Arabidopsis. *Plant Physiol.* 148, 2121–2133. doi:10.1104/pp.108.126912

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 26 October 2021
doi: 10.3389/fgene.2021.773597

Nucleotide Diversity of the Maize *ZmCNR13* Gene and Association With Ear Traits

Zhihao Zuo[1,2,3†], Yue Lu[1,2,3†], Minyan Zhu[1,2,3], Rujia Chen[1,2,3], Enying Zhang[4], Derong Hao[5], Qianfeng Huang[1,2], Hanyao Wang[1], Yanze Su[1], Zhichao Wang[1], Yang Xu[1,2,3], Pengcheng Li[1,2,3], Chenwu Xu[1,2,3] and Zefeng Yang[1,2,3]**

[1]Jiangsu Key Laboratory of Crop Genetics and Physiology Key Laboratory of Plant Functional Genomics of the Ministry of Education Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding, Agricultural College of Yangzhou University, Yangzhou, China, [2]Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou, China, [3]Joint International Research Laboratory of Agriculture and Agri-Product Safety of Ministry of Education of China, Yangzhou University, Yangzhou, China, [4]College of Agronomy, Qingdao Agricultural University, Qingdao, China, [5]Jiangsu Yanjiang Institute of Agricultural Sciences, Nantong, China

OPEN ACCESS

Edited by:
Ahmed Sallam,
Assiut University, Egypt

Reviewed by:
Zhikai Liang,
University of Minnesota Twin Cities, United States
Karansher Singh Sandhu,
Washington State University, United States

***Correspondence:**
Zefeng Yang
zfyang@yzu.edu.cn
Chenwu Xu
cwxu@yzu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *10 September 2021*
Accepted: *14 October 2021*
Published: *26 October 2021*

Citation:
Zuo Z, Lu Y, Zhu M, Chen R, Zhang E, Hao D, Huang Q, Wang H, Su Y, Wang Z, Xu Y, Li P, Xu C and Yang Z (2021) Nucleotide Diversity of the Maize ZmCNR13 Gene and Association With Ear Traits. Front. Genet. 12:773597. doi: 10.3389/fgene.2021.773597

The maize (*Zea mays* L.) *ZmCNR13* gene, encoding a protein of *fw2.2-like* (*FWL*) family, has been demonstrated to be involved in cell division, expansion, and differentiation. In the present study, the genomic sequences of the *ZmCNR13* locus were re-sequenced in 224 inbred lines, 56 landraces and 30 teosintes, and the nucleotide polymorphism and selection signature were estimated. A total of 501 variants, including 415 SNPs and 86 Indels, were detected. Among them, 51 SNPs and 4 Indels were located in the coding regions. Although neutrality tests revealed that this locus had escaped from artificial selection during the process of maize domestication, the population of inbred lines possesses lower nucleotide diversity and decay of linkage disequilibrium. To estimate the association between sequence variants of *ZmCNR13* and maize ear characteristics, a total of ten ear-related traits were obtained from the selected inbred lines. Four variants were found to be significantly associated with six ear-related traits. Among them, SNP2305, a non-synonymous mutation in exon 2, was found to be associated with ear weight, ear grain weight, ear diameter and ear row number, and explained 4.59, 4.61, 4.31, and 8.42% of the phenotypic variations, respectively. These results revealed that natural variations of *ZmCNR13* might be involved in ear development and can be used in genetic improvement of maize ear-related traits.

Keywords: maize, nucleotide polymorphism, ear-related traits, ZmCNR13, association analysis

INTRODUCTION

Maize (*Zea mays* L.), one of the most important cereal crops, is cultivated worldwide as sources of food, animal feed, and industrial materials. It was suggested that the total global maize production was 1,148.4 million tons in 2019, which was far greater than those of rice (*Oryza sativa* L.) and wheat (*Triticum aestivum* L.)[1]. However, improving kernel yield (KY) is still a primary mission in maize breeding (Li et al., 2018). KY is a complex quantitative trait affected

[1]http://www.fao.org/faostat/zh/#data/QCL.2019

TABLE 1 | Descriptive statistics and ANOVA results of the maize ear-related traits.

Trait	Mean	*SD*	Min	Max	*CV*	*F* Value (G)	*F* Value (E)	*F* Value (G×E)	h^2 (%)
EW (g)	77.04	29.10	6.02	214.10	37.77	13.52[a]	178.37[a]	4.15[a]	28.04
EGW (g)	63.09	25.73	3.66	184.00	40.79	13.72[a]	128.18[a]	4.27[a]	28.03
EL (cm)	12.14	2.44	3.77	21.54	20.14	17.30[a]	545.33[a]	3.67[a]	38.08
ED (cm)	3.91	0.50	1.01	8.63	12.75	16.87[a]	113.09[a]	2.79[a]	47.15
ERN	13.58	2.40	6.00	22.00	17.65	21.57[a]	71.86[a]	2.24[a]	52.97
KNR	21.06	5.88	1.00	42.00	27.93	11.65[a]	128.32[a]	3.26[a]	28.56
HKW (g)	25.58	5.95	10.20	44.44	23.25	35.97[a]	2223.19[a]	7.92[a]	58.30
KL (mm)	9.46	1.39	4.93	33.73	14.72	9.78[a]	101.88[a]	2.02[a]	41.50
KW (mm)	8.12	0.98	5.22	25.20	12.11	10.44[a]	37.11[a]	1.45[a]	45.80
KT (mm)	5.08	0.91	2.58	15.23	17.82	8.71[a]	610.32[a]	2.91[a]	22.65

[a]*Indicates statistical significance at $p < 0.001$ level.*

Abbreviationsin the table are as follows: CV, coefficient of variation,; G, genotype; E, environment; G×E, genotype-environment interaction; h^2, broad-sense heritability; EW, ear weight; EGW, ear grain weight; EL ear length; ED, ear diameter; ERN, ear row number; KNR, kernel number per row; HKW, hundred kernel weight; KL, kernel length; KW, kernel width and KT, kernel thickness.

by a variety of genetic and environmental factors and has low heritability (Beavis et al., 1994; Yan et al., 2006). Compared with KY, the heritability of yield components is relatively higher and is less affected by environmental factors (Messmer et al., 2009; Li et al., 2013; Raihan et al., 2016). Therefore, it is more effective to select some yield components than to directly select KY in the breeding process (Sidwell et al., 1976). Among the yield components, kernel size has a crucial effect on kernel weight estimated by kernel length (KL), kernel width (KW), and kernel thickness (KT). In addition, ear length (EL), ear diameter (ED), ear row number (ERN), and kernel number per row (KNR) are essential traits determining the kernel number (Zhang et al., 2017). All of these ear-related traits are quantitative traits regulated by multiple genes and environmental factors (Liu et al., 2012). Many genes regulating maize ear-related traits have been identified, such as *fea2* (Bommert et al., 2013), *fea3* (Je et al., 2016), *KNR1* (Wang et al., 2019) and *KNR4* (Liu et al., 2015) for kernel row number, *td1* (Bommert et al., 2005), *bif2* (McSteen and Hake 2001) and *ba1* (Ritter et al., 2002) for tassel morphology.

Tomato *fruit-weight 2.2* (*fw2.2*) was detected as an essential locus in controlling fruit weight and size (Frary et al., 2000; Nesbitt and Tanksley 2001). In plants, the homologs of *fw2.2-like (FWL)* genes, encoding proteins with a conserved PLAC8 (named after a series of human placental specific protein with unknown function) domain, were suggested to play essential roles in cell division and organ size control (Libault and Stacey 2010). A total of eight members of *FWL* gene family in rice were detected. Among them, *OsFWL1* and *OsFWL3* genes modulate rice grain length by regulating cell number, and *OsFWL4* gene is a negative regulator of tiller number and plant yield (Xu et al., 2013; Gao et al., 2020). In soybean, the silencing of *GmFWL1* expression resulted in a significant decrease in the number of nodules and changes in the structure of cell chromatin (Libault et al., 2010). In maize, a total of 13 *FWL* gene family members were identified, and named as *Cell Number Regulator (CNR)* genes (Guo et al., 2010). The *ZmCNR1* gene was illustrated to possess a plant-specific cell proliferation function affecting plant and fruit weight. In addition, it was also suggested that the expression of *ZmCNR2* negatively correlate with tissue growth activity and hybrid seedling vigor (Guo et al., 2010). These observations revealed that plant *FWL/CNR* genes play critical roles in plant development.

The maize *CNR* gene *ZmCNR13* was firstly identified through a *narrow odd dwarf* (*nod*) mutant (Rosa et al., 2017). Further evidence revealed that the *ZmCNR13* gene possessed the function in regulating cell division and differentiation and then affected both vegetative and reproductive development of maize (Rosa et al., 2017). However, the effect of this gene in the regulation of maize ear-related traits remains largely unknown, and there is no association analysis between the nucleotide polymorphisms of the maize *ZmCNR13* gene and yield-related traits. In this study, we investigated the nucleotide polymorphism of the maize *ZmCNR13* locus, and estimated the association between the sequence polymorphisms of this gene and some ear-related traits.

MATERIALS AND METHODS

Plant Materials, Experimental Design, and Analysis of Phenotypic Data

A total of 224 inbred lines (**Supplementary Table S1**) have been selected for phenotypic observation in this study. These lines had been grown in the field in a randomized block design with two replicates at Sanya (18°23′ N, 109°44′ E) in 2015, 2016 and Yangzhou (32°39′ N, 119°42′ E) in 2017. Each line was planted in a sequential row patterns with 15 plants, 3.5 m long and 0.4 m between adjacent rows. After harvesting and drying, three well-developed ears have been selected to measure ear traits, including ear weight (EW), ear grain weight (EGW), EL, ED, ERN, KNR, hundred kernel weight (HKW), KL, KW, and KT. ANOVA was performed using "aov" function in R software. The "lme4" (Bates et al., 2015) package was used to calculated the broad-sense heritability (h^2) for these ear-related traits. The

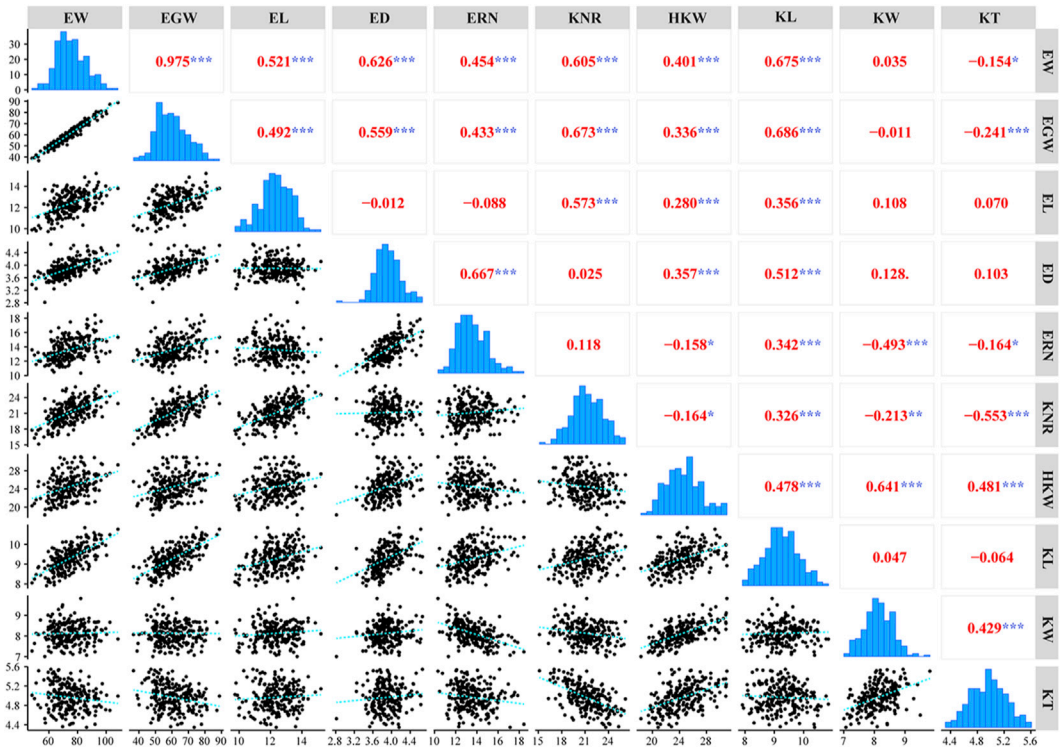


FIGURE 1 | Phenotypic distribution and Pearson correlation coefficients for ten ear-related traits (***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$).

TABLE 2 | Summary of parameters for the analysis of nucleotide polymorphisms.

Parameters	Upstream	5′UTR	Exons	Introns	3′UTR	Downstream	Full-length
Total length of amplicons (bp)	582	371	1,287	3,080	360	517	6,197
Number of all of the sequence variants	30	17	55	336	26	50	501
Frequency of all of the sequence variants	0.0515	0.0458	0.0427	0.1091	0.0722	0.0967	0.0808
Number of polymorphic sites	21	13	51	273	21	37	415
Frequency of polymorphic sites per bp	0.0361	0.0350	0.0396	0.0886	0.0583	0.0716	0.0670
Number of indels sites	31	11	60	120	37	48	306
Number of indels events	9	4	4	53	5	13	86
Average Indel length	3.4444	2.75	15	2.2642	7.4	3.6923	3.5581
Frequency of indels per bp	0.0155	0.0108	0.0031	0.0172	0.0139	0.0251	0.0139
π	0.00902	0.01093	0.00971	0.02090	0.02254	0.02391	0.01719
θ	0.00816	0.00613	0.00765	0.01702	0.01132	0.01544	0.01312
Tajima's *D*	0.2807	1.9086	0.7962	0.7203	2.6444[a]	1.5757	0.9854
Fu and Li's *D*	−1.2487	1.4749	1.0715	1.2195	1.2724	0.4676	1.1647
Fu and Li's *F*	−0.7670	1.9666[a]	1.1408	1.1546	2.1924[b]	1.1370	1.2750

[a]*Indicates a statistical significance at $p < 0.05$ level.*
[b]*Indicates a statistical significance at $p < 0.01$ level. "UTR" indicates untranslated region.*

observed values of these traits in three environments were used to calculate the best linear estimated prediction (BLUP) (Piepho et al., 2008) using the package 'lme4'. The calculation model of BLUP is: $y_{ij} = \mu + f_j + e_i + \epsilon_{ij}$, where y_{ij} is the observed value of the phenotype of j line in i environment, μ is the phenotypic mean of j line in all environments, f_j is the genetic effect of j line, e_i is the environment effect of i and ε_{ij} is the error of the observed value of the j line in i environment. The descriptive statistics and correlation coefficients were estimated using the "GGally" package in R software.

DNA Extraction and *ZmCNR13* Re-Sequencing

A total of 224 inbred lines, 56 landraces and 30 teosintes (**Supplementary Table S1**) were collected for target capture

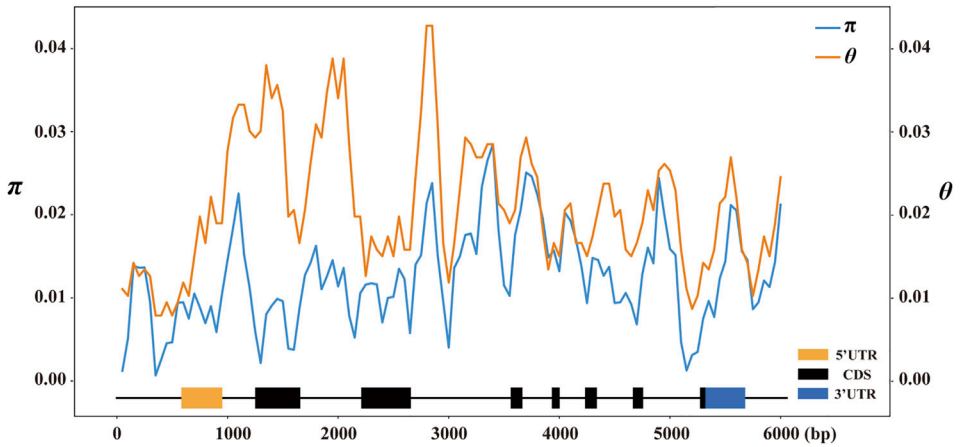


FIGURE 2 | Nucleotide diversity (π and θ) estimated along the sequences of maize *ZmCNR13*. π and θ were calculated using the method of sliding windows of 200 bp with a step of 50 bp.

TABLE 3 | Summary of nucleotide polymorphisms and neutrality test of *ZmCNR13*.

Populations	C	$\pi \times$ 1,000	$\theta \times$ 1,000	Tajima's *D*	Fu and Li's *D*	Fu and Li's *F*
teosintes	0.813	31.94	42.25	−0.9542	−1.3684	−1.4553
landraces	0.903	17.02	19.01	−0.3774	−0.5176	−0.5548
inbreds	0.914	17.19	13.12	0.9854	1.1647	1.2750
all	0.777	18.61	31.00	−1.2499	−4.5758[b]	−3.2829[b]

[a]*Indicates a statistical significance at $p < 0.05$ level.*
[b]*Indicates a statistical significance at $p < 0.01$ level.*

sequencing. Fresh and young leaves were collected from each line at the seeding stage, and a modified cetyl trimethyl ammonium bromide (CTAB) method was used to extract genomic DNA (Allen et al., 2006). DNA sequences of the *ZmCNR13* locus in the selected lines were resequenced using the target sequence capture sequencing technology on the NimbleGen platform (Choi et al., 2009) by BGI (Beijing Genomics Institute) Life Tech Co. The genomic sequences and positions of the *ZmCNR13* (GRMZM2G027821) locus in the inbred line B73 (AGPv3.21) were used as the references for target capture sequencing.

Analysis of Genotypic Data

The software Clustal X (Larkin et al., 2007) was used for multiple sequence alignment of the *ZmCNR13*. The nucleotide polymorphisms and allelic diversities of all tested lines were analyzed by DNASP5.0 software (Librado and Rozas 2009). Nucleotide diversity (π) in the *ZmCNR13* gene, which is defined as the mean number of nucleotide differences per site between any two DNA sequences, was estimated using R package "PopGenome" (Pfeifer et al., 2014). Linkage disequilibrium (LD) decay was measured by the squares of correlation coefficients (R^2) for all pairs of SNPs using the program PopLDdecay v3.41 (Zhang et al., 2019) with default parameters.

Marker-Trait Association Analysis in Inbred Lines

The method of genotyping-by-sequencing (GBS) was used to identify the genotypes of the selected lines (Li et al., 2019a). A total of 361,675 SNPs, which were distributed across the entire maize genome, were used to calculate the population structure and kinship. The population structure was estimated through the method of principal components (PCs). In this analysis, the top five PCs, which can explain 23.56% genetic variation, were used for association mapping. In addition, pair-wise coefficients of relatedness (kinship matrix) was calculated by TASSEL5.0 (Bradbury et al., 2007). In order to increase the accuracy of association analysis, two models were used for marker-trait association analysis. The general linear model (GLM) (Pritchard et al., 2000; Zhao et al., 2007) controlling for population structure (PCs), and mixed linear model (MLM) (Yu et al., 2006) controlling for both population structure (PCs) and kinship, were calculated by TASSEL5.0. A total of 398 markers in *ZmCNR13* with minor allele frequency (MAF) higher than 0.05 were used for association analysis. The *p*-value thresholds were empirically set to 1/398 and 0.05/398 for MLM and GLM, respectively, using the Bonferroni correction method (Bland and Altman 1995).

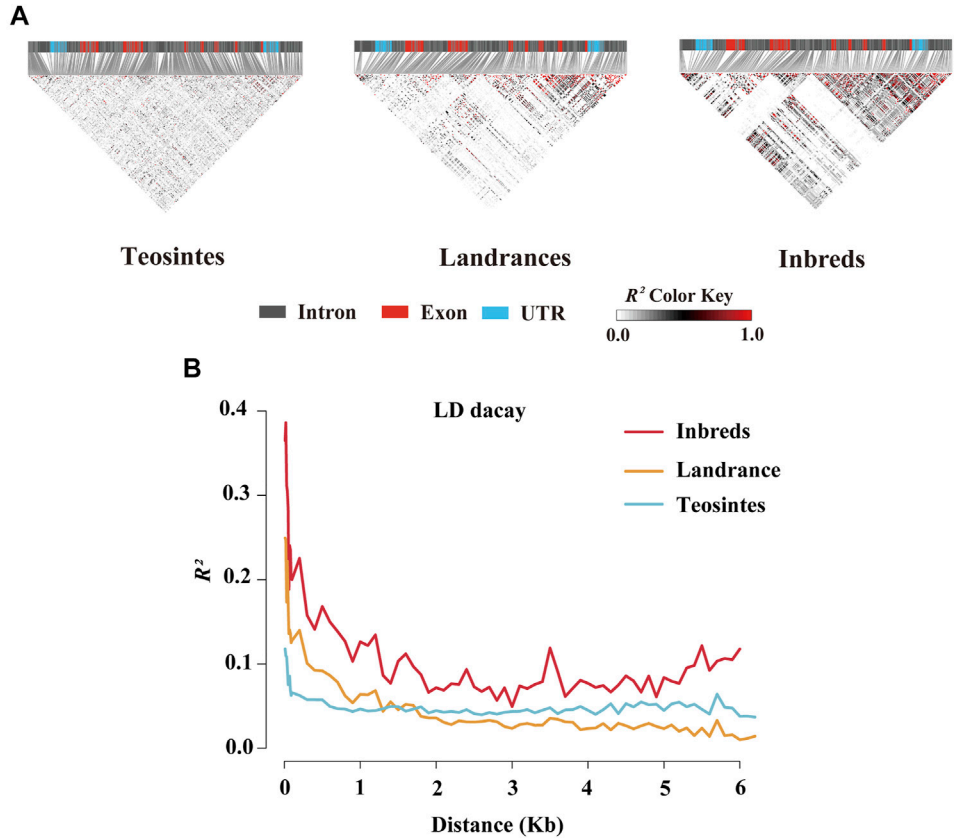


FIGURE 3 | Linkage disequilibrium (LD) analysis of *ZmCNR13* gene in three populations. **(A)** LD model of *ZmCNR13* gene in three populations. **(B)** LD decay for three populations measured by R^2.

RESULTS

The Phenotypic Variations Among Maize Inbred Lines

In this study, a total of ten ear-related traits, including EW, EGW, EL, ED, ERN, KNR, HKW, KL, KW, and KT, were obtained in a population of 224 maize inbred lines (**Table 1**). Coefficients of variation of these traits ranged from 12.11 to 40.79%, suggesting abundant phenotypic diversity among the tested inbred lines. ANOVA analyses also revealed that all these ear-related traits showed significant difference among inbred lines, suggesting that this population hold genetic characteristics for association analysis. In addition, we also noticed that both the environments and genotype-environment interaction had a significant impact on all these traits. The values of broad-sense heritability were further estimated. The results revealed that most of these traits possessed high heritability. The estimated heritability for ED, ERN, HKW, KL, and KW is higher than 40%. To evaluate the correlation relationship among these ten ear-related traits, paired correlation analysis was carried out, and the estimations of correlation coefficient (r) between any two traits were obtained (**Figure 1**). Significant correlations were observed between most parameters. Among them, EW/EGW had the highest correlation with r = 0.975. Meanwhile, the 12 of 45 pairwise correlations for ear-related traits, including ED/EL, ERN/EL, KNR/ED, KNR/ERN, KW/EW, KW/EGW, KW/EL, KW/ED, KW/KL, KT/EL, KT/ED and KT/KL didn't reach the significant level. These results indicate that different genetic mechanisms might affect the ear traits of maize.

Sequence Polymorphisms of the Maize *ZmCNR13* Gene

To evaluate the sequence polymorphisms of the *ZmCNR13* gene, the full-length sequences of this locus were re-sequenced in 224 inbred lines. After multiple sequence alignment, a full-length 6,197 bp sequence was obtained, including 582 bp upstream region, 5,098 bp coding region containing seven exons and seven introns, and 517 bp downstream region. A total of 501 variants were identified in the genomic sequence, including 415 SNPs and 86 indels covering 306 sites. On average, SNPs and Indels occurred per 12.37 bp and 72.06 bp, respectively. The highest frequency of variation sites was detected in the intron region (11.28 bp for SNP and 58.11 bp for InDels). The frequency of variation sites in the exon region was the lowest (23.23 bp for SNP and 321.75 bp for InDels). For all the tested lines, the overall nucleotide diversity (π) of the *ZmCNR13* locus was 0.01719, where the estimated π values of intron regions were relatively higher than exon regions. The nucleic

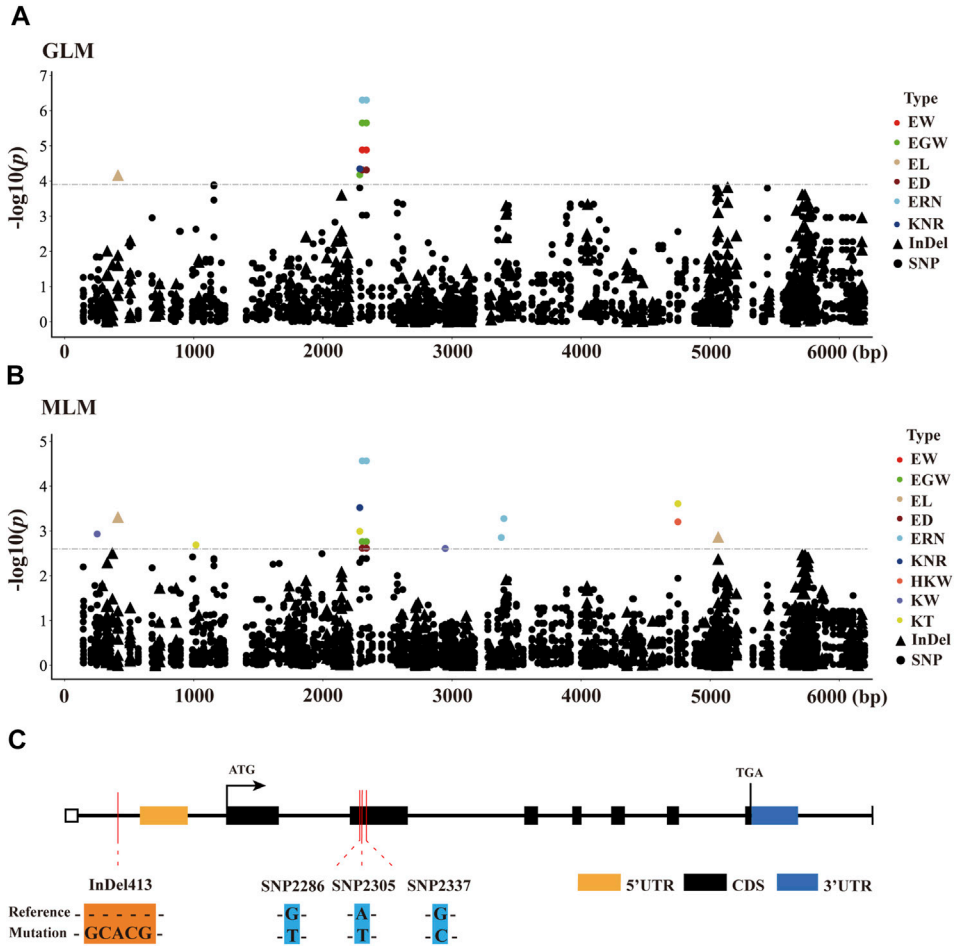


FIGURE 4 | Candidate-gene based association of *ZmCNR13* gene in inbred lines. **(A,B)** Manhattan plot using the MLM and GLM model. The *p*-value threshold was set at 1/398 (MLM) and 0.05/398 (GLM). Triangles and dots represent InDels and SNPs, respectively. **(C)** Schematic of the *ZmCNR13* gene structure and allelic variation.

acid polymorphisms of exon regions were low (θ = 0.00765), while that of other regions were relatively higher (θ = 0.01702 for intron regions and 0.01544 for downstream) (**Table 2**). In addition, π and θ were calculated using the sliding window of 200 bp with a step length of 50 bp. Two peaks in the intron3 and intron7 revealed that these regions were more diverse than others (**Figure 2**).

Nucleotide Diversity and Selection of *ZmCNR13* in Inbred Lines, Landraces and Teosintes

To further estimate the genetic diversity of the *ZmCNR13* locus, the sequence variation parameters of this gene in three different populations were analyzed and compared. The sequence conservation (C) of the *ZmCNR13* gene was similar in landraces (C = 0.903) and inbred lines (C = 0.914), and the lowest value of 0.813 in teosintes. Correspondingly, compared with teosintes, landraces and inbred lines showed lower nucleotide sequence polymorphisms (θ = 42.25 for teosintes, 19.01 for landraces and 13.12 for inbred lines). The neutrality of *ZmCNR13* gene was tested by Tajima's *D*, Fu and Li's *D**, and Fu and Li's *F**. The Tajima's *D* values of the three populations didn't achieve a significant level. Furthermore, the Tajima's *D* was positive in the inbred line population, indicating that the gene was lack of rare alleles in this population (**Table 3** and **Supplementary Figure S1**). There is no prominent LD block of the *ZmCNR13* gene in teosintes, and the degree of LD between mutation sites is relatively low. Compared with teosintes, the LD degree of landraces is enhanced to a certain extent, and smaller blocks begin to appear. In inbred lines, further the degree of LD is enhanced and larger blocks appear (**Figure 3A**). We also estimated the attenuation of LD with physical distance in three populations. The result revealed that the LD attenuation rate of inbred lines was slower than that of landraces and teosintes, and R^2 decreased to less than 0.2 after 200 bp, while those of landraces and teosintes decreased to less than 0.2 when it was less than 100 bp (**Figure 3B**). Taken together, although the gene didn't escape from neutral evolution, bottleneck effect of population has led to the lower nucleotide polymorphisms and LD decay in inbred lines.

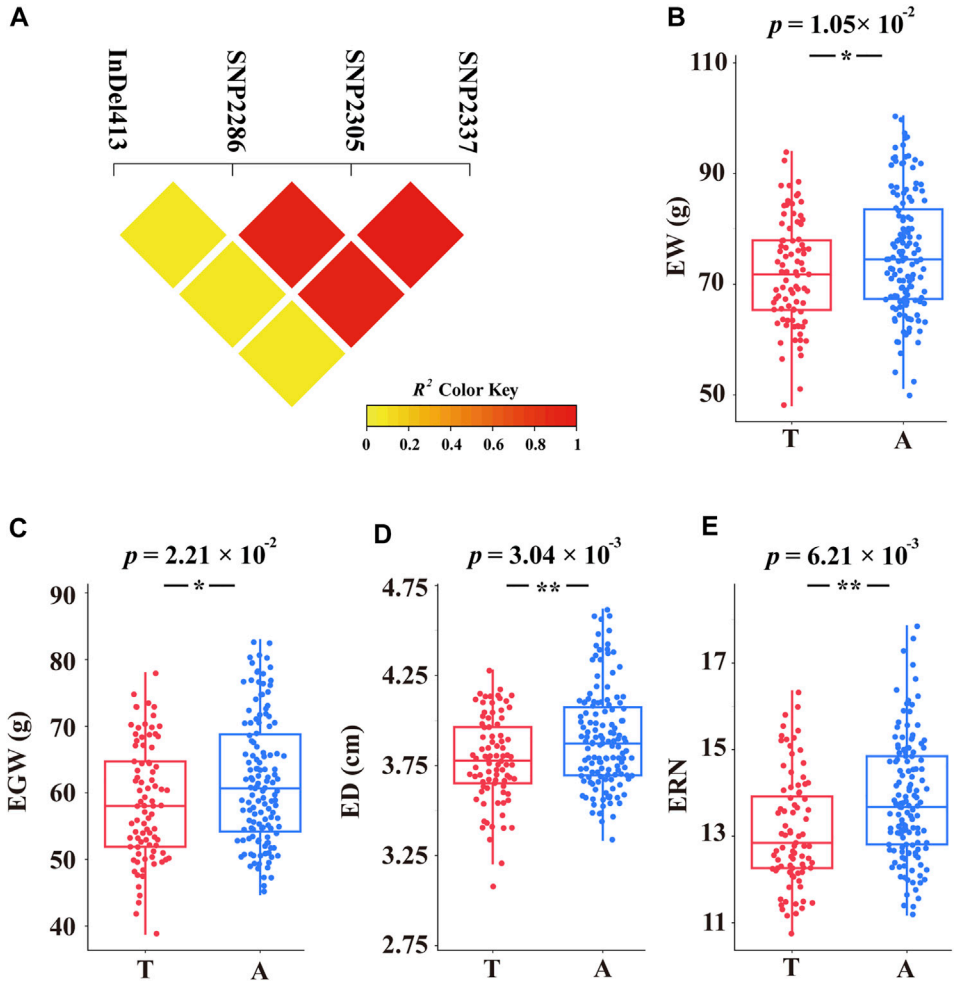


FIGURE 5 | Natural variations in *ZmCNR13* were significantly associated with four ear-related traits. **(A)** LD heatmap for four significant variants associated with ear traits. **(B–E)** Comparison of EW, EGW, ED and ERN between different alleles of SNP2305. *p* value for *t* test comparing two groups carrying different alleles were indexed on the top (**, $p < 0.01$; *, $p < 0.05$).

Association Analysis of Ear-Related Traits With *ZmCNR13*

To identify significant variants associated with ear-related traits, a total of 398 markers in *ZmCNR13* with minor allele frequency (MAF) higher than 0.05 were used for association analysis. Both methods of GLM and MLM were used for marker-trait association analysis. Four polymorphic sites (InDel413, SNP2286, SNP2305 and SNP2337) were captured jointly by two models, which were found to be significantly associated with six ear traits (EW, EGW, EL, ED, ERN, and KNR). Among them, two sites (SNP2305 and SNP2337) located in exon 2 jointly controlled four ear traits (EW, EGW ED and ERN). Furthermore, we noticed that SNP2305 belongs to non-synonymous mutation, where the transition of T to A will lead to the changes of amino acid serine to threonine. In addition, SNP2305 were found to be associated with EW, EGW, ED and ERN, and could explain 4.31–8.42% and 7.06–10.41% phenotypic variations under MLM and GLM, respectively (**Figure 4** and **Supplementary Table S2**).

LD analysis showed that SNP2286, SNP2305 and SNP2337 had relatively high linkage across inbred lines. Interestingly, SNP2305 and SNP2337 had complete linkage ($R^2 = 1$) (**Figure 5A**). We further classified haplotypes based on the variation of SNP2305, and divided the inbred lines into two major groups. *T*-test revealed that EW, EGW, ED and ERN showed significant difference between two groups. The allele A group possessed significantly higher values of EW, EGW, ED and ERN than the allele T (**Figures 5B–E**). On this basis, combined with correlation analysis, we noticed that the changes among the four traits were synergistic. Therefore, we can infer that allele A of SNP2305 has positive regulatory effects on EW, EGW, ED and ERN.

DISCUSSION

In the present study, association analyses were employed to illustrate the relationship between the maize *ZmCNR13* and

ear-related traits. Association analysis, also known as linkage disequilibrium mapping or association mapping, is a method based on linkage disequilibrium to identify the association between genetic polymorphisms and phenotypic variations (Mackay and Powell 2007). Compared with linkage analysis based on parental population, association analysis has higher resolution, and it can be directly used to analyze more than two alleles in natural populations (Salvi and Tuberosa 2005; Mackay and Powell 2007). As an extension of genome-wide association mapping, candidate gene association analysis is mainly used to identify genetic variations and excellent haplotypes significantly associated with target traits (Yan et al., 2011). Maize is a typical cross-pollinated plant with a high recombination rate, rich genetic diversity and LD decay distance of about 1Kb, so it is an ideal material for association analysis (Tenaillon et al., 2001; Whitt et al., 2002). Many loci affecting maize traits have been identified by this method, such as *Zmisa2* (Yang et al., 2014) and *ZmBT1* (Xu et al., 2014) for starch properties, *ZmYS1* (Yang et al., 2015) for kernel mineral concentrations, *ZmCKX5* (Wang et al., 2021) and *ZmMADS60* (Li et al., 2020) for root morphology, *ZmHKT1* (Li et al., 2019a) and *ZmPGP1* (Li et al., 2019b) for plant architecture.

Abundant genetic diversity is the basis for crop improvement (Yan et al., 2011). In this study, nucleotide polymorphisms of the *ZmCNR13* gene were analyzed in inbred lines through re-sequencing. A total of 501 variations including 415 SNPs and 86 InDels were detected, and most of them concentrated in the intron regions. The exon region had one SNP per 25.71 bp, while the intron region reached one SNP per 11.28 bp. The decrease of nucleotide polymorphism in exon region suggested that these regions might be influenced by greater selection pressure. Moreover, it is worth noted that the LD decays faster in landraces and teosintes than inbred lines, suggesting that there were genetic bottleneck effects (Wang et al., 1998) in the population of inbred lines, although no obvious selection signature was detected through neutral test.

Ear-related traits are the most important components of kernel yield of maize. Illustrating the genetic background of the genes related to ear traits and digging their elite variations will be of great importance in high-yield breeding maize (Zhu et al., 2018). In the present study, we revealed that a non-synonymous mutation in exon 2 (SNP2305) of the *ZmCNR13* gene was found to be significantly associated with four ear-related traits, including EW, EGW, ED and ERN. In addition, we further noticed that the inbred lines carrying SNP2305A possess higher EW, EGW, ED and ERN than those carrying SNP2305T. The validity of candidate gene association analysis has been repeatedly confirmed (Yan et al., 2011). For instance, the association between polymorphisms of *dwarf8* and flowering time was detected by some independent works (Thornsberry et al., 2001; Andersen et al., 2005; Camus-Kulandaivelu et al., 2006). To confirm the results of association analysis, we compared the phenotypes between the inbred lines carrying SNP2305A and SNP2305T in three environments. The results revealed that there are statistical differences between them for all these traits in different environments (**Supplementary Figure S2**), suggestive the genetical credibility of the association results. These observations revealed that the superior allelic variations of *ZmCNR13* possess potential application values in maize genetic improvement.

In conclusion, the maize *ZmCNR13* gene was re-sequenced in 224 inbred lines, 56 landraces and 30 teosintes. Although this gene was escaped from artificial selection during maize genetic improvement, a non-synonymous mutation in exon2 was found to be associated with ear-related traits, including EW, EGW, ED and ERN. The superior allelic variations of *ZmCNR13* has potential application values in maize genetic improvement.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here:PRJNA764471.

AUTHOR CONTRIBUTIONS

ZZ, CX, and ZY. conceived and designed the research. ZZ, YL, MZ, RC, EZ, DH, QH, and HW. conducted the experiments. ZZ, YL, YS, ZW, YX, and PL analyzed the data. ZZ, CX, and ZY wrote the manuscript. All authors read and approved the manuscript.

FUNDING

This work was supported by the grant from the National Natural Science Foundation of China (32070558, 32061143030, 31801028, 31972487, 32100448), a project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), and Natural Science Foundation of Jiangsu Province (BK20210799).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.773597/full#supplementary-material

REFERENCES

Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S., and Thompson, W. F. (2006). A Modified Protocol for Rapid DNA Isolation from Plant Tissues Using Cetyltrimethylammonium Bromide. *Nat. Protoc.* 1 (5), 2320–2325. doi:10.1038/nprot.2006.384

Andersen, J. R., Schrag, T., Melchinger, A. E., Zein, I., and Lübberstedt, T. (2005). Validation of Dwarf8 Polymorphisms Associated with Flowering Time in Elite European Inbred Lines of maize (Zea mays L.). *Theor. Appl. Genet.* 111 (2), 206–217. doi:10.1007/s00122-005-1996-6

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *J. Stat. Soft.* 67 (1):1–48. doi:10.18637/jss.v067.i01

Beavis, W. D., Smith, O. S., Grant, D., and Fincher, R. (1994). Identification of Quantitative Trait Loci Using a Small Sample of Topcrossed and F4 Progeny from maize. *Crop Sci.* 34 (4), 882–896. doi:10.2135/cropsci1994.0011183X003400040010x

Bland, J. M., and Altman, D. G. (1995). Statistics Notes: Multiple Significance Tests: the Bonferroni Method. *BMJ* 310 (6973), 170. doi:10.1136/bmj.310.6973.170

Bommert, P., Lunde, C., Nardmann, J., Vollbrecht, E., Running, M., Jackson, D., et al. (2005). Thick Tassel Dwarf1 Encodes a Putative maize Ortholog of the Arabidopsis CLAVATA1 Leucine-Rich Repeat Receptor-like Kinase. *Development* 132 (6), 1235–1245. doi:10.1242/dev.01671

Bommert, P., Nagasawa, N. S., and Jackson, D. (2013). Quantitative Variation in maize Kernel Row Number Is Controlled by the FASCIATED EAR2 Locus. *Nat. Genet.* 45 (3), 334–337. doi:10.1038/ng.2534

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

Camus-Kulandaivelu, L., Veyrieras, J.-B., Madur, D., Combes, V., Fourmann, M., Barraud, S., et al. (2006). Maize Adaptation to Temperate Climate: Relationship between Population Structure and Polymorphism in the Dwarf8 Gene. *Genetics* 172 (4), 2449–2463. doi:10.1534/genetics.105.048603

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., et al. (2009). Genetic Diagnosis by Whole Exome Capture and Massively Parallel DNA Sequencing. *Pnas* 106 (45), 19096–19101. doi:10.1073/pnas.0910672106

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Knaap, E. v. d., Cong, B., et al. (2000). fw2.2 : A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size. *Science* 289 (5476), 85–88. doi:10.1126/science.289.5476.85

Gao, Q., Li, G., Sun, H., Xu, M., Wang, H., Ji, J., et al. (2020). Targeted Mutagenesis of the rice FW 2.2-Like Gene Family Using the CRISPR/Cas9 System Reveals OsFWL4 as a Regulator of Tiller Number and Plant Yield in rice. *Ijms* 21 (3), 809. doi:10.3390/ijms21030809

Guo, M., Rupe, M. A., Dieter, J. A., Zou, J., Spielbauer, D., Duncan, K. E., et al. (2010). Cell Number Regulator1 Affects Plant and Organ Size in Maize: Implications for Crop Yield Enhancement and Heterosis. *Plant Cell* 22 (4), 1057–1073. doi:10.1105/tpc.109.073676

Je, B. I., Gruel, J., Lee, Y. K., Bommert, P., Arevalo, E. D., Eveland, A. L., et al. (2016). Signaling from maize Organ Primordia via FASCIATED EAR3 Regulates Stem Cell Proliferation and Yield Traits. *Nat. Genet.* 48 (7), 785–791. doi:10.1038/ng.3567

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X Version 2.0. *Bioinformatics* 23 (21), 2947–2948. doi:10.1093/bioinformatics/btm404

Li, C., Li, Y., Sun, B., Peng, B., Liu, C., Liu, Z., et al. (2013). Quantitative Trait Loci Mapping for Yield Components and Kernel-Related Traits in Multiple Connected RIL Populations in maize. *Euphytica* 193 (3), 303–316. doi:10.1007/s10681-013-0901-7

Li, M., Zhong, W., Yang, F., and Zhang, Z. (2018). Genetic and Molecular Mechanisms of Quantitative Trait Loci Controlling maize Inflorescence Architecture. *Plant Cel Physiol* 59 (3), 448–457. doi:10.1093/pcp/pcy022

Li, P., Ge, Z., Wang, H., Wei, J., Wang, Y., Xu, Y., et al. (2020). Nucleotide Diversity and Association Analysis of ZmMADS60 with Root Length in the maize Seedling Stage. *Agronomy* 10 (3), 342. doi:10.3390/agronomy10030342

Li, P., Pan, T., Wang, H., Wei, J., Chen, M., Hu, X., et al. (2019a). Natural Variation of ZmHKT1 Affects Root Morphology in maize at the Seedling Stage. *Planta* 249 (3), 879–889. doi:10.1007/s00425-018-3043-2

Li, P., Wei, J., Wang, H., Fang, Y., Yin, S., Xu, Y., et al. (2019b). Natural Variation and Domestication Selection of ZmPGP1 Affects Plant Architecture and Yield-Related Traits in maize. *Genes* 10 (9), 664. doi:10.3390/genes10090664

Libault, M., and Stacey, G. (2010). Evolution of FW2.2-like (FWL) and PLAC8 Genes in Eukaryotes. *Plant Signaling Behav.* 5 (10), 1226–1228. doi:10.4161/psb.5.10.12808

Libault, M., Zhang, X.-C., Govindarajulu, M., Qiu, J., Ong, Y. T., Brechenmacher, L., et al. (2010). A Member of the Highly Conserved FWL (Tomato FW2.2-like) Gene Family Is Essential for Soybean Nodule Organogenesis. *Plant J.* 62 (5), 852–864. doi:10.1111/j.1365-313X.2010.04201.x

Librado, P., and Rozas, J. (2009). DnaSP V5: a Software for Comprehensive Analysis of DNA Polymorphism Data. *Bioinformatics* 25 (11), 1451–1452. doi:10.1093/bioinformatics/btp187

Liu, L., Du, Y., Shen, X., Li, M., Sun, W., Huang, J., et al. (2015). KRN4 Controls Quantitative Variation in maize Kernel Row Number. *Plos Genet.* 11 (11), e1005670. doi:10.1371/journal.pgen.1005670

Liu, R., Jia, H., Cao, X., Huang, J., Li, F., Tao, Y., et al. (2012). Fine Mapping and Candidate Gene Prediction of a Pleiotropic Quantitative Trait Locus for Yield-Related Trait in Zea mays. *PLoS One* 7 (11), e49836. doi:10.1371/journal.pone.0049836

Mackay, I., and Powell, W. (2007). Methods for Linkage Disequilibrium Mapping in Crops. *Trends Plant Sci.* 12 (2), 57–63. doi:10.1016/j.tplants.2006.12.001

McSteen, P., and Hake, S. (2001). Barren Inflorescence2 Regulates Axillary Meristem Development in the maize Inflorescence. *Development* 128 (15), 2881–2891. doi:10.1242/dev.128.15.2881

Messmer, R., Fracheboud, Y., Bänziger, M., Vargas, M., Stamp, P., and Ribaut, J.-M. (2009). Drought Stress and Tropical maize: QTL-By-Environment Interactions and Stability of QTLs across Environments for Yield Components and Secondary Traits. *Theor. Appl. Genet.* 119 (5), 913–930. doi:10.1007/s00122-009-1099-x

Nesbitt, T. C., and Tanksley, S. D. (2001). Fw2.2 Directly Affects the Size of Developing Tomato Fruit, with Secondary Effects on Fruit Number and Photosynthate Distribution. *Plant Physiol.* 127 (2), 575–583. doi:10.1104/pp.010087

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., and Lercher, M. J. (2014). PopGenome: an Efficient Swiss Army Knife for Population Genomic Analyses in R. *Mol. Biol. Evol.* 31 (7), 1929–1936. doi:10.1093/molbev/msu136

Piepho, H. P., Möhring, J., Melchinger, A. E., and Büchse, A. (2008). BLUP for Phenotypic Selection in Plant Breeding and Variety Testing. *Euphytica* 161 (1), 209–228. doi:10.1007/s10681-007-9449-8

Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association Mapping in Structured Populations. *Am. J. Hum. Genet.* 67 (1), 170–181. doi:10.1086/302959

Raihan, M. S., Liu, J., Huang, J., Guo, H., Pan, Q., and Yan, J. (2016). Multi-environment QTL Analysis of Grain Morphology Traits and fine Mapping of a Kernel-Width QTL in Zheng58 × SK maize Population. *Theor. Appl. Genet.* 129 (8), 1465–1477. doi:10.1007/s00122-016-2717-z

Ritter, M. K., Padilla, C. M., and Schmidt, R. J. (2002). The maize Mutant Barren Stalk1 Is Defective in Axillary Meristem Development. *Am. J. Bot.* 89 (2), 203–210. doi:10.3732/ajb.89.2.203

Rosa, M., Abraham-Juárez, M. J., Lewis, M. W., Fonseca, J. P., Tian, W., Ramirez, V., et al. (2017). The Maize MID-Complementing Activity Homolog Cell Number Regulator13/Narrow odd Dwarf Coordinates Organ Growth and Tissue Patterning. *Plant Cell* 29 (3), 474–490. doi:10.1105/tpc.16.00878

Salvi, S., and Tuberosa, R. (2005). To Clone or Not to Clone Plant QTLs: Present and Future Challenges. *Trends Plant Sci.* 10 (6), 297–304. doi:10.1016/j.tplants.2005.04.008

Sidwell, R. J., Smith, E. L., and McNew, R. W. (1976). Inheritance and Interrelationships of Grain Yield and Selected Yield-Related Traits in a Hard Red Winter Wheat Cross 1. *Crop Sci.* 16 (5), 650–654. doi:10.2135/cropsci1976.0011183X001600050013x

Tenaillon, M. I., Sawkins, M. C., Long, A. D., Gaut, R. L., Doebley, J. F., and Gaut, B. S. (2001). Patterns of DNA Sequence Polymorphism along Chromosome 1 of maize (Zea mays Ssp. Mays L.). *Proc. Natl. Acad. Sci.* 98 (16), 9161–9166. doi:10.1073/pnas.151244298

Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D., and Buckler, E. S. (2001). Dwarf8 Polymorphisms Associate with Variation in Flowering Time. *Nat. Genet.* 28 (3), 286–289. doi:10.1038/90135

Wang, H., Sun, H., Xia, H., Wu, T., Li, P., Xu, C., et al. (2021). Natural Variation and Domestication Selection of ZmCKX5 with Root Morphological Traits at the Seedling Stage in maize. *Plants* 10 (1), 1. doi:10.3390/plants10010001

Wang, J., Caballero, A., and Hill, W. G. (1998). The Effect of Linkage Disequilibrium and Deviation from Hardy-Weinberg Proportions on the Changes in Genetic Variance with Bottlenecking. *Heredity* 81 (2), 174–186. doi:10.1046/j.1365-2540.1998.00390.x

Wang, J., Lin, Z., Zhang, X., Liu, H., Zhou, L., Zhong, S., et al. (2019). krn1 , a Major Quantitative Trait Locus for Kernel Row Number in maize. *New Phytol.* 223 (3), 1634–1646. doi:10.1111/nph.15890

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

Xu, J., Xiong, W., Cao, B., Yan, T., Luo, T., Fan, T., et al. (2013). Molecular Characterization and Functional Analysis of "fruit-weight2.2-like" Gene Family in rice. *Planta* 238, 643–655. doi:10.1007/s00425-013-1916-y

Xu, S., Yang, Z., Zhang, E., Jiang, Y., Pan, L., Chen, Q., et al. (2014). Nucleotide Diversity of maize ZmBT1 Gene and Association with Starch Physicochemical Properties. *PLoS One* 9 (8), e103627. doi:10.1371/journal.pone.0103627

Yan, J.-b., Tang, H., Huang, Y.-q., Zheng, Y.-l., and Li, J.-s. (2006). Quantitative Trait Loci Mapping and Epistatic Analysis for Grain Yield and Yield Components Using Molecular Markers with an Elite maize Hybrid. *Euphytica* 149 (1), 121–131. doi:10.1007/s10681-005-9060-9

Yan, J., Warburton, M., and Crouch, J. (2011). Association Mapping for Enhancing maize (Zea mays L.) Genetic Improvement. *Crop Sci.* 51 (2), 433–449. doi:10.2135/cropsci2010.04.0233

Yang, Z., Ma, S., Hu, Y., Zhang, E., Xie, Z., Xu, S., et al. (2015). Association Analysis of the maize Gene ZmYS1 with Kernel mineral Concentrations. *Plant Mol. Biol. Rep.* 33 (5), 1327–1335. doi:10.1007/s11105-014-0836-8

Yang, Z., Zhang, E., Jiang, Y., Xu, S., Pan, L., Chen, Q., et al. (2014). Sequence Polymorphisms in Zmisa2 Gene Are Significantly Associated with Starch Pasting and Gelatinization Properties in maize (Zea mays L.). *Mol. Breed.* 34 (4), 1833–1842. doi:10.1007/s11032-014-0142-z

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A Unified Mixed-Model Method for Association Mapping that Accounts for Multiple Levels of Relatedness. *Nat. Genet.* 38 (2), 203–208. doi:10.1038/ng1702

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2019). PopLDdecay: a Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files. *Bioinformatics* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875

Zhang, C., Zhou, Z., Yong, H., Zhang, X., Hao, Z., Zhang, F., et al. (2017). Analysis of the Genetic Architecture of maize Ear and Grain Morphological Traits by Combined Linkage and Association Mapping. *Theor. Appl. Genet.* 130 (5), 1011–1029. doi:10.1007/s00122-017-2867-7

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., et al. (2007). An Arabidopsis Example of Association Mapping in Structured Samples. *Plos Genet.* 3 (1), e4. doi:10.1371/journal.pgen.0030004

Zhu, X.-M., Shao, X.-Y., Pei, Y.-H., Guo, X.-M., Li, J., Song, X.-Y., et al. (2018). Genetic Diversity and Genome-wide Association Study of Major Ear Quantitative Traits Using High-Density SNPs in maize. *Front. Plant Sci.* 9, 966. doi:10.3389/fpls.2018.00966

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 November 2021
doi: 10.3389/fgene.2021.728472

The Genetic Architecture of Grain Yield in Spring Wheat Based on Genome-Wide Association Study

Yuyao Li[1,2†], Jingquan Tang[3†], Wenlin Liu[3], Wenyi Yan[2], Yan Sun[3], Jingyu Che[4], Chao Tian[4], Hongji Zhang[3] and Lihe Yu[1,2*]*

[1] Heilongjiang Bayi Agricultural University, Daqing, China, [2] Heilongjiang Academy of Agricultural Sciences, Harbin, China, [3] Crop Resources Institute, Heilongjiang Academy of Agricultural Sciences, Harbin, China, [4] Keshan Branch, Heilongjiang Academy of Agricultural Sciences, Qiqihar, China

Uncovering the genetic architecture for grain yield (GY)–related traits is important for wheat breeding. To detect stable loci for GY-related traits, a genome-wide association study (GWAS) was conducted in a diverse panel, which included 251 elite spring wheat accessions mainly from the Northeast of China. In total, 52,503 single nucleotide polymorphisms (SNPs) from the wheat 55 K SNP arrays were used. Thirty-eight loci for GY-related traits were detected and each explained 6.5–16.7% of the phenotypic variations among which 12 are at similar locations with the known genes or quantitative trait loci and 26 are likely to be new. Furthermore, six genes possibly involved in cell division, signal transduction, and plant development are candidate genes for GY-related traits. This study provides new insights into the genetic architecture of GY and the significantly associated SNPs and accessions with a larger number of favorable alleles could be used to further enhance GY in breeding.

Keywords: GWAS, marker-assisted selection, single nucleotide polymorphism, *Triticum aestivum*, grain yield

OPEN ACCESS

Edited by:
Ahmed Sallam,
Assiut University, Egypt

Reviewed by:
Karansher Singh Sandhu,
Washington State University, United States
Mohamed El-Soda,
Cairo University, Egypt

****Correspondence:***
Hongji Zhang
fumai@163.com
Lihe Yu
yulihe2002@126.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *21 June 2021*
Accepted: *30 August 2021*
Published: *15 November 2021*

Citation:
Li Y, Tang J, Liu W, Yan W, Sun Y, Che J, Tian C, Zhang H and Yu L (2021) The Genetic Architecture of Grain Yield in Spring Wheat Based on Genome-Wide Association Study. Front. Genet. 12:728472. doi: 10.3389/fgene.2021.728472

BACKGROUND

Common wheat is the most important food crop worldwide and provides nearly 20% of the total caloric input to the global population. Grain yield (GY) improvement is one of the challenging goals in wheat breeding due to the complex genetic architecture and low heritability (He et al., 2010; Tester and Langridge, 2010; Gao et al., 2017; Wang et al., 2019). Heilongjiang and Jilin of China are the major spring wheat-producing regions, and yield potential in this region has been improved largely in the past decades (He et al., 2010; Zhou et al., 2014; Qin et al., 2015). However, the wheat production in this region is facing various threats, such as decreased groundwater, resulting in declining growing area, the irrigation frequency of wheat, and the bottleneck for yield potential of new cultivars through conventional breeding (He et al., 2010; Tester and Langridge, 2010).

Grain yield is a complex trait and influenced by many factors, particularly the genetic factors. GY-related traits included the spike number per unit area (SN), kernel number per spike (KNS), and thousand-kernel weight (TKW) (Ellis et al., 2005; Zhou et al., 2007; Rasheed et al., 2016; Gao et al., 2017; Li et al., 2018). Marker-assisted selection (MAS) is an effective tool for the further

Abbreviations: BLUE, best linear unbiased estimation; GY, grain yield; GWAS, genome-wide association study; h^2, broad-sense heritability; KASP, kompetitive allele-specific PCR; KNS, kernel number per spike; LD, linkage disequilibrium; MAS, marker-assisted selection; QTL, quantitative trait loci; R^2, phenotypic variance explained; RIL, recombinant inbred line; SL, spike length; SN, spike number per unit area; SNP, single nucleotide polymorphism; TKW, thousand-kernel weight.

improvement of yield potential. Also, MAS is a key technique to increase the yield of wheat (He et al., 2010; Rasheed et al., 2016).

The effectiveness and reliability of MAS depends on the number of available genes and tightly linked markers for target traits. Until now, more than 70 genes have been cloned in wheat, among which 40 are associated with GY and related traits (Cui et al., 2014; Rasheed et al., 2016; Nadolska-Orczyk et al., 2017; Li et al., 2018; Wang et al., 2019). For all the cloned genes, about 150 functional markers or kompetitive allele-specific PCR (KASP) were developed (Rasheed et al., 2016). Besides this, more than 100 loci identified by genome-wide association study (GWAS) or biparental linkage mapping for GY-related traits are reported (Cui et al., 2014; Gao et al., 2017; Würschum et al., 2017; Li et al., 2018). However, identifying the novel genes or loci for GY is still important for wheat production.

Single nucleotide polymorphisms (SNP) provide an effective way to identify candidate genes for various traits (Zhu et al., 2008; Wang et al., 2014; Rasheed et al., 2016). Recently, the wheat 55, 90, and 660 K SNP arrays are gradually replacing simple sequence repeats (SSR) and diversity array technology (DArT) markers in genetic analysis in yield, disease resistance, end-use quality, and biotic or abiotic stress tolerance–related traits (Jin et al., 2016; Li et al., 2016; Liu et al., 2016, 2017; Valluru et al., 2017; Quan et al., 2021). Linkage analysis based on biparental populations and association mapping based on natural populations are two main ways to uncover the genetic analysis of complex traits (Zhu et al., 2008; Liu et al., 2017). Compared with linkage analysis, association mapping is based on linkage disequilibrium (LD) and offers an effective and reliable approach to uncover the genetic architecture of complex traits (Zhu et al., 2008; Liu et al., 2017). GWAS uses the natural germplasms, including wild types, landraces, released cultivars, and improved accessions, as materials and bypasses the time of developing biparental populations (Sela et al., 2014; Shi et al., 2017). Furthermore, traditional linkage analysis focuses on specific traits, whereas association mapping can be used to analyze various traits based on the same genotype data (Zhu et al., 2008). Nowadays, GWAS is commonly applied in genetic analysis of complex traits in wheat, such as GY-related traits, disease-related traits (stripe rust, leaf rust, or powdery mildew), and biotic and abiotic stress (Cui et al., 2011, 2014; Azadi et al., 2015; Beyer et al., 2019; Quan et al., 2021).

In China, Heilongjiang and Jilin are the main zones for spring wheat. In this study, 251 spring wheat accessions mainly originating from Heilongjiang and Jilin (1930–2020s) were selected. The aims of this study were to (1) detect the loci for GY and related traits in spring wheat, and (2) search for candidate genes for GY-related traits for further study.

MATERIALS AND METHODS

Plant Materials and Field Trials

The diverse panel used in the present study contained 251 varieties mainly from Heilongjiang or Jilin province of China (**Supplementary Table 1**). The diverse panel was grown at Haerbin and Keshan in Heilongjiang province during the 2019 and 2020 cropping seasons. A randomized complete block design with three replicates was employed in field trials. Each plot comprised four 2.0-m rows spaced 20 cm apart with 40 seeds in each row. Agronomic management was performed according to local practices at each location.

Phenotyping and Statistical Analysis

Six phenotypic traits related to GY were evaluated in all four environments, including the GY, spike number per unit area (SNU), SN, spike length (SL), KNS, and TKW. The middle two lines of plants were harvested in each plot at physiological maturity, and GY was measured after the seed water content dried to 14% and was expressed as kg ha^{-1}. The investigation of the other five traits and statistical analysis were according to Li et al. (2018). BLUP estimation for six traits among four environments was calculated using the MIXED procedure (PROCMIXED) in SAS v9.3 (SAS Institute)[1] following the formula:

$$y = \mathrm{Xb} + \mathrm{Zu} + \mathrm{e}.$$

Of these, y is the observed phenotype, Xb is the fixed effects (environment), Zu is the random effect (genotype), and e is the residual effect.

Genotyping, Population Structure, and Linkage Disequilibrium

The 251 accessions were genotyped using the wheat 55 K SNP arrays. SNP markers with missing data > 20% and minor allele frequency (MAF) < 0.05 were removed to avoid spurious alleles. The physical position for subsequent GWAS analysis followed the IWGSC RefSeq v2.1[2].

Polymorphic and evenly distributed markers (one marker/LD block) were used to conduct population structure analysis. In this study, after filtering by MAF and missing data, 3000 polymorphic SNP markers (about 5 Mb for the whole genome of common wheat according to previous studies) evenly distributed on 21 chromosomes were analyzed in Structure v2.3.4 (Pritchard et al., 2000)[3].

To verify the result, PCA and neighbor-jointing (NJ) trees were also estimated using the software Tassel v5.0 (Breseghello and Sorrells, 2006), respectively. After filtering using the Tassel v5.0, 5000 evenly distributed SNP markers were chosen to calculate LD for whole genomes using the full matrix and sliding window options in Tassel v5.0 (Breseghello and Sorrells, 2006). The details for population structure and LD decay analysis followed Liu et al. (2017). Variance components were used to calculate broad sense heritability (h_b^2) of GY-related traits as $h_b^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{ge}^2/\mathrm{r} + \sigma_\varepsilon^2/\mathrm{re})$, where σ_g^2, σ_{ge}^2, and σ_ε^2 represent the genotype, genotype × environment interaction, and residual error variances, respectively, and e and r are the numbers of environments and replicates per environment, respectively.

[1] http://www.sas.com

[2] https://urgi.versailles.inra.fr/blast_iwgsc/blast.php

[3] http://pritchardlab.stanford.edu/structure.html

Genome-Wide Association Mapping

To control background variation and eliminate the spurious marker-trait associations (MTAs), the mixed linear model (MLM, PCA + K model) in Tassel v5.0 was used in consideration of the kinship matrix and population structure as follows:

$$y = \mu + x\beta + u + e.$$

Of these, y is the vector of observed phenotype, μ is the mean, x is the genotype, β is the effect of the SNP, u is the random effects due to genetic relatedness with Var (u) = σ^2g K and Var (e) = σ^2e, and K is the kinship matrix across all genotypes. The kinship matrix was treated as a random-effect factor and calculated by the software Tassel v5.0, whereas the PCA was considered as a fixed-effect factor and inferred by the Tassel V5.0 in MLM analysis. Due to Bonferroni–Holm correction for multiple testing (α = 0.05) being too conserved, and no significant MTAs were detected with this criterion, markers with an adjusted -log10 (*P*-value) ≥ 3.0 were regarded as significant markers. Manhattan plots and Q-Q plots were drawn based on R language (R 3.6.5).

The Identification of Candidate Genes

Loci existing in two or more environments are considered stable. To identify candidate genes, the flanking sequences corresponding to the SNP markers (including the SNPs located in the LD decay interval for peak markers) significantly associated with GY-related traits are used in BLASTn and BLASTx searches against NCBI[4] and ENA databases. Also, the annotation information for IWGSC 2.1 was also used to identify candidate genes.

RESULTS

Phenotypic Evaluation

Significant and continuous variations of GY and related traits were observed in the diverse panel (**Supplementary Table 1** and **Supplementary Figure 1**). The mean values of SNU, SN, SL, KNS, TKW, and GY were 131.1 (88.7–177.9), 16.6 (12.9–20.7), 10.6 (6.8–13.8), 36.0 (23.1–50.6), 32.4 g (20.5–41.2 g), and 4623.5 kg.hm^{-1} (2233.0–6987.4 kg.hm^{-1}). ANOVA showed highly significant effects ($P < 0.01$) of genotypes, environments, and genotype × environment interactions on all traits (**Table 1**) that were exhibited. The broad sense heritability (h^2) for SNU, SN, SL, KNS, TKW, and GY were 0.64, 0.52, 0.65, 0.71, 0.63, and 0.49, respectively, which indicates most of these traits are stable and mainly determined by genetic factors.

In this study, GY showed significant ($P < 0.01$) and positive correlations with SNU, SN, SL, KNS, and TKW (r ranged from 0.18 to 0.75). SNU showed significant ($P < 0.01$) and positive correlations with GY ($r = 0.18$), whereas significant and negative correlations with SN, SL, KNS, and TKW (r ranged from 0.19 to 0.24); SN showed significant ($P < 0.01$) and positive correlations with SL, KNS, and TKW (r ranged from 0.10 to 0.82); SL exhibited significant ($P < 0.01$) and positive correlations with KNS and TKW (r ranged from 0.21 to 0.71) (**Supplementary Table 2**).

Marker Coverage and Genetic Diversity of the Physical Map

After filtering unqualified markers, 52,503 polymorphic SNPs were employed for construction of physical map and GWAS analysis (**Supplementary Table 3** and **Supplementary Figure 2**). Among the polymorphic SNP markers, 34.9, 35.6, and 29.6% were from the A, B, and D genomes, respectively, indicating that the D genome has the lowest polymorphism (**Supplementary Table 3**). Of these, the chromosome 6B had more SNPs at 2691, whereas the 4D possessed only 1118 SNPs. The total length of the physical map is 14,058.8 Mb with an average marker density of 0.273 Mb per marker.

Population Structure and Linkage Disequilibrium

Population analysis indicates that all 251 wheat accessions were divided into three subgroups: subgroups I, II, and III. Of these, subgroup I contained 126 varieties mainly from Heilongjiang province ranging from the 1950s to 1980s, such as Kehong, Kehan 8, Kejian, and Xinshuguang 5; subgroup II had 75 varieties mainly from Heilongjiang province ranging from the 1990s to 2010s, such as Kechun 2, Kechun 9, Longmai 2, and Longmai 8; and subgroup III comprised 50 varieties mainly from Jilin province of China and foreign counties (United States, Canada, and Japan) (**Figure 1**). Furthermore, the average LD for the whole genome was 8 Mb (**Supplementary Figure 3**).

[4]http://www.ncbi.nlm.nih.gov/

TABLE 1 | ANOVA analysis for GY-related traits in 251 spring wheat accessions.

		F-value					
Source of variation	**df**	**SNU**	**SN**	**KNS**	**SL**	**TKW**	**GY**
Genotypes	250	120.4**	24.4**	94.4**	29.2**	85.8**	108.5**
Environments	3	380.9**	98.5**	302.1**	115.6**	258.5**	325.4**
Replicates (nested in environments)	2	18.2**	5.3**	13.2**	6.8**	13.6**	21.2**
Genotypes*Environments	749	9.2**	4.3**	5.6**	4.8**	6.3**	12.2**
Error	1425						

** and ** indicate significant at 0.05 and 0.01 level.*

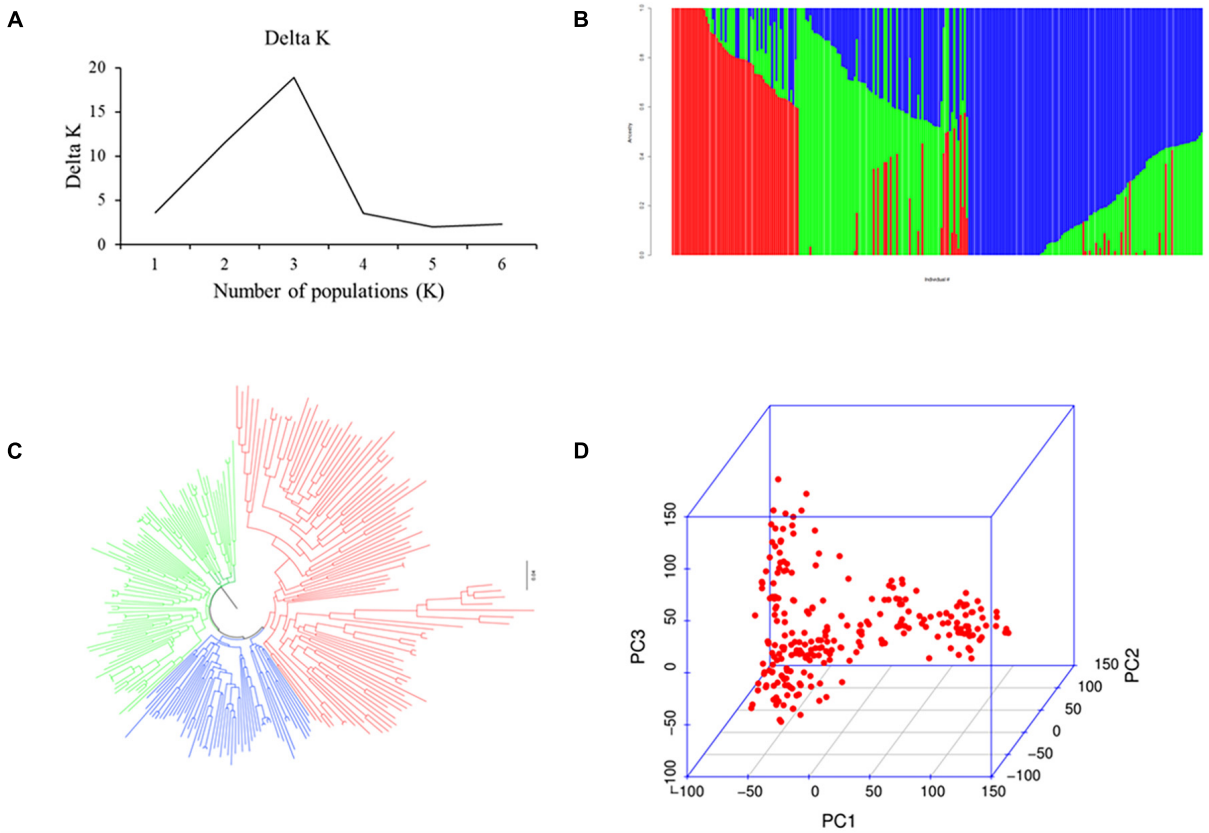


FIGURE 1 | Population analysis for the 251 spring wheat accessions. **(A)** Delta K for structure analysis. **(B)** Population structure analysis. **(C)** Neighbor-joining (NJ) tree. **(D)** Principal components analysis (PCA) plots.

GWAS

In total, 38 loci (162 MTAs) were identified associated with GY-related traits in this study (**Table 2**, **Figure 2**, **Supplementary Table 4**, and **Supplementary Figure 4**). Among these, 6 (1D, 3D, 5A, 6A, 6D, and 7B), 7 (2D, 3A, 3B, 5A, and 6A), 21 (1A, 1D, 2D, 3A, 3B, 3D, 4A, 4B, 4D, 5A, 5B, 6A, 6B, and 6D), 7 (1B, 2B, 5A, 6B, and 7D), 5 (1A, 4A, 4B, 1A, 4A, 1A, and 4A), and 12 (2B, 3A, 3D, 5D, 6A, 6D, 3D, 4A, 1D, and 2A) loci were detected for SNU, SN, SL, KNS, TKW, and GY, respectively and each explaining 7.7–9.3% of the phenotypic variances. Of these loci, two on chromosome 3D (22.8 Mb) and 6D (482.3–482.6 Mb) showed pleiotropic effects on SNU and GY; a locus on chromosome 5A (517.5–529.3 Mb) controlled both the SL and SNU; four loci on chromosome 2D (23.2–37.9 Mb), 3A (10.9–21.9 and 595.6 Mb), and 5A (20.7 Mb) had pleiotropic effects on SL and SN. Besides this, five loci on chromosome 1A (28.6–30.7, 307.1–321.1, and 338.9–346.0 Mb), 4A (653.4–687.1 Mb), and 4B (15.8–17.3 Mb) had pleiotropic effects on SL and TKW; two loci on chromosome 1D (252.4–284.1 Mb) and 3D (600.0–600.4 Mb) had significant effects on GY and KNS; and a locus on chromosome 2B (592.8 Mb) controlled both the KNS and GY. Another locus on chromosome 6A (2.5–4.0 Mb) had pleiotropic effects on GY, SN, and SL, whereas the loci of 6B (132.4–133.0 Mb), showed pleiotropic effects on SL and KNS.

Of the loci identified in this study, 28 were identified in the BLUP. Besides this, five loci (249.7–250.3 Mb on 1D, 10.7–21.7 Mb on 3A, 640.8–652.0 Mb on 4A, 516.7–528.5 Mb on 5A, and 460.8–464.8 Mb on 6D) existed in two or more environments and could not be detected in the BLUP. Besides this, four loci (30.2 Mb on 2A, 544.2–544.9 Mb on 5D, 615.2 Mb on 6A, 291.1 Mb on 6D, and 580.8–581.0 Mb on 7B) with strong environmental specificity and only identified in specific environment. Also, we observed a new QTL at 536.4–536.8 Mb on chromosome 6A, which is only identified in BLUP and not observed by analyzing individual environments.

Candidate Genes

In this study, six candidate genes for GY traits were identified (**Table 3**). Two cytokinin ribosides (*TraesCS2B02G397600* and *TraesCS3B02G281000*) were identified in the LD decay of the loci on 2B (592.8 Mb) and 3B (480.8–481.4 Mb). Another gene encoding an E3 ubiquitin transferase (*TraesCS3A02G344600*) was identified in the LD decay of the loci on chromosome 3A (596.0 Mb). For the loci on chromosome 3A (696.4–700.9 Mb) and 6A (2.5–4.0 Mb), candidate gene (*TraesCS3A02G459800*) for F-box proteins and serine/threonine-protein kinases (*TraesCS7B02G328800*) were identified, respectively. Besides

TABLE 2 | Loci for GY-related traits in 251 spring wheat accessions by association analysis.

Peak Marker[a]	Chromosome	Position (bp)[b]	*P*-value	R^2 (%)	Interval (Mb)	Trait[c]	Favorable allele	References
AX-110180733	1A	28676837	4.5E-07	13.5	28.6–30.7	TKW/SL	C	Wang et al. (2011) and Li et al. (2019)
AX-108817901	1A	309183870	1.4E-05	9.3	307.1–321.1	TKW/SL	C	
AX-109971512	1A	338900660	5.2E-05	8.5	338.9–346.0	SL	C	
AX-108850659	1A	370774742	1.4E-05	9.3	370.7–370.8	SL	C	
AX-110363533	1A	587798859	1.8E-05	9.2	587.8–594.7	SN	G	
AX-110953049	1B	103157084	1.7E-05	9.1	100.8–103.2	YM/NS	G	
AX-86175573	1D	284057669	6.1E-05	8.2	252.4–284.1	YM/NS	C	
AX-94767476	1D	430118481	6.3E-06	10.1	430.1–430.3	SL	C	
AX-111041231	2A	34739520	2.3E-05	9.0	34.7	YM	A	Li et al. (2019)
AX-111470278	2B	592759916	3.8E-06	10.4	592.8	SN/YM	C	
AX-86163393	2D	37927966	1.9E-05	9.0	23.9–37.9	SL/NSS	C	
AX-110090611	2D	556224959	1.6E-06	11.5	556.2–556.3	SL	C	
AX-111707600	3A	10841219	3.1E-05	11.7	10.9–21.9	YM/NSS	G	Li et al. (2019)
AX-110922897	3A	596035891	3.4E-06	10.5	596.0	NSS/SL	G	
AX-109490522	3A	696351940	8.0E-06	9.8	696.4–700.9	SL	A	Azadi et al. (2015) and Li et al. (2019)
AX-109379472	3B	481361479	3.6E-05	8.8	480.4–481.4	NSS	G	
AX-108791993	3B	589834517	3.0E-05	8.8	589.9	SL	A	
AX-94879852	3D	22763685	6.4E-05	8.2	22.8	NS/YM	A	
AX-109181055	3D	600142463	1.6E-06	11.4	600.0–600.4	NS/YM	A	Li et al. (2019)
AX-111662342	4A	640213293	4.6E-05	8.0	640.2	YM	A	
AX-110577792	4A	674369223	2.3E-05	8.8	653.4–687.1	TKW/SL	A	
AX-109901470	4B	15795054	2.8E-05	8.7	15.8–17.3	SL/TKW	G	
AX-109294476	4D	13447428	3.1E-05	8.9	13.5	SL	C	
AX-110185031	5A	20775669	4.2E-06	10.3	20.7	SL/NSS	G	Li et al. (2019)
AX-108851118	5A	519200575	2.7E-05	8.8	517.5–529.3	SL/NS	G	Li et al. (2018)
AX-109600316	5A	705150342	3.4E-06	10.4	705.2	SN	G	Gao et al. (2015) and Li et al. (2019)
AX-110606057	5B	697524601	7.2E-06	9.9	677.9–697.5	SL	C	Li et al. (2019)
AX-110510530	5D	548322186	9.4E-05	7.6	547.6–548.3	YM	C	
AX-94478216	6A	3983883	1.7E-06	11.2	2.5–4.0	YM/NSS/SL	T	
AX-110172686	6A	538918974	3.9E-05	8.9	538.9–539.3	NS	A	
AX-109862690	6A	619409256	2.9E-07	16.7	619.4	YM	A	
AX-111569068	6B	132435739	3.3E-05	8.5	132.4–133.0	SL/SN	A	
AX-94621559	6D	8684658	6.0E-05	8.1	8.6	SL/SN	T	
AX-111694627	6D	312832201	1.4E-05	9.3	312.8	YM	T	
AX-108976043	6D	486310967	7.8E-05	7.8	482.3–486.3	NS/YM	T	Li et al. (2019)
AX-109321162	7B	585610186	5.1E-05	8.1	585.4–586.4	NS	G	
AX-111622533	7D	21913685	5.8E-05	8.1	21.9	SN	G	
AX-109984815	7D	142702326	3.5E-05	8.6	138.9–148.9	SN	G	

[a] *The peak marker of the loci.*
[b] *According to the IWGSC v2.1.*
[c] *SNU, spike number per unit area; SN, spikelet number; SL, spike length; KNS, kernel number per spike; TKW, thousand-kernel weight; GY, grain yield.*

this, the gene *TraesCS6A02G301800* encoding trehalose 6-phosphatase (T6P) was identified as the candidate gene for the loci 7B (586.4 Mb).

DISCUSSION

The 251 spring wheat accessions were divided into three subgroups (**Figure 1**), and the characterization of the subgroups was largely consistent with geographic origins, released years, and pedigrees. For example, most of the accessions from Heilongjiang province ranging from the 1950s to 1980s belonged to subgroup 1, the accessions from Heilongjiang province ranging from the 1990s to 2010s belonged to subgroup 2, and subgroup 3 mainly included the accessions from the Jilin province and foreign counties. A significant population structure existed in the diverse panel, and previous studies indicate that the lack of appropriate correction for population structure can lead to spurious MTAs

(Zhu et al., 2008). Thus, to eliminate spurious MTAs, an MLM model with subpopulation data (Q) (fixed-effect factors) and kinship matrix (random-effect factor) were conducted. Also, the LD decay influenced several factors, such as population structure, allele frequency, recombination rate, and selection, and seriously affects the precision of association mapping. In this panel, the LD decay for the whole genome was about 8 Mb, consistent with previous reports (Liu et al., 2017), and indicates that the number of markers is enough for the subsequent association analysis.

Comparison With the QTL or Gene in Previous Studies

The genes or loci associated with GY-related traits were extensively reported previously. In this study, association of GY and related traits were performed based on the wheat high-density physical map. GY is a typical quantitative inheritance complex trait and significantly influenced by various environments (Gao et al., 2015; Li et al., 2018). The GY-related loci (gene) is distributed on all 21 chromosomes in wheat (Kumar et al., 2007; Reif et al., 2011; Lee et al., 2014; Lopes et al., 2015; Sukumaran et al., 2015; Gao et al., 2017; Li et al., 2018). Azadi et al. (2015) report a QTL for GY-related traits on chromosome 1A, which is tightly linked with the SSR marker *gwm357* and located between the two GY QTL mapped by Cuthbert et al. (2008) and Huang et al. (2004). Also, a locus for KNS at chromosome 1A (around 26.8–40.5 Mb) is identified (Wang et al., 2011; Li et al., 2019). According to the IWGSC V2.0 and the consensus map by Maccaferri et al. (2015), the loci in 1A identified by Azadi et al. (2015) and Li et al. (2019) are overlapped with the 1A locus (28.6–30.7 Mb) for TKW and SL identified in the present study. Thus, these loci in 1A (*AX-108817901* 307.1–321.1 Mb) for TKW or SL (*AX-109971512* 338.9–346.0 Mb) for SL (*AX-108850659* 370.7–370.8 Mb) for SL and (*AX-110363533* 587.8–594.7 Mb) for SN were novel. Besides this, the loci for YM and NS identified in the 1B chromosome (*AX-110953049* 100.8–103.2 Mb) and two loci in 1D chromosome for YM and NS (*AX-110953049* 252.4–284.1 Mb) and SL (*AX-94767476* 430.1–430.3 Mb) may be novel.

Li et al. (2019) identified a locus for GY at chromosome 2A (33.3–34.9 Mb) and 3A (21.2–26.9 and 702.6–712.2 Mb) by association mapping in 166 wheat accessions. Also, we identified few loci for GY-related traits in chromosome 2A and 3A. Of these loci identified by Li et al. (2019), the locus at 2A (32.3–33.9 Mb) was overlapped with the QTL for GY at 2A (34.7 Mb) in this study; the locus at 3A (22.9–39.3 Mb) were nearly with the QTL for GY at 3A (10.9–21.9 Mb), whereas the locus at 3A (700.1–705.1 Mb) coincides with the locus for KNS on 3A (696.4–700.9 Mb) in our study. Besides this, Azadi et al. (2015) and Gao et al. (2015) identified a locus at the 3A chromosome nearly the 3A locus (696.4–700.9 Mb) identified in this study. On chromosome 3D, Li et al. (2019) identified a locus for KNS by GWAS in 166 common wheat accessions (574.8–576.8 Mb) and explained 7.1–9.9% of the phenotypic variations, which overlapped with the loci for NS and SL (3D, 600.0–600.4 Mb) identified in our study. The loci on chromosome 2B (*AX-111470278* 592.8 Mb) for SN and YM, 2D (*AX-86163393* 23.9–37.9 Mb) for SL and NSS, and *AX-110090611* (556.2–556.3 Mb) for SL are likely to be new loci for GY-related traits. Besides this, the loci identified in chromosome 3A (*AX-110922897* 3A 596.0 Mb) for NSS and SL, 3B (*AX-109379472* 480.4–481.4 Mb) for NSS, *AX-108791993* (589.9 Mb) for SL and 3D (*AX-94879852* 22.8 Mb) for NS and YM appear to be new. Also, the loci identified in chromosome 4A (*AX-111662342* 4A 640.2 Mb) for YM, *AX-110577792* (653.4–687.1 Mb) for TKW and SL, 4B (*AX-109901470* 15.8–17.3 Mb) for SL and TKW, and *4D* (*AX-109294476* 13.5 Mb) are new loci for GY-related traits.

The 5A chromosome enriched a series of loci for GY-related traits. Reif et al. (2011) identified a GY QTL on 5A linked with SSR marker *barc151*. Liu et al. (2014) also mapped a SL QTL on 5A. Li et al. (2018) identified a QTL *QSL.caas-5AL.2* for SL in the Gaocheng 8901 × Zhou 8425B population. Besides this, Li et al. (2019) identified a locus for TKW by GWAS in 166 common wheat accessions at 5A (709.4–711.3) and explained phenotypic variations 8.2–13.0%. The locus for SL identified in this study at chromosome 5A (705.2 Mb) is overlapped with the loci discussed above. Azadi et al. (2015) detected a QTL for KNS on 5B, linked with DArT markers *wPt-3661*, whereas the 5B QTL is around the loci for SL (677.9–697.5 Mb) identified in the present study, and these might be the same. Also, the SL locus (677.9–697.5 *Mb*) on 5B is nearly the loci for KL, KW, and TKW detected by Chen et al. (2016); Mohler et al. (2016), and Sun et al. (2017), respectively, indicating that this should be an important locus in determining kernel weight. Li et al. (2019) identified a locus for TKW and SN by GWAS in 166 wheat accessions at 5B (696.5–699.7 Mb) and explained 8.2–13.0% of the phenotypic variations, which are nearly the loci at chromosome 5B identified in our study. Also, Li et al. (2019) detected a locus for TKW and SN by GWAS at 6D (482.3–486.3 Mb). We also have identified a locus for NS and GY on chromosome 6D (482.3–486.3 Mb). Thus, the loci on chromosome 5D (*AX-110510530* 547.6–548.3 Mb) for YM and 6A (*AX-94478216* 2.5–4.0 Mb) for YM, NSS, and SL and the loci (*AX-110172686* 538.9–539.3 Mb) for NS and (*AX-109862690* 619.4 Mb) for YM are likely to be new. The loci for GY-related traits identified in chromosome 6B (*AX-111569068* 132.4–133.0 Mb) for SL and SN, 6D (*AX-94621559* 8.6 Mb) for SL and SN, and *AX-111694627* (312.8 Mb) for YM were novel. Besides this, we identified 3 loci on the 7B and 7D chromosomes for GY-related traits. The loci on chromosome 7B (*AX-109321162* 585.4–586.4 Mb) for NS and 7D (*AX-111622533* 21.9 Mb; *AX-109984815* 138.9–148.9 Mb) for SN appear to be new.

Among the 38 loci for GY and related traits, 10 loci discussed above (1A: 28.6–30.7 Mb, 2A: 30.2 34.7 Mb, 3A: 10.9–21.9 Mb, 3A: 696.4–700.9 Mb, 3D: 600.0–600.4 Mb, 5A: 10.3–20.7 Mb, 5A: 517.5–529.3 Mb, 5A: 702.6–705.2 Mb, 5B: 677.9–697.5 Mb, and 6D: 482.3–486.3 Mb) should be the same as the QTL reported in previous studies. The stable loci validated by both GWAS and QTL mapping between ours and previous studies indicate that they are widespread in varieties and may be more powerful and stable in various varieties. Moreover, the methods of GWAS used in the present study are proven to be reliable and efficient in detecting loci for GY and related traits.

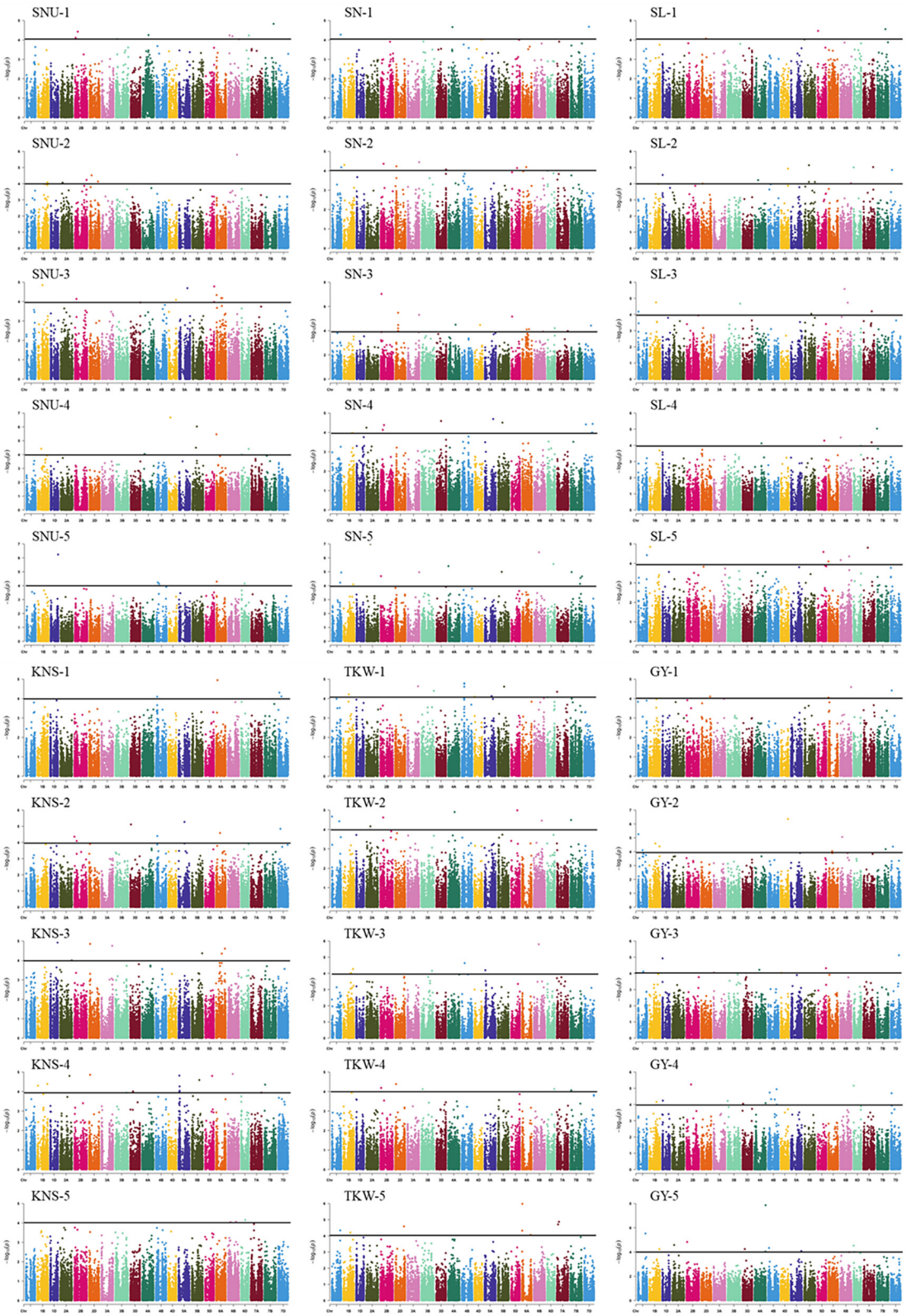


FIGURE 2 | Manhattan plot for GY-related traits in 251 wheat accessions analyzed by the MLM in Tassel v5.0. SNU, spike number per unit area; SN, spikelet number; SL, spike length; KNS, kernel number per spike; TKW, thousand-kernel weight; GY, grain yield. The 1, 2, 3, 4, and 5 indicated the Haerbin, 2018, 2019; Keshan, 2018, 2019, and the best linear unbiased prediction (BLUP).

Different QTL should be used in different regions. According to the results from association analysis, most of the loci (28) identified in this study existed on two or more environments (including BLUP), indicating high stability, and could be applied in various regions; these loci are only identified in an individual environment (30.2 Mb on 2A, 544.2–544.9 Mb on 5D, 615.2 Mb on 6A, 291.1 Mb on 6D, and 580.8–581.0 Mb on 7B) could be used in a specific environment, whereas those loci (536.4–536.8 Mb on chromosome 6A) only detected in BLUP may provide new insights into the genetic mechanism of GY-related traits.

Candidate Gene Analysis

To identify the candidate genes for GY-related traits, the flanking sequences of SNP markers (including the markers from the LD decay interval) significantly associated with GY-related traits were used as queries to BLAST against the National Center for Biotechnology Information (NCBI) and European Nucleotide Archive (ENA) databases. In total, six candidate genes were identified (**Table 3**) for further research. Of these, two cytokinin ribosides (*TraesCS2B02G397600* and *TraesCS3B02G281000*) were identified in the LD decay of the loci on chromosome 2B (592.8 Mb) and 3B (480.4–481.4 Mb). The cytokinin is a positive regulator of shoot growth and negative regulator of root growth (Han et al., 2014) and has been strongly implicated in many aspects affecting GY-related traits, particularly the kernel number and size (Yamburenko et al., 2017). Besides this, cytokinin also plays crucial roles in the response to biotic and abiotic stressors (Cortleven et al., 2019). *TraesCS3A02G344600,* coded an E3 ubiquitin transferase, which were identified for the loci on chromosome 3A (596.0 Mb). Previous studies report that the E3-Ubiquitin protein ligases are a large protein family that is important in plant growth and development (Craig et al., 2009; Park et al., 2010). Wang et al. (2020) indicates that a E3 ligase gene, *TaSDIR1-4A*, contributes to the determination of grain size in common wheat. For the loci on chromosome 3A (696.4–700.9 Mb) and 6A (2.5–4.0 Mb), candidate genes for F-box proteins (*TraesCS3A02G459800*) and serine/threonine-protein kinases (*TraesCS7B02G328800*) were identified, respectively. F-box proteins are a large protein family and plays crucial roles in cell-cycle progression, transcriptional regulation, flower formation, signal transduction, and many other cellular processes in plants (Hong et al., 2012; Kaur et al., 2017; Marcotuli et al., 2018; Guérin et al., 2021). Furthermore, *AX-108951749* on 2B and *IWA2223* on 5AL encodes the serine/threonine-protein kinases. The serine/threonine-protein kinases play crucial roles in cell-cycle progression, transcriptional regulation, flower formation, and signal transduction (Rehman et al., 2019; Jia et al., 2020). The gene *TraesCS6A02G301800* encodes trehalose 6-phosphatase (T6P) on the loci 7B (585.6 Mb), which regulates carbon assimilation and sugar status in plants. In addition, previous studies report that T6P has also been demonstrated to play an essential role in plant development under abiotic stress (Diab et al., 2013; Gahlaut et al., 2019).

TABLE 3 | The details for the candidate genes of GY-related traits.

Chromosome	Region (Mb)[a]	Candidate gene[b]	Annotation[c]
2B	592.8	TraesCS2B02G397600	cytokinin riboside
3A	696.4–700.9	TraesCS3A02G344600	E3 ubiquitin transferase
3A	596.0	TraesCS3A02G459800	F-box proteins
3B	480.4–481.4	TraesCS3B02G281000	cytokinin riboside
6A	2.5–4.0	TraesCS6A02G301800	trehalose 6-phosphatase
7B	585.6	TraesCS7B02G328800	serine/threonine-protein kinases

[a,b,c]*According to IWGSC v2.1.*

Potential Implications in Wheat Breeding

The conventional breeding approach has led to improved GY on wheat, and breeding selection is time-consuming and not very efficient (He et al., 2010). Significant additive effects are reported between GY traits and a number of favorable alleles, indicating that pyramiding favorable alleles will improve GY traits (Li et al., 2018). The markers associated with GY detected should facilitate MAS. Loci with pleiotropic and consistent effects across each environment should be amenable to MAS. Besides this, the loci validated between ours and previous studies by QTL-mapping or GWAS indicated these loci are stable in various varieties and should be applied in further study. In this study, accessions with agronomic characters and more favorable alleles, such as Kefeng 6, Beimai 15, Kehan 18, Longfumai 6, Longfumai 196, Kechun 7, Longmai 23, Kechun 1, Longfumai 4, and Longmai 13, are recommended as parental lines for improvement of GY-related traits in wheat breeding.

CONCLUSION

In this study, association mapping for GY and related traits (SNU, SL, SN, KNS, and TKW) were conducted in a diverse panel, including 251 spring wheat varieties mainly from China. In total, 38 loci were identified, and each explained 6.5–16.7% of the phenotypic variations. Of these, 12 are overlapped with known genes or QTL, and 26 are likely to be new. Besides this, six candidate genes for GY-related traits involved in grain development, plant hormone signal transduction, and starch biosynthesis were identified. The stable loci and associated markers and varieties with favorable traits and alleles could be used in further wheat breeding. These new loci provide a new sight in genetic architecture of GY-related 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/s.

AUTHOR CONTRIBUTIONS

YL and LY designed the research. JT analyzed the physiology data. YL and WY drafted the manuscript. YL, YS, JC, and CT

performed the experiment. HZ and LY revised the manuscript. All authors have read, edited, and approved the current version of the manuscript.

FUNDING

This work was funded by the "Key Projects of the Ministry of Science and Technology: Mutation Breeding of Main Crops" (2016YFD0102101). The Heilongjiang Agricultural Innovation Project (HNK2019CX04-06).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.728472/full#supplementary-material

REFERENCES

Azadi, A., Mardi, M., Hervan, E. M., Mohammadi, S. A., Moradi, F., Tabatabaee, M. T., et al. (2015). QTL mapping of yield and yield components under normal and salt-stress conditions in bread wheat (*Triticum aestivum* L.). *Plant Mol. Biol. Rep*. 33, 102–120. doi: 10.1007/s11105-014-0726-0

Beyer, S., Daba, S., Tyagi, P., Bockelman, H., Brown-Guedira, G., and Mohammadi, M. (2019). Loci and candidate genes controlling root traits in wheat seedlings-a wheat root GWAS. *Funct. Integr. Genomics* 19, 91–107. doi: 10.1007/s10142-018-0630-z

Breseghello, F., and Sorrells, M. E. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172, 1165–1177. doi: 10.1534/genetics.105.044586

Chen, G. F., Zhang, H., Deng, Z. Y., Wu, R. G., Li, D. M., Wang, M. Y., et al. (2016). Genome-wide association study for kernel weight-related traits using SNPs in a Chinese winter wheat population. *Euphytica* 212, 173–185. doi: 10.1007/s10681-016-1750-y

Cortleven, A., Leuendorf, J. E., Frank, M., Pezzetta, D., Bolt, S., and Schmülling, T. (2019). Cytokinin action in response to abiotic and biotic stresses in plants. *Plant Cell Environ.* 42, 998–1018. doi: 10.1111/pce.13494

Craig, A., Ewan, R., Mesmar, J., Gudipati, V., and Sadanandom, A. (2009). E3 ubiquitin ligases and plant innate immunity. *J Exp. Bot.* 60, 1123–1132.

Cui, F., Li, J., Ding, A., Zhao, C., Wang, L., Wang, X., et al. (2011). Conditional QTL mapping for plant height with respect to the length of the spike and internode in two mapping populations of wheat. *Theor. Appl. Genet*. 122, 1517–1536.

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

Cuthbert, J. L., Somers, D. J., Brûlé-Babel, A. L., Brown, P. D., and Crow, G. H. (2008). Molecular mapping of quantitative trait loci for yield and yield components in spring wheat (*Triticum aestivum* L.). *Theor. Appl. Genet*. 117, 595–608. doi: 10.1007/s00122-008-0804-5

Diab, A. A., Nada, A. M., and Ashoub, A. (2013). Molecular cloning, expression, sequence analysis and in silico comparative mapping of trehalose 6-phosphate gene from Egyptian durum wheat. *Int. J. Biotechnol. Mol. Biol. Res.* 4, 9–23. doi: 10.5897/ijbmbr2012.0147

Ellis, M. H., Rebetzke, G. J., Azanza, F., Richards, R. A., and Spielmeyer, W. (2005). Molecular mapping of gibberellin-responsive dwarfing genes in bread wheat. *Theor. Appl. Genet*. 111, 423–430. doi: 10.1007/s00122-005-2008-6

Gahlaut, V., Jaiswal, V., Singh, S., Balyan, H. S., and Gupta, P. K. (2019). Multi-locus genome wide association mapping for yield and its contributing traits in hexaploid wheat under different water regimes. *Sci. Rep*. 9:19486.

Gao, F., Ma, D., Yin, G., Rasheed, A., Dong, Y., Xiao, Y., et al. (2017). Genetic progress in grain yield and physiological traits in Chinese wheat cultivars of Southern Yellow and Huai Valley since 1950. *Crop Sci*. 57, 760–773. doi: 10.2135/cropsci2016.05.0362

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

Guérin, C., Mouzeyar, S., and Roche, J. (2021). The landscape of the genomic distribution and the expression of the f-box genes unveil genome plasticity in hexaploid wheat during grain development and in response to heat and drought stress. *Int. J. Mol. Sci.* 22:3111. doi: 10.3390/ijms22063111

Han, Y., Zhang, C., Yang, H., and Jiao, Y. (2014). Cytokinin pathway mediates APETALA1 function in the establishment of determinate floral meristems in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A*. 111, 6840–6845. doi: 10.1073/pnas.1318532111

He, Z. H., Xia, X. C., and Bonjean, A. P. A. (2010). *Wheat Improvement in China*. Mexico: CIMMYT, 51–68.

Hong, M. J., Kim, D. Y., Kang, S. Y., Kim, D. S., Kim, J. B., and Seo, Y. W. (2012). Wheat F-box protein recruits proteins and regulates their abundance during wheat spike development. *Mol. Biol. Rep*. 39, 9681–9696. doi: 10.1007/s11033-012-1833-3

Huang, X. Q., Kempf, H., Ganal, M. W., and Röder, M. S. (2004). Advanced backcross QTL analysis in progenies derived from a cross between a German elite winter wheat variety and a synthetic wheat (*Triticum aestivum* L.). *Theor. Appl. Genet*. 109, 933–943. doi: 10.1007/s00122-004-1708-7

Jia, H., Li, M., Li, W., Liu, L., Jian, Y., Yang, Z., et al. (2020). A serine/threonine protein kinase encoding gene KERNEL NUMBER PER ROW6 regulates maize grain yield. *Nat. Commun.* 11:988.

Jin, H., Wen, W., Liu, J., Zhai, S., Zhang, Y., Yan, J., et al. (2016). Genome-wide QTL mapping for wheat processing quality parameters in a Gaocheng 8901/Zhoumai 16 recombinant inbred line population. *Front. Plant Sci*. 7:1032. doi: 10.3389/fpls.2016.01032

Kaur, S., Dhugga, K. S., Beech, R., and Singh, J. (2017). Genome-wide analysis of the cellulose synthase-like (Csl) gene family in bread wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 17:193. doi: 10.1186/s12870-017-1142-z

Kumar, N., Kulwal, P. L., Balyan, H. S., and Gupta, P. K. (2007). QTL mapping for yield and yield contributing traits in two mapping populations of bread wheat. *Mol. Breeding* 19, 163–177. doi: 10.1007/s11032-006-9056-8

Lee, H. S., Jung, J. U., Kang, C. S., Heo, H. Y., and Park, C. S. (2014). Mapping of QTL for yield and its related traits in a doubled haploid population of Korean wheat. *Plant. Biotechnol. Rep.* 8, 443–454. doi: 10.1007/s11816-014-0337-0

Li, C., Bai, G., Chao, S., Carver, B., and Wang, Z. (2016). Single nucleotide polymorphisms linked to quantitative trait loci for grain quality traits in wheat. *Crop J.* 4, 1–11.

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

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.

Liu, J., He, Z., Rasheed, A., Wen, W., Yan, J., Zhang, P., et al. (2017). Genome-wide association mapping of black point reaction in common wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 17, 1–12.

Liu, J., He, Z., Wu, L., Bai, B., Wen, W., Xie, C., et al. (2016). Genome-wide linkage mapping of QTL for black point reaction in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet*. 129, 2179–2190. doi: 10.1007/s00122-016-2766-3

Lopes, M. S., Dreisigacker, S., Peña, R. J., Sukumaran, S., and Reynolds, M. P. (2015). Genetic characterization of the wheat association mapping initiative

(WAMI) panel for dissection of complex traits in spring wheat. *Theor. Appl. Genet*. 128, 453–464. doi: 10.1007/s00122-014-2444-2

Maccaferri, M., Ricci, A., Salvi, S., Milner, S. G., Noli, E., Martelli, P. L., et al. (2015). A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. *Plant Biotechnol. J.* 13, 648–663. doi: 10.1111/pbi.12288

Marcotuli, I., Colasuonno, P., Blanco, A., and Gadaleta, A. (2018). Expression analysis of cellulose synthase-like genes in durum wheat. *Sci. Rep*. 8:15675.

Mohler, V., Albrecht, T., Castell, A., Diethelm, M., Schweizer, G., and Hartl, L. (2016). Considering causal genes in the genetic dissection of kernel traits in common wheat. *J. Appl. Genet*. 57, 467–476. doi: 10.1007/s13353-016-0349-2

Nadolska-Orczyk, A., Rajchel, I. K., Orczyk, W., and Gasparis, S. (2017). Major genes determining yield-related traits in wheat and barley. *Theor. Appl. Genet.* 130, 1081–1098. doi: 10.1007/s00122-017-2880-x

Park, Y. Y., Lee, S., Karbowski, M., Neutzner, A., Youle, R. J., and Cho, H. (2010). Loss of MARCH5 mitochondrial E3 ubiquitin ligase induces cellular senescence through dynamin-related protein 1 and mitofusin 1. *J. Cell Sci.* 123, 619–626.

Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association mapping in structured populations. *Am. J. Hum. Genet*. 67, 170–181. doi: 10.1086/302959

Qin, X., Zhang, F., Liu, C., Yu, H., Cao, B., Tian, S., et al. (2015). Wheat yield improvements in China: past trends and future directions. *Field Crop. Res.* 177, 117–124. doi: 10.1016/j.fcr.2015.03.013

Quan, X., Dong, L. J., Zhang, N., Xie, C., Li, H., Xia, X., et al. (2021). Genome-wide association study uncover the genetic architecture of salt tolerance-related traits in common wheat (*Triticum aestivum* L.). *Front. Genet.* 12:663941. doi: 10.3389/fgene.2021.663941

Rasheed, A., Wen, W., Gao, F., Zhai, S., Jin, H., Liu, J., et al. (2016). Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theor. Appl. Genet.* 129, 1843–1860. doi: 10.1007/s00122-016-2743-x

Rehman, S. U., Wang, J., Chang, X., Zhang, X., Mao, X., and Jing, R. (2019). A wheat protein kinase gene *TaSnRK2.9-5A* associated with yield contributing traits. *Theor. Appl. Genet* 132, 907–919. doi: 10.1007/s00122-018-3247-7

Reif, J. C., Maurer, H. P., Korzun, V., Ebmeyer, E., Miedaner, T., and Würschum, T. (2011). Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. *Theor. Appl. Genet*. 123, 283–292. doi: 10.1007/s00122-011-1583-y

Sela, H., Ezrati, S., Ben-Yehuda, P., Manisterski, J., Akhunov, E., Dvorak, J., et al. (2014). Linkage disequilibrium and association analysis of stripe rust resistance in wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) population in Israel. *Theor. Appl. Genet*. 127, 2453–2463. doi: 10.1007/s00122-014-2389-5

Shi, W., Hao, C., Zhang, Y., Cheng, J., Zhang, Z., Liu, J., et al. (2017). A combined association mapping and linkage analysis of kernel number per spike in common wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 8:1412. doi: 10.3389/fpls.2017.01412

Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., and Reynolds, M. P. (2015). Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor. Appl. Genet.* 128, 353–363. doi: 10.1007/s00122-014-2435-3

Sun, C. W., Zhang, F. Y., Yan, X. F., Zhang, X. F., Dong, Z. D., Cui, D. Q., et al. (2017). Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai Valley of China. *Plant Biotechnol. J.* 15, 953–969. doi: 10.1111/pbi.12690

Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822.

Valluru, R., Reynolds, M. P., Davies, W. J., and Sukumaran, S. (2017). Phenotypic and genome-wide association analysis of spike ethylene in diverse wheat genotypes under heat stress. *New Phytol.* 214, 271–283.

Wang, J., Wang, R., Mao, X., Zhang, J., Liu, Y., Xie, Q., et al. (2020). RING finger ubiquitin E3 ligase gene *TaSDIR1-4A* contributes to determination of grain size in common wheat. *J. Exp. Bot.* 71, 5377–5388.

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

Wang, S. C., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796.

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

Würschum, T., Langer, S. M., Longin, C. F. H., Tucker, M. R., and Leiser, W. L. (2017). A modern Green Revolution gene for reduced height in wheat. *Plant J.* 92, 892–903. doi: 10.1111/tpj.13726

Yamburenko, M. V., Kieber, J. J., and Schaller, G. E. (2017). Dynamic patterns of expression for genes regulating cytokinin metabolism and signaling during rice inflorescence development. *PLoS One* 12:e0176060. doi: 10.1371/journal.pone.0176060

Zhou, K. Q., Wang, G. D., Li, Y. H., Liu, X. B., Herbert, S. J., and Hashemi, M. (2014). Assessing variety mixture of continuous spring wheat (*Triticum aestivum* L.) on grain yield and flour quality in Northeast China. *Int. J. Plant Prod*. 8, 91–105.

Zhou, Y., He, Z. H., Sui, X. X., Xia, X. C., Zhang, X. K., and Zhang, G. S. (2007). Genetic improvement of grain yield and associated traits in the northern China winter wheat region from 1960 to 2000. *Crop Sci*. 47, 245–253. doi: 10.2135/cropsci2006.03.0175

Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1, 5–20.

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 19 November 2021
doi: 10.3389/fgene.2021.767883

Discovery and Validation of a Recessively Inherited Major-Effect QTL Conferring Resistance to Maize Lethal Necrosis (MLN) Disease

Ann Murithi[1,2], Michael S. Olsen[1], Daniel B. Kwemoi[3], Ogugo Veronica[1], Berhanu Tadesse Ertiro[1], Suresh L. M.[1], Yoseph Beyene[1], Biswanath Das[1], Boddupalli M. Prasanna[1] and Manje Gowda[1]**

[1]International Maize and Wheat Improvement Center (CIMMYT), Nairobi, Kenya, [2]Department of Plant Science and Crop Protection, University of Nairobi, Nairobi, Kenya, [3]National Crops Resources Research Institute (NaCRRI), Namulonge, Uganda

OPEN ACCESS

Edited by:
Ahmed Sallam, Assiut University, Egypt

Reviewed by:
Tao Zhong, China Agricultural University, China
Jianfeng Weng, Chinese Academy of Agricultural Sciences (CAAS), China

***Correspondence:**
Michael S. Olsen
M.Olsen@cgiar.org
Manje Gowda
M.Gowda@cgiar.org

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *31 August 2021*
Accepted: *26 October 2021*
Published: *19 November 2021*

Citation:
Murithi A, Olsen MS, Kwemoi DB, Veronica O, Ertiro BT, L. M. S, Beyene Y, Das B, Prasanna BM and Gowda M (2021) Discovery and Validation of a Recessively Inherited Major-Effect QTL Conferring Resistance to Maize Lethal Necrosis (MLN) Disease. Front. Genet. 12:767883. doi: 10.3389/fgene.2021.767883

Maize lethal necrosis (MLN) is a viral disease with a devastating effect on maize production. Developing and deploying improved varieties with resistance to the disease is important to effectively control MLN; however, little is known about the causal genes and molecular mechanism(s) underlying MLN resistance. Screening thousands of maize inbred lines revealed KS23-5 and KS23-6 as two of the most promising donors of MLN resistance alleles. KS23-5 and KS23-6 lines were earlier developed at the University of Hawaii, United States, on the basis of a source population constituted using germplasm from Kasetsart University, Thailand. Both linkage mapping and association mapping approaches were used to discover and validate genomic regions associated with MLN resistance. Selective genotyping of resistant and susceptible individuals within large F_2 populations coupled with genome-wide association study identified a major-effect QTL (*qMLN06_157*) on chromosome 6 for MLN disease severity score and area under the disease progress curve values in all three F_2 populations involving one of the KS23 lines as a parent. The major-effect QTL (*qMLN06_157*) is recessively inherited and explained 55%–70% of the phenotypic variation with an approximately 6 Mb confidence interval. Linkage mapping in three F_3 populations and three F_2 populations involving KS23-5 or KS23-6 as one of the parents confirmed the presence of this major-effect QTL on chromosome 6, demonstrating the efficacy of the KS23 allele at *qMLN06.157* in varying populations. This QTL could not be identified in population that was not derived using KS23 lines. Validation of this QTL in six F_2 populations with 20 SNPs closely linked with *qMLN06.157* was further confirmed its consistent expression across populations and its recessive nature of inheritance. On the basis of the consistent and effective resistance afforded by the KS23 allele at *qMLN06.157*, the QTL can be used in both marker-assisted forward breeding and marker-assisted backcrossing schemes to improve MLN resistance of breeding populations and key lines for eastern Africa.

Keywords: maize lethal necrosis, genome-wide association study, F2 population, selective genotyping, disease resistance

INTRODUCTION

Increasing trade and travel coupled with weak phytosanitary systems are accelerating the global spread of devastating crop pests and diseases (McDonald and Stukenbrock, 2016; De Groote et al., 2021). The maize lethal necrosis (MLN) is one such transboundary maize disease that emerged in eastern Africa in 2011 (Mahuku et al., 2015; Prasanna et al., 2020). The disease was first reported in Kenya in 2011 and later spread rapidly to neighboring countries in eastern Africa. MLN results from the synergistic interaction of *Maize chlorotic mottle virus* (MCMV) with any of the cereal viruses of the Potyviridae family including *Sugarcane mosaic virus* (SCMV), *Wheat streak mosaic virus (WSMV)*, and *Maize dwarf mosaic virus* (MDMV) (Mahuku et al., 2015; Prasanna et al., 2020). MLN in eastern Africa was found to be due to synergistic interaction between MCMV and SCMV (Wangai et al., 2012). Maize is affected by MLN at all growth stages, resulting in chlorotic mottling of leaves, severe stunting, and necrosis (Wangai et al., 2012). An MLN survey in Kenya in 2012–2013 reported 26,000 ha of maize succumbed to MLN and about 95% of commercially available maize varieties were susceptible (Semagn et al., 2015; De Groote, et al., 2016). Subsequently, a community survey in 2013 done by De Groote et al. (2016) to assess the impact of MLN in Kenya reported a loss of 0.5 million tons of maize production, valued at US$180 million. The disease was exacerbated because of practices such as monocropping, besides lack of resistant maize varieties, and complicated nature of MLN spread and development (Beyene et al., 2017). Management and control of MLN can be achieved through effective integration of host plant resistance, vector control, and cultural practices (Prasanna et al., 2020). A review by Marenya et al. (2018) showed that, although improving agronomic practices through maize-legume rotation would be a useful approach for MLN control, application of such a method may not be feasible over large geographic areas in eastern Africa. Marenya et al. (2018) highlighted the importance of using MLN tolerant maize varieties and estimated the benefits to the tune of US$245–756 million and US$195–675 million in Ethiopia and Kenya, respectively.

Intensive evaluation of elite global maize germplasm (CIMMYT elite breeding panels, USDA virus-resistant line collection, and maize landrace accessions from CIMMYT gene bank) during 2013–2015 revealed very few sources of resistance to MLN and MCMV, with most of the germplasm showing moderate to high susceptibility (Semagn et al., 2015; Nyaga et al., 2020; Prasanna et al., 2020). Therefore, the identification of genetically diverse resistance sources and introgression of resistance into elite maize lines is considered as a high priority for maize breeding program in Africa. Understanding the nature of resistance and the genomic regions associated with MLN resistance is important for accelerating the transfer of resistance into diverse genetic backgrounds (Gowda et al., 2015, 2018). Linkage mapping and genome-wide association study (GWAS) are two major mapping strategies used for dissection of genetic architecture of traits (Holland, 2007). Earlier efforts through GWAS led to the mapping and localization of genomic regions associated with MLN resistance; these genomic regions with major and minor effects were spread across several chromosomes (Gowda et al., 2015; Nyaga et al., 2020).

Selective genotyping (SG) (Saunak et al., 2009), a strategy to genotype only individuals with extreme phenotypic values, is a useful approach for managing genotyping costs of a QTL discovery project. In simulation studies, Sun et al. (2010) suggested that detection power of large-effect QTL (>10% of trait variation) is high even when genotyping tails as small as 5% of the population. SG was initially proposed to increase the power of detecting rare alleles with large effects. This approach has been applied to identify associations between quantitative traits and genetic markers in human genetic studies including attention-deficit hyperactivity disorder (Cornish et al., 2005) and body mass index (Herbert et al., 2006), and in plants for traits such as sugarcane orange rust (McCord et al., 2019). Using a SG strategy, Uemoto et al. (2011) identified 32 loci associated with oleic acid (C18:1) in the intramuscular fat of the trapezius muscles in Japanese black cattle. Barnett et al. (2013) extended the extreme case control methods to identify rare variants in sequence-based association studies.

Results from the evaluation of thousands of maize inbred lines against MLN under artificial inoculation at the CIMMYT-managed MLN Screening Facility at KALRO-Naivasha, Kenya, identified extremely few lines with MLN resistance (Prasanna et al., 2020). Two of the most promising sources of MLN resistance identified were KS23-5 and KS23-6, originating from the University of Hawaii, United States, on the basis of source germplasm from the Kasetsart University, Thailand. The KS23 lines were identified as sources of MCMV resistance in Hawaii (Brewbaker, 2009) and added to a diverse collection of virus resistance lines at the Ohio State University, United States (Jones et al., 2018). The two KS23 lines developed only mild MLN and MCMV symptoms late into disease rating period under both field and screenhouse conditions at the MLN Screening Facility in Naivasha, Kenya (Mahuku et al., 2015).

The aim of this study was to identify genomic region(s) associated with MLN resistance within the KS23 genetic background and to validate these genomic regions in diverse bi-parental populations. The specific objectives were 1) to identify QTL associated with MLN resistance in five F_2 populations and three F_3 populations involving KS23 lines as donor parents and 2) to validate QTL in independent F_2 populations involving KS23 lines as resistant parents. To accomplish these objectives, two sets of populations were developed on the basis of KS23 lines, and the two QTL mapping strategies—linkage mapping and GWAS—were applied to map and characterize the genomic regions associated with MLN resistance. The identified KS23 alleles that confer MLN resistance were validated using $F_{2:3}$ populations.

MATERIALS AND METHODS

Germplasm and Phenotype Evaluation

Maize inbred lines KS23-5 and KS23-6 (Kaeppler et al., 1998; Brewbaker, 2009) were obtained from the North Central

Germplasm Introduction Station, United States, and maintained at CIMMYT. Five bi-parental F_2 populations were used for SG: SG1 (KS23-5 x CZL00025), SG2 (KS23-5 x CML545), SG3 (KS23-6 x CML539), SG4 (CML494 x CZL068), and SG5 (DTP-F46 x CML442). Populations 4 and 5 involved CIMMYT lines with moderate resistance to MLN (CML494 and DTP-F46). Inbred lines were crossed, and their F_1 self-pollinated to generate F_2 seed during 2014. In the 2015 main season, over 2,500 F_2 individuals across all populations were planted at the MLN Screening Facility at Naivasha, Kenya. Each individual F_2 plant was tagged, and tissue sampled prior to inoculation with MLN. Individual plants that showed any symptoms of stress or disease prior to inoculation were excluded from the trial to prevent confounding effects in identification of the susceptible tail of each population. Inoculated plants were rated at five successive time points and 10% of individuals within each population which succumbed earliest and 10% which exhibited strong resistance were selected for genotyping.

Three independent F_3 bi-parental QTL mapping (QM) populations involving KS23 lines were screened and evaluated under MLN artificial inoculation at MLN Screening Facility, Naivasha, Kenya. QM1 (CML543*2/CML444:DH5 x KS23-5) comprised 150 F_3 families, QM2 (CML543*2/CML444:DH6 X KS23-6) comprised 138 F_3 families, and QM3 (CML543 x KS523-5) comprised 108 F_3 families. The three populations were evaluated under MLN artificial inoculation in 3-m plots for two seasons in 2016 and 2017 in an alpha lattice experimental design, with two replications per entry. Standard agronomic management practices were followed.

QTL validation populations (VPs) were generated to evaluate the effect of the KS23 allele across various genetic backgrounds. Six bi-parental populations were developed in 2016: VP1 (KS23-6 x CML5458), VP2 (KS23-6 x CML539), VP3 (KS23-6 x CKDHL0186), VP4 (KS23-6 x CKDHL0221), VP5 (KS23-6 x CML 442), and VP6 (KS23-6 x CML537). F_2 plants from each population were genotyped using 20 SNP markers, within or flanking the MLN-resistant QTL region, spanning a 56-cm interval position 155–159 Mb on chromosome 6. SNP genotyping was performed at Corteva, Johnston, Iowa, United States; F_2 plants that had no recombination events across the target interval were self-pollinated, and F_3 ears were classified as homozygous for the KS23 allele, homozygous for the alternate allele, or heterozygous. At least 90 individuals from each population (30 ears selected for each marker class) were evaluated in 2017. The trial was evaluated in an alpha lattice incomplete block design with two replicates per entry at the MLN Screening Facility, Naivasha, for one season.

The detailed protocol for MLN artificial inoculation was explained in earlier publications (Gowda et al., 2015, 2018; Nyaga et al., 2020; Awata et al., 2020; Awata et al., 2021). In brief, artificial inoculation of the materials in all the field trials was done following the standard protocols developed by CIMMYT, as below. Inoculum for each virus type, MCMV and SCMV, were prepared separately. Preparation of the inoculant began with growing of susceptible maize plants in pots in two separate greenhouses, one for MCMV and the other for SCMV. Plants were infected with each of the viruses, previously isolated from infected plants, respectively. Identified symptomatic leaves from infected plants were harvested and a diagnostic assay, enzyme-linked immunosorbent assay was conducted in the MLN Screening Facility laboratory at Naivasha to ensure the purity of the viruses. Infected leaves were harvested, weighed (4.8 kg of MCMV and 1.2 kg of SCMV) and chopped separately. Leaves were homogenized in cold 0.1 M potassium phosphate buffer at pH 7.0 and sieved to remove plant debris. Extracted MCMV and SCMV homogenate was mixed in a large mixing tank at a ratio of 1:4 (MCMV: SCMV) and Celite (an abrasive) added at 1 g/L. Inoculation of the test materials was done at the V4-6 stage using a motorized backpack mist blower (Solo 423 Mist Blower, 12 L capacity), and a repeat of the inoculation was performed after 1 week. Inoculated plots were rated for MLN severity using a scale of 1–9, where 1 = no MLN symptoms, 3 = fine chlorotic streaks on new/emerging leaves, 5 = severe chlorotic mottling throughout plant, 7 = excessive chlorotic mottling and leaf necrosis or presence of "dead heart" symptoms, and 9 = complete plant necrosis. Disease severity (DS) was taken beginning 3 weeks after the second inoculation and at 7-day intervals thereafter. Five ratings were taken for the SG experiment, and four ratings were taken for the $F_{2:3}$ VPs. After analyzing DS ratings, we selected the third score (40 days after inoculation) for further analysis because of its higher heritability estimate and full expression of disease symptoms. Area under the disease progress curve (AUDPC) was calculated for each plot to provide a measure of the progression of MLN severity across time. AUDPC data were calculated by using the formula $\sum_{i=1}^{i=1}[(ti+1-ti)(yi+yi+1)/2]$, where "t" is time in days of each reading, "y" is the percentage of affected foliage at each reading, and "n" is the number of readings (Campbell and Madden, 1990; Madden, 2007).

Selective Genotyping Experiment

F_2 individuals with extreme MLN responses from each of the five SG populations were selected and genotyped using DArTSeq at Canberra, Australia. Approximately 27,000 SNPs were used to perform a whole genome scan to detect genomic variations that signal association with MLN resistance. Genotypic data from ~27,000 SNP markers were filtered to generate ~20,000 SNPs with < 20% missing genotypes and minor and maximum allele frequency of >1% filtering criteria was used on TASSEL ver. 5.0. Phenotypic values for MLN_DS and AUDPC were used for the analysis. DS scores of two populations without the KS23 background were also included in the GWAS analysis.

TASSEL ver. 5.0 was used to run GWAS analysis (Bradbury et al., 2007). To detect genomic variation underlying the observed responses in terms of MLN_DS and AUDPC values, a MLM analysis combining kinship and population structure as covariates was performed. The analysis was done by making two groups—the first three F_2 populations having KS23-5 or KS23-6 as one of the parents, and other group solely based on white maize germplasm (non-KS23) developed at CIMMYT. Detailed information of population structure was described using the first three PCs using the EIGSTRAT method described by Price et al. (2006). The pattern generated from the quantile-quantile plot of the model, and at which point the observed F test deviated from the expected F test statistics, was

used to determine the p value threshold ($p < 5 \times 10^{-7}$). All the significant SNPs were simultaneously fitted into a linear model to obtain R^2_{adj}, to determine how much the detected SNP contributed to the total phenotypic variance. The BP position of the significantly associated SNPs was used to perform BLAST searches against the maize B73 reference genome, RefGen_v2 (http://acdstagging.org/v2/genes.php).

Phenotypic Data Analysis

Each crop season in which MLN evaluation was undertaken was treated as an individual environment. Analysis of variance (ANOVA) was done using multiple environments traits analysis package incorporated in R software (META-R; Gregorio et al., 2015) that integrates both fixed and random factors. Variance components (σ^2_G, σ^2_{GxE}, σ^2_e) for individual and combined environments were estimated. Correlation between trait and environment and summary statistics (means, SE, range, LSD, and CV) were also generated using standard procedures implemented in META-R. Broad sense heritability (H^2) was calculated as the ratio of genotypic variance (δ^2G) to the phenotypic variance. Best linear unbiased predictions were also obtained by using location and the reps as fixed effects and genotype and incomplete block as random effects.

Linkage Mapping

The three F_3 bi-parental QM populations were genotyped with 447 SNP markers spaced across the genome using Kompetitive Allele Specific PCR (KASP) genotyping systems of LGC Company (Semagn et al., 2013). Five F_2 populations both tails were genotyped with Diversity Array Technology Pty Ltd (DArT). Genetic maps for the five F_2 populations were also made in same way as done for the F_3 populations. Markers with duplicate genotypes, monomorphic markers, and those with >10% missing genotypes were excluded from the analyses. QTL IciMapping was used to remove the highly correlated SNPs, resulting in retention of 360, 361, 360, 781, 770, 750, 780, and 781 high-quality SNPs in QM1, QM2, QM3, SG1, SG2, SG3, SG4, and SG5, respectively. These SNPs were used to construct linkage maps using the MAP function, by selecting the most significant markers using stepwise regression. Linkage maps for each population were constructed using IciMapping V 4.0 with Kosambi method for map distance calculation, logarithm of odds (LOD) score set at threshold of 3.0, and maximum distance of 30 cm between two loci to declare linkage between two markers. MLN_DS and AUDPC scores generated for each population were used to detect QTL on the basis of Inclusive Composite Mapping (ICIM) implemented in the software. The step of ICIM was set to 1 cm, and LOD threshold of 3.0 was used to declare putative QTL. Both additive and dominance effects of each QTL were estimated, and the favorable allele-contributing parent was identified on the basis of the sign of additive effect of each QTL. Phenotypic variation explained by individual QTL and total variation explained by the QTLs were estimated. QTL naming was done with letter "q" indicating QTL, followed by abbreviation of trait name, the chromosome number, and approximate marker position in Mb along the chromosome using the B73 v2 reference physical map.

RESULTS

Phenotypic Data Analysis

The response of both resistant and susceptible parents used in development of five F2 populations were clearly visible in the field (**Supplementary Table S1**). Among the SG populations, those involving KS23 sources skewed more toward susceptibility possibly due to recessive inheritance of the resistance in these F_2 populations, whereas the two non-KS23 populations showed normally distributed phenotypic data for both MLN_DS and AUDPC values (**Figure 1**). The distribution of MLN_DS within each SG population ranged from 2 to 9 with population mean values ranging from 4.8 in SG1 to 7.0 in SG3 (**Table 1**).

The three QM populations each showed normal distribution of phenotypic data for MLN_DS and AUDPC values (**Figure 1**). The MLN_DS values ranged from 2.5 to 7.24, from 2.2 to 7.4, and from 2.73 to 8.70 with the mean of 4.67, 4.82, and 4.76 in QM1, QM2, and QM3, respectively. AUDPC values ranged from 72 to 204, from 67 to 205, and from 78 to 242 with an average of 131, 132, and 133 in QM1, QM2, and QM3, respectively. Analyses of variance in each of the QM populations revealed significant ($p < 0.01$) genotypic variance and genotype × environment interaction variances for both MLN_DS and AUDPC values (**Table 2**). The heritability estimates were moderate to high with 0.66, 0.78, and 0.69 for MLN_DS and with 0.71, 0.80, and 0.73 for AUDPC values in QM1, QM2, and QM3, respectively.

QTL Mapping With Selective Genotyping in F_2 Populations

For each F_2 and F_3 populations, linkage map, the number of progenies or families, markers, map lengths, and average genetic distances between the markers for each biparental population are presented in **Supplementary Table S1**. Major-effect QTL was identified in all the five SG populations (**Table 3**; **Figure 2**). From SG1, two major-effect QTLs were identified on chromosome 6 with LOD scores of 60.6 and 11.8, respectively, explaining >55% of the phenotypic variance for MLN_DS. These QTL were also detected for AUDPC values. In SG2, among the three QTL detected for MLN_DS, the QTL on chromosome 6 explained up to 58% of the total variation and was also detected for AUDPC values. For SG3, three and eight QTL were detected, which together explained 86.8% and 90.7% of the total variation for MLN_DS and AUDPC values, respectively. The major QTL (*qMLN6-155*) where favorable alleles are contributed by resistant parent KS23-6 individually explained >55% of variation for both MLN_DS and AUDPC values. The other two SG populations that are lacking KS23 parents detected completely different QTL on chromosomes 3 and 6. A major QTL was detected consistently for both MLN_DS and AUDPC values on chromosome 3 (*qMLN3-142*), which explained >16% of total variation in SG4. In SG5, minor effect QTL were identified in different genomic regions compared to those of detected in the KS23-based populations. Comparison across populations revealed one common genomic region on chromosome 6, between 150 and 160 Mb, with favorable alleles contributed either by KS23-5 and KS23-6.

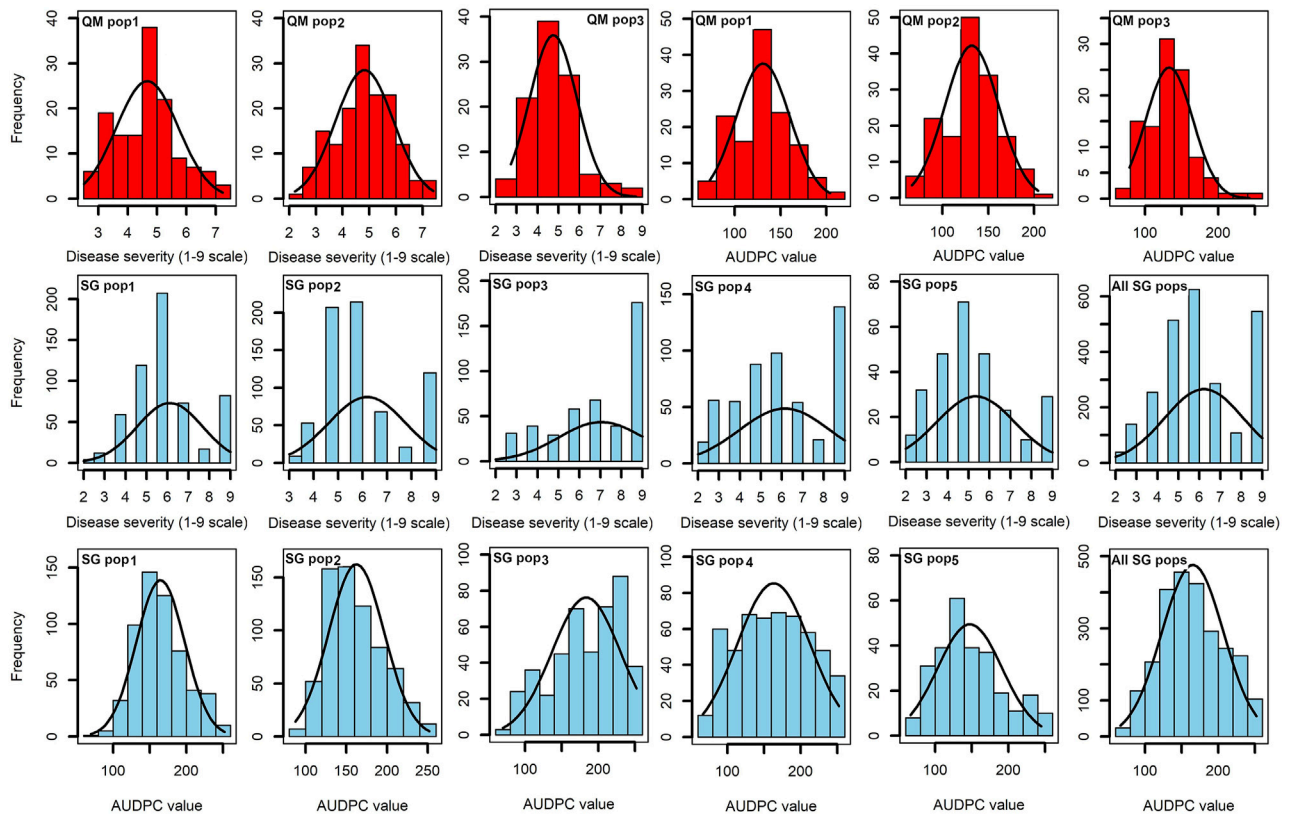


FIGURE 1 | Phenotypic distribution of MLN disease severity (on a scale of 1–9) and AUDPC values in three QM populations (in red) and five SG populations (in blue).

TABLE 1 | MLN disease severity and descriptive statistics of the SG populations under artificial inoculation at the MLN Screening Facility, Naivasha, Kenya.

Population	R Parent	S Parent	F_2 pop mean	N[a]	Ind[b]	R Tail (%)[c]
SG1–(KS23-5 x CZL0025)	2	8	4.79	273	31/29	11.3
SG2–(KS23-5 x CML545)	2	9	6.57	530	71/67	13.4
SG3–(KS23-6 x CZL3018)	2	8	7.01	445	43/64	9.7
SG4–(CZL0068 x CML494)	3	8	6.19	692	35/37	5.1
SG5–(CML442 x DTP-F46)	3	9	6.12	573	34/71	5.9

[a]*Number of F_2 plants from which the F_2 mean MLN_DS was derived. The R and S tails were drawn from this set of F_2 plants for genotyping.* [b]*Number of individuals in the R and S tails, respectively.* [c]*Percentage of the population represented by the R tail.*

TABLE 2 | Means, variance, and heritability estimates for MLN resistance in the three F_3 populations used for linkage mapping.

Trait	Mean	σ^2_G	σ^2_{GxE}	σ^2_e	H^2	$LSD_{0.05}$	CV
QM1–(CML543*2/CML444:DH5) x KS23-6							
MLN_DS	4.67	0.69[a]	0.32[a]	0.81	0.66	1.01	19.33
AUDPC	130.97	573.67[a]	211.95[a]	530.42	0.71	23.57	17.59
QM2–(CML543*2/CML444:DH6) x KS23-5							
MLN_DS	4.82	0.88[a]	0.11[a]	0.79	0.78	0.92	18.44
AUDPC	132.09	662.85[a]	87.57[a]	501.27	0.80	25.46	16.95
QM3–CML543/KS523-5							
MLN_DS	4.75	0.73[a]	0.34[a]	0.63	0.69	1.03	16.66
AUDPC	133.20	621.31[a]	261.48[a]	386.49	0.73	23.34	14.76

[a]*Significant at* $p < 0.01$.

For SG1, a major QTL (*qMLN6-157*) on chromosome 6 with a LOD score of 20.9 and explaining 60.6% of the phenotypic variation was significantly associated with marker *S6_155,436,477* (**Figure 3**; **Table 3**). In SG1, both MLN_DS and AUDPC scores for heterozygotes were like those of individuals with two alleles from susceptible parent CZL0025 (**Figure 3**). In SG2, a major-effect QTL *qMLN-155* that was flanked by two SNPs *S6_151,486,592* and *S6_155,632,957* detected on chromosome 6 was identified with LOD scores of about 50 that explained >60% of the phenotypic variation (**Figure 3**; **Table 3**). For this QTL, both MLN_DS and AUDPC scores for individuals with two alleles from CML545 were similar as of heterozygotes skewed toward susceptibility (**Figure 3**; **Table 3**). Similarly, also in SG3 for the major QTL on chromosome 6, we observed that the score of MLN_DS and AUDPC values for all heterozygotes was skewed

TABLE 3 | QTL associated with MLN disease severity, and AUDPC value in five selective genotyping (SG) F2 populations.

Trait	QTL name	Chr	Position (cM)	LOD	PVE (%)	Add	Dom	TPVE (%)	QTL confidence interval		Favorable allele
									Left M	Right M	
SG1–KS23-5 x CZL0025											
MLN_DS	*qMLN6-157*	6	638	20.93	60.62	2.95	0.05	77.87	S6_155,436,477	S6_161,415,596	KS23-5
	qMLN6-157	6	714	6.64	11.85	1.46	–0.77		S6_152,772,323	S6_160,037,589	KS23-5
AUDPC	*qMLN6-157*	6	638	25.33	56.80	70.75	7.69	80.93	S6_155,436,477	S6_161,415,596	KS23-5
	qMLN6-157	6	713	10.16	16.59	35.64	–94.85		S6_152,772,323	S6_160,037,589	KS23-5
SG2–KS23-5 x CML545											
MLN_DS	*qMLN4-175*	4	154	26.12	8.43	–0.14	–4.05	85.23	S4_13,413,473	S4_176,397,245	CML545
	qMLN4-175	4	213	21.79	8.36	0.10	–3.93		S4_133,105,107	S4_176,397,245	KS23-5
	qMLN6-155	6	37	56.27	58.79	2.65	3.13		S6_151,486,592	S6_155,632,957	KS23-5
AUDPC	*qMLN6-155*	6	38	64.89	74.82	60.71	75.26	88.12	S6_151,486,592	S6_155,632,957	KS23-5
SG3–KS23-6 x CZL3018											
MLN_DS	*qMLN2-9*	2	330	6.95	9.14	0.07	–2.45	86.81	S2_7,327,544	S2_9,505,516	KS23-6
	qMLN6-155	6	303	33.42	55.54	1.82	1.45		S6_155,436,477	S6_1,56,119,960	KS23-6
	qMLN6-159	6	1,299	14.88	14.93	-0.07	–1.85		S6_159,591,558	S6_1,62,119,018	KS23-6
AUDPC	*qMLN4-155*	4	431	6.10	2.34	–53.22	–43.81	90.73	S4_6,233,065	S4_155,743,239	CZL3018
	qMLN4-155	4	1,044	5.62	2.81	–6.45	–53.80		S4_13,413,473	S4_232,131,798	CZL3018
	qMLN5-175	5	1,114	10.03	4.22	–70.87	–1.40		S5_174,672,234	S5_200,599,379	CZL3018
	qMLN6-156	6	305	54.62	66.64	61.02	42.45		S6_155,436,477	S6_1,56,119,960	KS23-6
	qMLN6-125	6	596	6.43	2.82	11.57	12.60		S6_78,194,387	S6_125,826,911	KS23-6
	qMLN6-15	6	973	5.65	3.44	–13.92	–60.27		S6_79,988,210	S6_14,732,584	CZL3018
	qMLN7-151	7	359	9.71	3.71	2.58	134.65		S7_150,785,968	S7_163,514,829	KS23-6
	qMLN7-73	7	578	5.24	2.61	–5.78	–54.35		S7_72,579,150	S7_129,968,248	CZL3018
SG4—CML494 x CZL0068											
MLN_DS	*qMLN3-142*	3	805	6.72	16.32	1.16	–2.46	50.36	S3_141,496,550	S3_152,845,401	CML494
	qMLN3-142	3	833	7.71	25.33	1.83	–0.96		S3_141,496,550	S3_223,306,308	CML494
	qMLN6-15	6	611	4.52	8.36	0.89	–3.31		S6_3,836,638	S6_128,503,466	CML494
AUDPC	*qMLN3-142*	3	835	19.98	20.93	60.20	–38.64	31.63	S3_141,496,550	S3_223,306,308	CML494
	qMLN6-15	6	578	3.17	8.22	36.26	34.85		S6_3,836,638	S6_128,503,466	CML494
SG5–CML442 x DTP-F46											
MLN_DS	*qMLN7-150*	7	650	23.06	11.01	–2.92	–0.89	15.21	S7_14,518,869	S7_157,898,866	DTPYC9F46
	qMLN8-163	8	203	18.15	10.93	–0.75	4.09		S8_22,861,113	S8_163,958,071	DTPYC9F46
AUDPC	*qMLN2-24*	2	788	18.05	5.59	–3.18	–95.83	19.01	S2_23,980,190	S2_152,032,245	DTPYC9F46
	qMLN4-155	4	730	18.93	5.58	–1.59	–97.93		S4_6,050,934	S4_162,863,037	DTPYC9F46
	qMLN5-140	5	301	18.59	5.57	–1.01	–97.32		S5_12,215,487	S5_145,497,230	DTPYC9F46
	qMLN7-150	7	650	19.99	5.67	–59.61	–32.99		S7_14,518,869	S7_157,898,866	DTPYC9F46
	qMLN8-163	8	204	18.95	5.77	–16.61	91.35		S8_22,861,113	S8_163,958,071	DTPYC9F46

Chr, chromosome; DS, disease severity; LOD, logarithm of odds; add, additive effect; PVE, phenotypic variance explained; fav allele, parental line contributing for favorable alleles for MLN_DS or AUDPC values, QTL name composed by the trait code followed by the chromosome number in which the QTL was mapped and a physical position of the QTL. QTL names are italicized.

toward susceptibility and was similar to the alleles contributed from CZL3018 (**Figure 3**). For SG4, a QTL on chromosome 3, centered on marker S3_141,496,550 with a LOD score of about 7, explained >20% of the phenotypic variation (**Figure 3**). For this population, both MLN_DS and AUDPC scores for individuals with either homozygous dominant or homozygous recessive were intermediate to those of the parents (**Figure 3**). In SG5, all the identified QTL are associated with the resistant parent DTPYC9F46 (**Table 3**).

QTL Mapping in F_3 Populations

For QM1, linkage analysis detected five QTLs for MLN_DS on chromosomes 6, 7, 9, and 10 with each QTL explaining 8.30%–30.40% of phenotypic variance and together explaining 65.49% of total phenotypic variance (**Table 4**). Four QTLs were identified for AUDPC value in the same populations on chromosomes 1, 6, 7, and 8 with each QTL explaining 3.42%–46.94% of phenotypic variance and together explaining 63.23% of total variation. One major-effect QTL on chromosome 6 (*qMLN6-142*) and another minor effect QTL on chromosome 7 (*qMLN7-6*) were consistent across traits. In QM2, three QTLs were detected on chromosome 6 that individually explained 5.83%–23.68% of phenotypic variance and together explaining 61.92% of total variance for MLN_DS. For AUDPC values, three QTLs detected on chromosome 6 together explained 61.03% of total phenotypic variance, and same QTLs were also detected for

MLN_DS except one QTL *qMLN6-125*. In QM3, three QTLs were detected for MLN_DS, and the same QTLs were also detected for AUDPC values. These QTL individually explained 8.17%–15.86% of phenotypic variance for MLN_DS and 7.60%–13.44% of phenotypic variance for AUDPC values. Whole genome scan for MLN resistance across all the populations revealed a constant peak with a high LOD score in chromosome 6. Although the QTL was identical in all the populations, the QTL intervals and markers varied between populations (**Figure 2**; **Table 4**).

The QTL, found at the interval flanked by markers *PZA02673_1*, *PZA00223_4*, and *PZA01618_2*, was located between bins 6.05 and 6.06, at positions between 85 and 156 Mb on B73 reference genome v2 (**Table 4**). In QM1 and QM2, these SNPs were detected at the LOD above 23, where a threshold of 3 was used to declare significant QTL. Whereas in QM3, of the three SNPs, *PZA02673* was detected at threshold above 3; all hits lower compared to the first two populations. *PZA00223_4* was detected in QM1 and QM2, *PZA01618_2* was only detected in QM2, whereas *PZA02673_1* was detected in QM1 and in QM3. This indicate that favorable alleles of PZA02673_1 and PZA 01618_2 are originated from the parent KS23-5. SNP *S6_156,386,857* was identified in QM1 at a LOD score of 31.03 only at the late stage of MLN_DS (data not shown), whereas *S6_157,568,432* was detected in QM2 at for both MLN_DS and AUDPC values with a LOD score above 20. The QTL was constantly detected across all the populations on chromosome 6 (**Table 4**). Similarly, among all the populations, the MLN-resistant QTL detected in QM1 explained the highest proportion of phenotypic variance followed by QM2 (**Table 4**).

To delimit the physical position of the MLN-resistant QTL, we again used Maize GDB to elucidate the position of the flanking markers against maize B73 reference genome v.2 (**Figure 2**). All the SNPs across the intervals spanned large chromosomal intervals ranging between 25 and 30 Mb. For example, the QTL *qMLN6_157* detected in QM2 is ranged from 129 to 157.5Mb, and same QTL was again detected in SG2 at 151–155 Mb with reduced confidence interval. The physical position indicated by SNPs *S6_157,568,432* and *S6_156,386,857* points to the location of the MLN-resistant QTL in chromosome 6 (**Figure 2**). This position coincided with the positions of the other flanking markers, especially *PZA0022_4*, which was found between *S6_156,461,452* and *S6_157,833,157* bps on maize B73 reference genome.

Genome-Wide Association Study in Multiple Segregating F_2 Populations With Selective Genotyping

Using the eigenstate values generated in TASSEL, a clear population structure was identified in all the groupings using the first three PCs (**Figure 4**). Clustering using the neighbor-joining method performed using TASSEL revealed five clusters representing each of the populations used (**Figure 4**). GWAS results for MLN_DS and AUDPC confirmed the major QTL for MLN resistance on chromosome 6. Manhattan plots from the analysis identified a highly significant genomic region on chromosome 6 both for MLN_DS and AUDPC, whereas some minor QTL were also identified on chromosome 8 (**Figures 5A,B**, **Supplementary Table S2**). The position of the QTL on chromosome 6 was consistent for both MLN_DS and AUDPC values in populations involving KS23; on the other hand, there was no corresponding significant hits in SG4 (CML494 x CZL068) and SG5 (DTP-F46 x CML442) combined population-based GWAS (**Figures 5C,D**). The summary of ~36 significant SNPs identified under MLN_DS scores is presented in **Supplementary Table S2**. The most significant SNPs with $p < 3.57 \times 10^{-24}$ and $p < 2.77 \times 10^{-22}$ were common across MLN_DS and AUDPC values (**Supplementary Table S2**). QTL *qMLN6_157* detected in QM2 is ranged from 129 to 157.5Mb, and same region was again detected in SG2 at 151–155Mb; in SG1 and SG3, QTL was detected between 152 and 161 Mb. With GWAS, we identified several SNPs in this region, but first best 10 SNPs were distributed between 155 and 158 Mb. Taking into account the most significant markers within the segment of 156–157 Mb, the MLN-resistant QTL was estimated to span about 1.7 Mb region on the long arm of chromosome 6.

QTL Validation

There was a significant change in the progression of MLN disease in all the VP families, determined by their genotypic class, with a more rapid progression in the families of susceptible genotypes. During second scoring, observable segregation patterns were seen in families that were heterozygous across the *qMLN_06.157* interval. Within each heterozygous family, susceptible individuals were observed a higher frequency than the resistant individuals. Genotypes in resistant families and those segregating for resistance maintained healthy status up to the fourth score.

Analysis of variance revealed significant effect of the MLN haplotypes on MLN scores (**Table 5**). The marker class means of the MLN disease scores varied across the population. The summary of statistics relating to each of the populations is provided in **Table 5**. The heritability estimates were moderate to high, with all population having heritability >0.7. The mean performance of populations varied, with CML539 (VP2) showing higher scores. Means from the contributions of the MLN-resistant locus in different marker classes are shown in **Figure 6**. As expected, the marker class of resistant haplotype (+/+) has lower scores, whereas high MLN scores are observed in the susceptible genotypes (−/−). The segregating class displayed higher scores, and in some population, this class had nearly the same score as the −/− genotypes (VP1 and VP5).

DISCUSSION

Transboundary diseases have a devastating effect on crop production and severely impact the food security and livelihoods of resource-constrained smallholders and their families (Savary et al., 2019). The MLN disease emerged as a major threat to maize producers in eastern Africa since 2011. Intensive multi-disciplinary and multi-institutional efforts over the last 7–8 years have significantly reduced the spread and

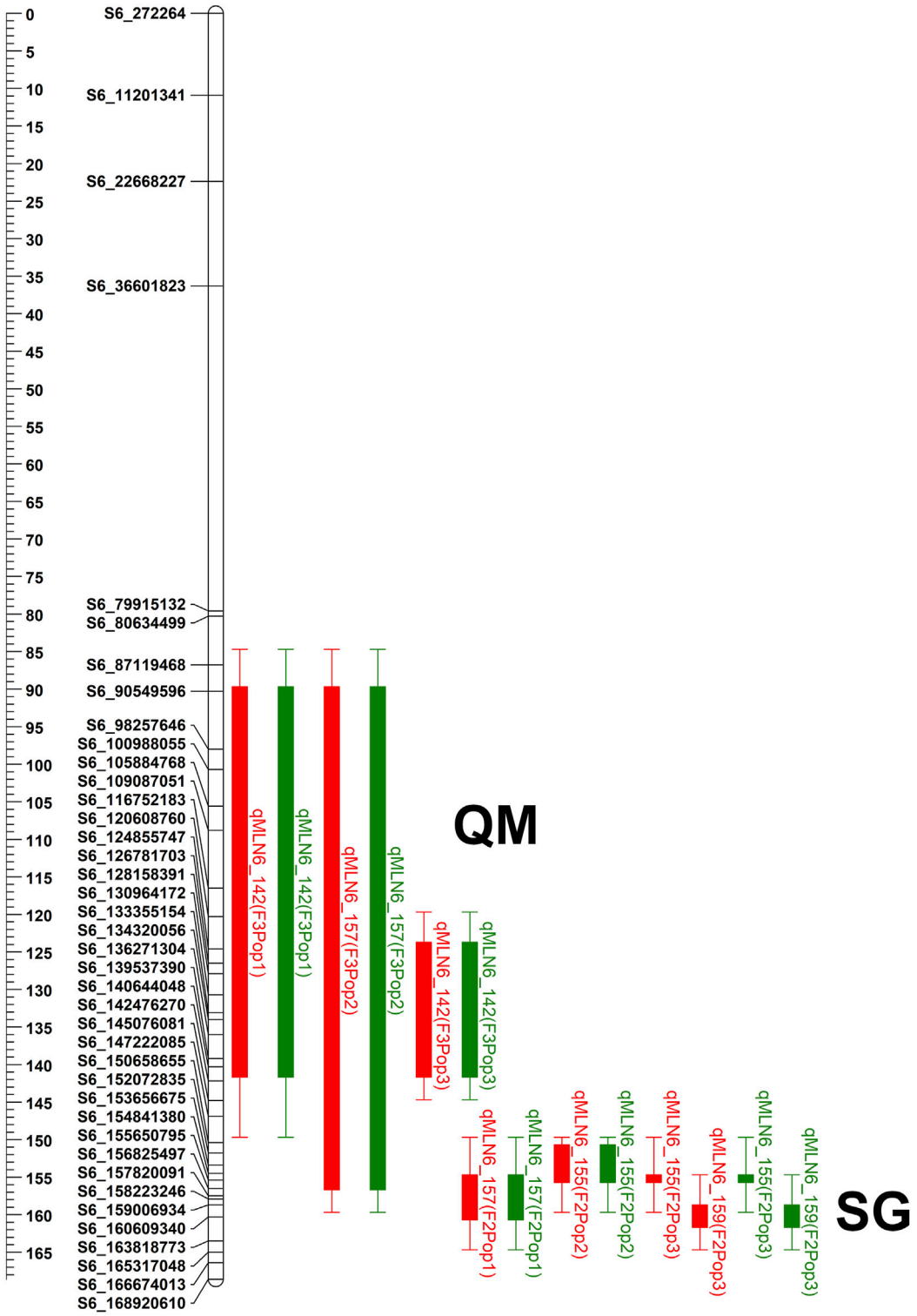


FIGURE 2 | The physical positions of the QTL detected on chromosome 6 in all three F3 QTL mapping (QM) populations and three F2 selective genotyping (SG) populations. QTL bar with red color represents for MLN_DS and green bar represents QTL for AUDPC value. QTL name was composed by the trait code followed by the chromosome number in which the QTL was mapped and a physical position of the QTL.

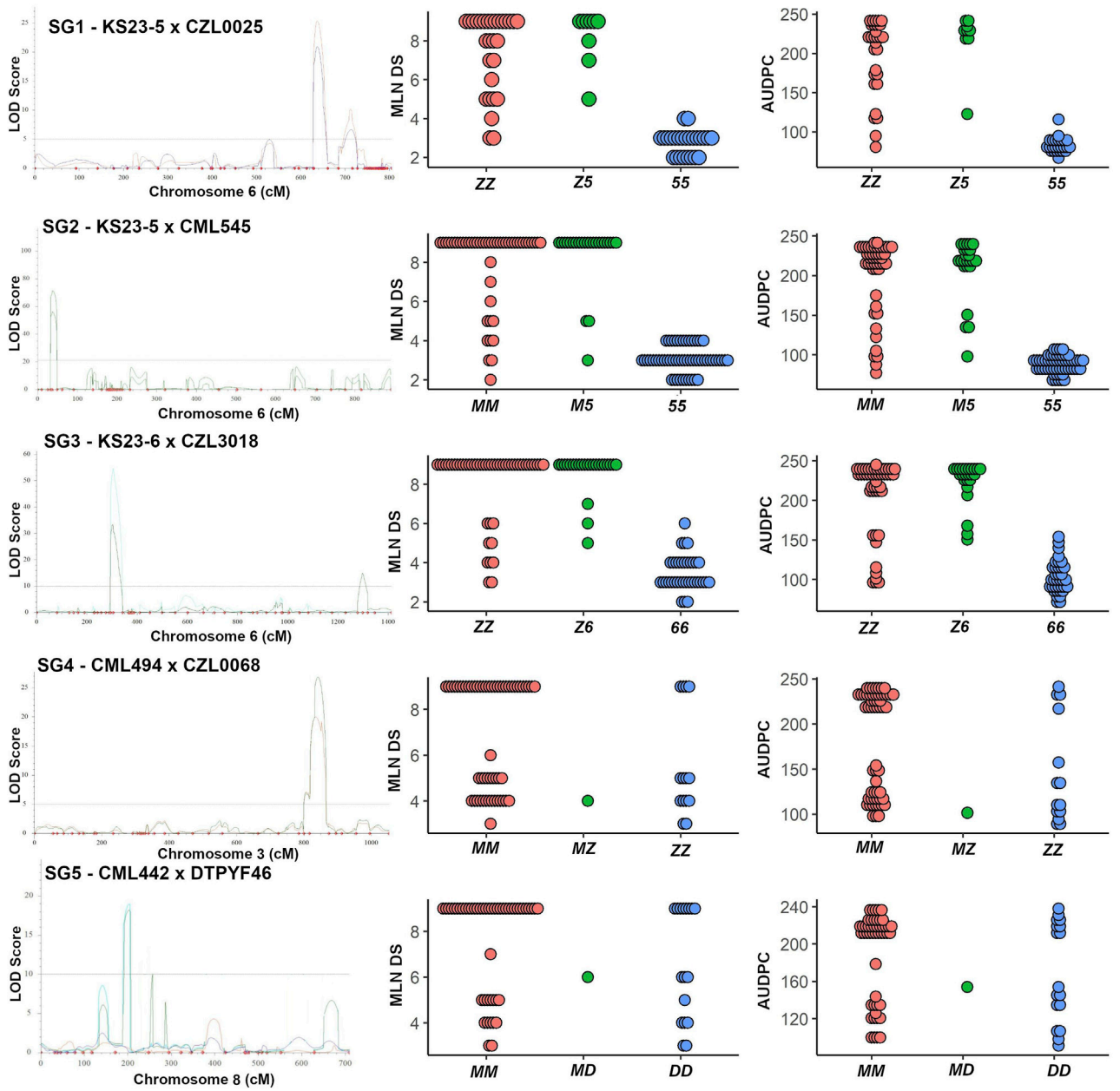


FIGURE 3 | Genomic regions influencing MLN resistance in three KS23 derived populations and non-KS23–based populations. A logarithm of odds (LOD) scan showing the major QTL identified on chromosome 6 (for first three KS23-based populations) and a dot plot for the closest marker for MLN_DS and AUDPC values are shown. Similarly, a LOD scan showing the major QTL identified on chromosome 3 and 8 (SG4 and SG5) and a dot plot for the closest markers are shown. In SG1, ZZ and 55 correspond to alleles from parent CZL0025 and KS23-5, respectively. Similarly, in SG2, MM and 55 correspond to alleles from CML545 and KS23-5; in SG3, ZZ and 66 correspond to CZL3018 and KS23-6; in SG4, MM and ZZ correspond to alleles from CML494 and CZL0068; and in SG5, MM and DD correspond to alleles contributed from CML442 and DTPYF46, respectively.

impact of the disease in Sub-Saharan Africa, but the threat still remains. Breeding and deployment of MLN-resistant elite varieties is an important component of MLN disease management strategy (Prasanna et al., 2020). Although resistance to MLN has been shown to be quantitatively inherited, we still understand very little about its genetic architecture in maize. In the present study, simultaneous mapping strategies were implemented to dissect the genetic architecture of resistance to MLN. On the basis of the excellent resistance offered by two exotic lines KS23-5 and KS23-6 against MLN under artificial inoculation in Kenya, multiple populations were developed and analyzed in this study using linkage mapping and GWAS. Furthermore, the QTL *qMLN6-157*, which was consistently detected at least in

TABLE 4 | QTLs identified for MLN disease severity and AUDPC values, their physical positions, and genetic effects in three F3 QTL mapping (QM) populations.

Trait	QTL name	Chr	Position (cM)	LOD	PVE (%)	Add	Dom	TPVE (%)	QTL confidence interval		QTL physical position		Fav allele
									Left M	Right M	Left M	Right M	
QM1 - CML543*2/CML444:DH5 x KS23-6													
MLN_DS	*qMLN6-142*	6	85	29.08	30.40	0.62	0.25	65.49	PZA00223_4	PZA02673_1	8,510,027	142,648,706	KS23-6
	qMLN7-6	7	11	3.64	8.71	0.21	0.66		PHM9162_135	PHM7898_10	5,632,196	9,124,219	KS23-6
	qMLN7-6	7	109	3.48	8.76	−0.06	0.66		PHM9162_135	PZA00424_1	5,632,196	138,551,416	DH5
	qMLN9-37	9	12	3.83	8.36	−0.08	0.90		PZA00708_3	S9_37,149,685	1,081,791	37,149,685	DH5
	qMLN10-98	10	20	3.99	8.50	0.03	0.77		S10_97,796,845	PHM15331_16	97,796,845	10,432,605	KS23-6
AUDPC	*qMLN1-147*	1	77	3.31	3.42	7.96	0.37	63.23	PZA01019_1	PZA03194_1	146,538,889	212,244,425	KS23-6
	qMLN6-142	6	84	29.59	46.94	27.60	10.50		PZA00223_4	PZA02673_1	8,510,027	142,648,706	KS23-6
	qMLN7-6	7	115	3.08	12.66	−2.06	26.49		PHM9162_135	PZA00424_1	5,632,196	138,551,416	DH5
	qMLN8-121	8	68	3.20	3.45	−3.89	−10.23		PHM5805_19	PZA00498_5	120,860,173	170,377,814	DH5
QM2—CML543*2/CML444:DH6 x KS23-5													
MLN_DS	*qMLN6-142*	6	21	4.33	5.83	1.01	−2.67	61.92	S6_89,823,772	PZA02673_1	89,823,772	142,648,706	KS23-5
	qMLN6-157	6	85	23.60	18.65	1.15	0.24		PZA00223_4	S6_157,568,432	8,510,027	157,568,432	KS23-5
	qMLN6-157	6	89	33.54	23.68	−1.34	−0.08		PZA01618_2	S6_157,568,432	129,927,781	157,568,432	DH6
	qMLN6-157	6	104	6.49	6.49	0.54	0.79		PZA01618_2	S6_157,568,432	129,927,781	157,568,432	KS23-5
AUDPC	*qMLN6-125*	6	70	4.01	2.87	−9.45	−1.30	61.03	PZA00223_4	S6_125,593,444	8,510,027	125,593,444	DH6
	qMLN6-157	6	85	22.38	26.16	29.96	7.80		S6_157,568,432	PZA00223_4	8,510,027	157,568,432	KS23-5
	qMLN6-157	6	87	25.12	30.44	−34.4	9.68		S6_157,568,432	PZA00223_4	8,510,027	157,568,432	DH6
QM3—CML543 x KS23-5													
MLN_DS	*qMLN6-142*	6	87	4.34	8.17	−0.35	0.25	41.21	S6_125,593,444	PZA02673_1	125,593,444	142,648,706	CML543
	qMLN6-142	6	93	8.70	13.30	0.51	0.34		PZA00910_1	PZA02673_1	124,203,565	142,648,706	KS23-5
	qMLN7-13	7	75	3.98	15.86	0.39	1.38		PZA02872_1	PZA00424_1	13,058,813	138,551,416	KS23-5
AUDPC	*qMLN6-142*	6	87	4.41	7.60	−15.14	9.63	26.03	S6_125,593,444	PZA02673_1	125,593,444	142,648,706	CML543
	qMLN6-142	6	93	8.01	11.77	21.53	12.70		PZA00910_1	PZA02673_1	124,203,565	142,648,706	KS23-5
	qMLN7-13	7	78	4.05	13.44	18.87	61.08		PZA02872_1	PZA00424_1	13,058,813	138,551,416	KS23-5

Chr, chromosome; DS, disease severity; LOD, logarithm of odds; add, additive effect; PVE, phenotypic variance explained; fav allele, parental line contributing the favorable allele for MLN resistance, QTL name composed by the trait code followed by the chromosome number in which the QTL was mapped and a physical position of the QTL. QTL names are italicized.

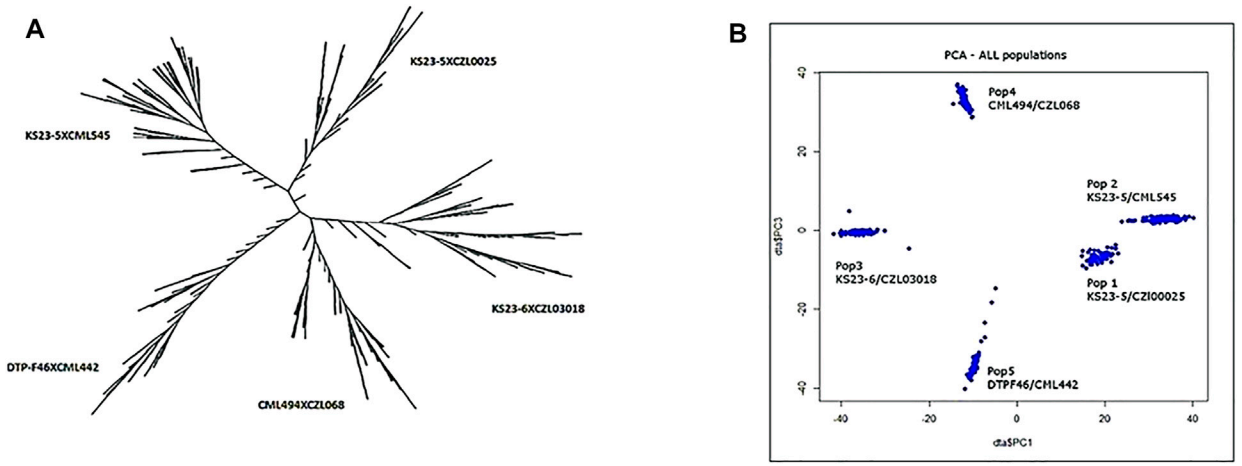


FIGURE 4 | Clustering pattern of 438 genotypes forming five populations using UPGMA method **(A)** and grouping of the genotypes using the first three PCs **(B)**.

one of the three QM F_3 population and five F_2 SG populations, was validated for its effect across various genetic backgrounds in five F_2 VP populations.

QTL for Maize Lethal Necrosis Resistance

Previous studies conducted to map MLN-resistant loci have identified dominantly inherited QTLs in several chromosomes in various mapping populations (Gowda et al., 2015, 2018; Awata et al., 2020; Jones et al., 2018; Nyaga et al., 2020; Sitonik et al., 2019). At the same time, screening thousands of genetically diverse maize inbred lines at the MLN Screening Facility at Naivasha, Kenya, since 2013 led to the identification of two MLN-resistant sister lines, KS23-5 and KS23-6. This is in sharp contrast to the susceptibility of most of the CIMMYT germplasm as well as those of partner institutions in Africa to MLN (Prasanna et al., 2020). The present study utilized the two MLN-resistant lines and undertook linkage mapping as well as GWAS to discover and validate major QTL for resistance to MLN. The bi-parental population-based linkage mapping increased the QTL detection power, whereas GWAS increased the resolution at the detected interval (Kibe et al., 2020).

All three KS23-6– and KS23-5–based populations revealed consistent and very large effect QTL (*qMLN6-157*), where R^2 values ranged from 14.93% to 60.61% for MLN_DS and 16.59%–74.82% for AUDPC (**Figure 2**, **Figure 3**; **Table 3**). Several factors including screening of large F_2 populations and use of artificial inoculation to make sure uniform disease pressure and getting high heritability all helped to accurately map the QTL. On the other hand, overestimation of QTL effects is possibly due to the SG approach used and the relatively small numbers of individuals genotyped. Whenever SG is done against recombinants, it reduces the effective recombination rate near QTL and potentially might cause bias in linkage map construction (Lin and Ritland 1996). However, the population size has a larger effect on linkage maps with multiple fold increase in calculated marker distances compared to map of F_3 populations (**Supplementary Table S1**). Silva et al. (2007) reported that population size has a larger effect on linkage maps with about a threefold increase in marker distances as population size is reduced from 800 to 100 RI lines. Furthermore, deletions or genome rearrangements, often spanning megabases, that occur between maize genotypes (Fu and Dooner 2002; Song and Messing 2003; Brunner et al., 2005; Springer et al., 2009) can cause markers to be out of place on position-based maps due to rearrangements in the genome and causing further increase in map distances. Jones et al. (2018) also reported similar large genetic maps when using F2 populations with SG approach. Further research with large, replicated field trials and fine mapping of the major-effect QTL will pave the way to find more reliable markers for the causal gene. Another major QTL on chromosome 3 was observed from the background of a population without KS23 as parent (**Figure 3**). The QTL identified for MLN resistance on chromosome 3 in non-KS23 populations overlaps with an earlier detected dominantly inherited MLN-resistant QTL (Gowda et al., 2018; Sitonik et al., 2019; Awata et al., 2020).

Using linkage analysis with QM populations, we reconfirmed the major QTL on the long arm of chromosome 6 conferring resistance to MLN (**Figure 2**; **Table 3** and **4**). Given the size of the mapping population, the confidence interval of MLN-resistant QTL discovered was large. For instance, the interval in QM3 was much larger (about 50 Mb) than that of QM1 and QM2. However, the phenotypic variance explained by this QTL at this interval was >10% across all the populations, indicating that the QTL has a major-effect on MLN resistance. The identification of this major QTL for MLN resistance in this experiment is consistent with the QTL identified with SG populations in the background of KS23.

GWAS has been widely used in the discovery of causal variants for resistance to many maize diseases, including MLN (Gowda et al., 2015; Sitonik et al., 2019; Kibe et al., 2020; Nyaga et al., 2020). Although GWAS has proven to be advantageous for discovery of minor alleles, the complexity of the population structure causes a high rate of false positives. The populations

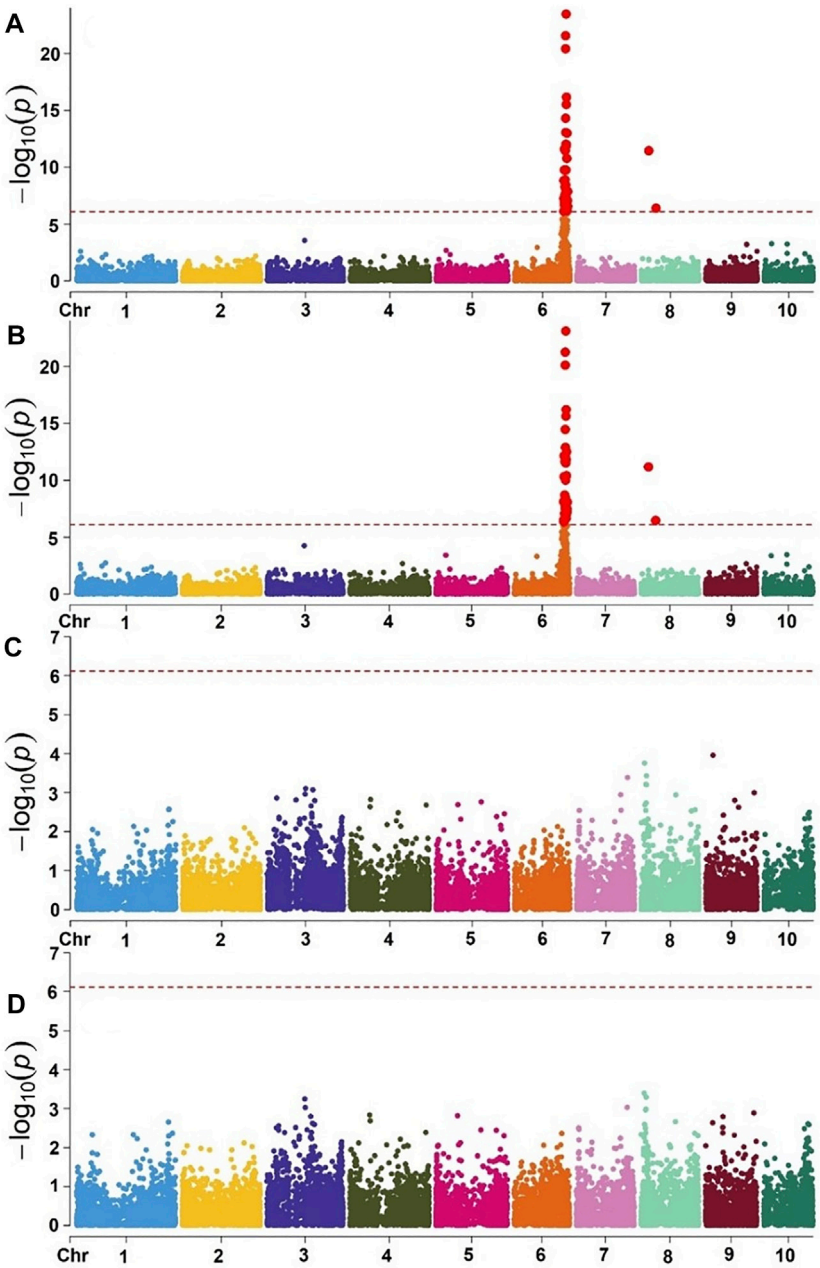


FIGURE 5 | Manhattan plot of GWAS using MLM in the selective genotyping populations. Combined genome-wide association scan for MLN_DS **(A)** and AUDPC values **(B)** based on the first three F2 populations (SG) with KS23 background. Manhattan plots for MLN_DS **(C)** and AUDPC values **(D)** based on two F2 populations (CML494 X CZL068 and DTP-F46 X CML442) with no KS23 background. The horizontal dotted line indicates genome-wide significance and the plots above the line represent SNP markers that showed significance above threshold of $p = 5 \times 10^{-7}$.

used here were developed from parents collected in various CIMMYT breeding programs in Africa as well as Latin America (Mexico) and KS23 lines. Along with population structure, family relatedness/kinship matrix (K matrix) was used in the present study to correct for any possible spurious associations. Principal components analysis identified five clusters that represent the genetic diversity of the populations used (**Figure 4**). These clusters represent allele frequency differences between the populations because of ancestral differences. For instance, the populations based on KS23-5 cluster were closer compared with other populations (**Figure 4**).

GWAS is known to have a high-resolution power due to possible historical recombination events accumulated within a mapping panel. On the contrary, the panel used here comprised F_2 segregating populations developed from parents developed in diverse regions. Furthermore, the mode of phenotypic and

TABLE 5 | Estimates of means, genetic variance, heritability, LSD and coefficient of variation in different populations.

Population		Mean	$\delta^2 g$	$\delta^2 e$	H^2	$LSD_{5\%}$	CV
CML548/KS23-6	MLN Early	3.99	0.76	0.33	0.82	1.11	14.42
	MLN Late	4.77	1.17	0.65	0.78	1.54	16.89
CML539/KS23-6	MLN Early	4.24	1.03	0.38	0.84	1.16	14.62
	MLN Late	5.54	1.63	0.69	0.83	1.6	15.00
CKDHL0186/KS23-6	MLN Early	3.51	0.79	0.35	0.82	1.12	16.93
	MLN Late	5.02	3.54	1.11	0.86	2.04	21.00
CKDHL0221/KS23-6	MLN Early	3.73	1.14	0.70	0.76	1.49	22.44
	MLN Late	5.15	3.86	0.97	0.89	1.86	19.16
CML442/KS23-6	MLN Early	3.98	0.96	0.29	0.87	1.04	13.46
	MLN Late	4.67	1.47	0.64	0.82	1.46	17.18
CML537/KS23-5	MLN Early	4.08	0.84	0.23	0.88	0.92	11.68
	MLN Late	4.75	1.1	0.45	0.83	1.24	14.13

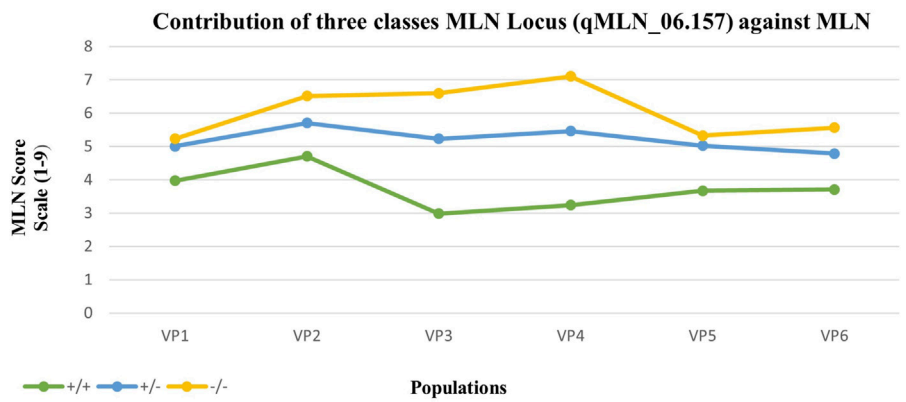


FIGURE 6 | Mean response to MLN inoculation of individuals from contrasting marker classes within six populations. +/+ are homozygous for the KS23 haplotype, +/– are heterozygous, and –/– are homozygous for the susceptible haplotype. All populations were genotyped with 20 SNP markers identified along the MLN-resistant QTL region on chromosome 6.

genotypic data collection used was whole population phenotyping and SG, which differed from traditional methods employed in many GWAS studies. SG is a method that uses extreme phenotypes in a population to maximize their genotypic dissimilarity (Chunfang et al., 2004). It is assumed that the extreme phenotypes harbor diverse alleles/loci that were brought together through intermating of diverse parents (Sun et al., 2010). SG has been used in number of linkage mapping studies in maize to identify QTL associated with resistance to Curvularia leaf spot (Hou et al., 2013), root-lodging resistance (Farkhari et al., 2013), and, more recently, localization of QTL for resistance/tolerance to MCMV (Jones et al., 2018).

Using the selective genotyping approach, the association study identified 20 SNPs significantly associated with MLN resistance on chromosome 6 (**Figure 5A–B**, **Table 6**, **Supplementary Table S2**). In contrast to the linkage mapping done here, GWAS drastically increased the resolution and enabled reduction of the confidence interval of MLN-resistant QTL from 30 Mbps to an interval of 1–5 Mbps (**Figures 2**, **5**). The groupings of the population, either having KS23 donor background or not, further reinforced the presence of a major-effect MLN-resistant QTL from KS23 background. For instance, although the clusters containing CIMMYT lines revealed two minor QTLs on chromosome 3 and 8, they did not show any significant association with the major QTL on chromosome 6 as opposed to the clusters having a K23 background (**Figures 5C,D**). The absence of this locus in the CIMMYT breeding lines analyzed in the study as well as the elite maize germplasm from the national agricultural research systems (NARS) institutions in eastern and southern Africa indicates that the QTL identified here is indeed unique to the KS23 background resulting in the favorable phenotype. The discovery of the major-effect QTL and the significant phenotypic variation explained by the QTL paves way for fast-tracking introgression of MLN resistance from KS23 lines into CIMMYT/NARS lines adapted to Sub-Saharan Africa. The QTL also reconfirms the MCMV resistance QTL reported in F_2 populations by Jones et al. (2018).

QTL Validation

For verification of the haplotype block associated with MLN resistance on chromosome 6 (*qMLN6_157*), a set of independent breeding populations were developed and tested to determine the proportion of phenotypic variance explained by the QTL. Six F_2 populations were genotyped with 20 SNP markers identified

TABLE 6 | Chromosomal positions and SNPs significantly associated with both MLN disease severity (DS) and area under the disease progress curve (AUDPC) and used in validation experiments.

SNP name	Chr	BP	MLM-P values	PVE	Fav allele	Allele effect
S6_155,632,957	6	155,632,957	4.30E-26	0.31	A	4.30
S6_157,168,501	6	157,168,501	1.22E-24	0.29	C	0.00
S6_157,914,681	6	157,914,681	1.60E-19	0.22243	A	50.05
S6_156,249,290	6	156,249,290	9.46E-16	0.17412	T	–50.03
S6_156,841,805	6	156,841,805	8.95E-15	0.16194	T	–2.30
S6_156,373,000	6	156,373,000	1.95E-13	0.14544	T	7.91
S6_155,646,296	6	155,646,296	4.15E-13	0.36368	G	–5.31
S6_155,436,477	6	155,436,477	1.64E-12	0.13418	T	–3.64
S6_151,486,592	6	151,486,592	4.19E-10	0.10537	T	–2.45
S6_155,990,350	6	155,990,350	3.95E-09	0.08379	T	0.00
S6_153,471,979	6	153,471,979	8.63E-09	0.08996	T	–2.45
S6_158,281,554	6	158,281,554	1.54E-08	0.08703	C	18.04
S6_149,124,264	6	149,124,264	1.83E-08	0.08618	G	29.77
S6_153,264,776	6	153,264,776	5.45E-07	0.32948	T	–55.6
S6_153,422,344	6	153,422,344	6.18E-07	0.27911	C	–42.4
S6_153,261,193	6	153,261,193	6.52E-07	0.32475	T	–54.3
S6_151,035,391	6	151,035,391	1.73E-06	0.15491	G	–1.95
S6_1,61,218,736	6	1,61,218,736	1.74E-06	0.15483	C	–19.4
S6_151,035,617	6	151,035,617	2.16E-06	0.15213	T	–3.78
S6_153,661,647	6	153,661,647	1.68E-05	0.24219	C	–6.27

Chr, chromosome; BP, physical position in base pairs; PVE, phenotypic variance explained; fav allele, allele associated with MLN, resistance.

along the MLN-resistant QTL (*qMLN6_157*) region on chromosome 6. Each population was partitioned into three marker class genotypes, homozygous for the resistant parent (donor) and homozygous for susceptible or segregating individuals. Disease progression in each of these classes was distinct, indicating the effect of *qMLN_06.157* under MLN. Low scores observed within the resistant class indicate a significant effect of the QTL to MLN resistance (**Table 5**; **Figure 6**). This reconfirms the reliability of QTL detected through linkage mapping and GWAS and increases the confidence on this QTL to include as part of marker assisted breeding to improve the MLN resistance. Nevertheless, further fine mapping of this region and finding tightly linked flanking markers can enhance the efficiency to improve MLN resistance as well as introgress this QTL into elite susceptible lines.

In maize, the genetics of virus resistance has been studied in both temperate and tropical germplasm (Redinbaugh et al., 2018). Few inbred lines with strong resistance to multiple viruses have been identified. For instance, line Oh1VI was identified as highly resistant to MDMV, SCMV, WSMV, Maize chlorotic dwarf virus, Maize fine streak virus (MFSV), Maize mosaic virus (MMV), and Maize necrotic streak virus (Zambrano et al., 2014; Redinbaugh et al., 2018). After the inoculation of viruses, resistance lines for these viruses showed either no or few symptoms with significantly reduced virus titer in plant tissues. On the contrary for MCMV, Jones et al. (2018) observed high titer in systemic tissues of the identified resistant lines. Genetic analyses of F_2 populations for MFSV, MMV, MDMV, SCMV, and WSMV suggest that resistance to these viruses is governed by both additive and dominant genes (Redinbaugh and Zambrano, 2014). Resistance QTLs with additive or dominant gene action for all these five viruses were clustered in the same genomic region at chromosome 3 and 6 (Zambrano et al., 2014; Awata et al., 2020). For SCMV resistance, two major dominantly inherited QTL/genes, namely, *Scmv1* and *Scmv2*, were identified and fine mapped on the short arm of chromosome 6, and near the centromere of chromosome 3, respectively (Tao et al., 2013; Redinbaugh et al., 2018). For expression of complete resistance, both resistance genes must act together time where resistance at all developmental stages is provided by *Scmv1*, and resistance expressed at later stages of plant development is governed by *Scmv2* (Xia et al., 1999; Ding et al., 2012). On the contrary, MCMV is governed by major QTL, and recessive nature was identified in two F_2 populations (Jones et al., 2018).

The resistance exhibited by MLN-resistant QTL (*qMLN6_157*) in this study suggests a natural recessive inheritance. Susceptible individuals were observed at a higher frequency in the segregating families compared with the resistant counterparts (**Table 5**; **Figure 6**). More so, the mean score of the segregating families was closer to that families selected to be homozygous susceptible (**Figure 6**). The recessive nature of the QTL is also inferred by phenotypic distribution of the F_2 population, which is skewed toward susceptible phenotype (**Figure 1**). Some differences in gene action on controlling the MLN resistance or tolerance were observed when the distribution of phenotypes in the F2 populations and the phenotypes of F2 plants heterozygous for closely linked markers to the MLN-resistant QTL were compared. In SG1, SG2, and SG3, the MLN_DS and AUDPC scores of plants heterozygous for the markers tightly linked to the QTL on chromosome 6 were similar to those of the susceptible parent (**Figure 3**), suggesting that tolerance is controlled by a recessive gene. On the contrary, in SG4 and SG5 populations, the distribution of MLN_DS and AUDPC scores appears to be more normally distributed, which is consistent with the identification of several QTL with smaller

contributions to tolerance in these lines. Similar observations on the recessive nature of inheritance of the QTL were made by Jones et al. (2018) using F_2 population developed from KS23 background, tested under MCMV. Both MLN_DS and the AUDPC scores of plants heterozygous for the markers are closely linked to the large effect QTL (*qMLN6-157*) on chromosome 6 (**Figure 3**), further suggesting that MLN resistance is controlled by a recessively inherited gene. The QTL validation experiment confirmed the robustness of the QTL and the potential benefit of introgression of the QTL in desired genetic backgrounds. The identity of the closely linked markers would be useful for indirect selection of the QTL especially in marker-assisted backcrossing as well as forward breeding methods. On the other hand, having single gene or QTL-based resistance is always risk of losing it quickly at any given point of time especially after introgression of the QTL to many elite lines. To avoid such scenario, it is suggested to combine the QTL (*qMLN6-157*) with other MLN resistant, dominantly inherited major-effect QTL from chromosome 3 (region of 130–140 Mb) and chromosome 6 (region of 5–20 Mb; Gowda et al., 2015, 2018; Sitonik et al., 2019; Nyaga et al., 2020; Awata et al., 2020; Awata et al., 2021).

CONCLUSION

The present study identified a novel QTL for MLN resistance in the genetic background of two sister lines KS23-5 and KS23-6, on the basis of a source population constituted at the University of Hawaii, United States, using germplasm from Kasetsart University, Thailand. These two lines showed excellent resistance to MLN when tested under artificial inoculation at the MLN Screening Facility at Naivasha, Kenya, and thus serve as trait donors for improving MLN resistance in maize breeding pipelines in Africa. Both linkage mapping and association mapping approaches were used in the study to discover and validate genomic regions associated with MLN resistance. Three F_3 bi-parental populations and five F_2 bi-parental populations were used in linkage mapping, and all the F_2 populations were used in SG-based association mapping. One major QTL unique to KS23 background, and recessively inherited, was identified on the long arm of chromosome 6, designated as *qMLN_06.157*. This QTL was consistently detected in at least one of three F_3 populations, and in the three KS23 derived F_2 populations and in the GWAS. This QTL could not be identified in populations that were not based on KS23 lines. The validation study confirmed that the QTL is consistent and expressing in KS23 genetic and different environmental background, as well as it is recessively inherited with a major-effect, indicating its great potential for application in marker-assisted breeding for MLN resistance. Furthermore, the unique nature of MLN resistance conferred by *qMLN_06.157* warrants further study using fine mapping and gene cloning to investigates gene(s) or casual variations within KS23 background that confers the favorable phenotype. Overall, *qMLN157* is a novel major-effect recessively inherited QTL and could be targeted for further fine mapping, for developing breeder-friendly diagnostic markers and for further introgression of MLN resistance into desired genetic backgrounds in 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

AM, MO, DK, BD, BP, YB, and MG conceived the experiments; BD, OV, DK, SL, BE, and AM conducted field evaluations and phenotyping; MG, MO, and OV coordinated sample genotyping; MG, DK, and AM carried out GWAS and linkage analyses; AM, BD, YB, MO, BP, DK, BE, SL, and MG interpreted the results and drafted the manuscript.

FUNDING

The research was supported by the Bill and Melinda Gates Foundation (B and MGF) and the United States Agency for International Development (USAID) through the Stress Tolerant Maize for Africa (STMA; B and MGF grant no. OPP1134248) Project, IMAS Project, HTPG Project, AGG (Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods; B and MGF Investment ID INV-003439) Project and the CGIAR Research Program on Maize (MAIZE). MAIZE receives W1&W2 support from the Governments of Australia, Belgium, Canada, China, France, India, Japan, Korea, Mexico, Netherlands, New Zealand, Norway, Sweden, Switzerland, United Kingdom, United States, and the World Bank. Our sincere thanks to CIMMYT technicians in Kenya; the Integrated Genotyping Service and Support (IGSS) platform, BecA, ILRI, Kenya, and Diversity Array Technology (DArT) Ltd. for the high-density genotyping and imputation service.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.767883/full#supplementary-material

REFERENCES

Awata, L. A. O., Beyene, Y., Gowda, M., L. M., S., Jumbo, M. B., Tongoona, P., et al. (2020). Genetic Analysis of QTL for Resistance to *Maize Lethal Necrosis* in Multiple Mapping Populations. *Genes* 11 (1), 32. doi:10.3390/genes11010032

Awata, L. A. O., Ifie, B. E., Danquah, E., Jumbo, M. B., Suresh, L. M., Gowda, M., et al. (2021). Introgression of Maize Lethal Necrosis Resistance Quantitative Trait Loci Into Susceptible Maize Populations and Validation of the Resistance Under Field Conditions in Naivasha, Kenya. *Front. Plant Sci.* 12, 649308. doi:10.3389/fpls.2021.649308

Barnett, I. J., Lee, S., and Lin, X. (2013). Detecting Rare Variant Effects Using Extreme Phenotype Sampling in Sequencing Association Studies. *Genet. Epidemiol.* 37, 142–151. doi:10.1002/gepi.21699

Beyene, Y., Gowda, M., Suresh, L. M., Mugo, S., Olsen, M., Oikeh, S. O., et al. (2017). Genetic Analysis of Tropical maize Inbred Lines for Resistance to maize Lethal Necrosis Disease. *Euphytica* 213, 224. doi:10.1007/s10681-017-2012-3

Boddupalli, P., Suresh, L. M., Mwatuni, F., Beyene, Y., Makumbi, D., Gowda, M., et al. (2020). Maize Lethal Necrosis (MLN): Efforts toward Containing the Spread and Impact of a Devastating Transboundary Disease in Sub-saharan Africa. *Virus. Res.* 282, 197943. doi:10.1016/j.virusres.2020.197943

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

Brewbaker, J. L. (2009). Registration of Nine maize Populations Resistant to Tropical Diseases. *J. Plant Reg.* 3 (1), 10–13. doi:10.3198/jpr2008.07.0396crc

Brunner, S., Fengler, K., Morgante, M., Tingey, S., and Rafalski, A. (2005). Evolution of DNA Sequence Nonhomologies Among maize Inbreds. *Plant Cell* 17, 343–360. doi:10.1105/tpc.104.025627

Campbell, C. L., and Madden, L. V. (1990). *Introduction to Plant Disease Epidemiology*. New York, NY: John Wiley & Sons.

Chunfang, J., Hong, L., Attie, A. D., Churchill, G. A., Bulutuglo, D., and Yandell, B. S. (2004). Selective Phenotyping for Increased Efficiency in Genetic Mapping Studies. *Genetics* 168 (4), 2285–2293.

Cornish, P. L., Knowles, S. R., Marchesano, R., Tam, V., Shadowitz, S., Juurlink, D. N., et al. (2005). Unintended Medication Discrepancies at the Time of Hospital Admission. *Arch Intern. Med.* 165 (4), 424–429. doi:10.1001/archinte.165.4.424

De Groote, H., Munyua, B. G., Palmas, S., Suresh, L. M., Bruce, A. Y., and Kimenju, S. (2021). Using Panel Community Surveys to Track the Impact of Crop Pests over Time and Space - the Case of Maize Lethal Necrosis (MLN) Disease in Kenya from 2013 to 2018. *Plant Dis.* 105 (5), 1259–1271. doi:10.1094/PDIS-08-20-1730-SR

De Groote, H., Oloo, F., Tongruksawattana, S., and Das, B. (2016). Community-survey Based Assessment of the Geographic Distribution and Impact of maize Lethal Necrosis (MLN) Disease in Kenya. *Crop Prot.* 82, 30–35. doi:10.1016/j.cropro.2015.12.003

Ding, J., Li, H., Wang, Y., Zhao, R., Zhang, X., Chen, J., et al. (2012). Fine Mapping of *Rscmv2*, a Major Gene for Resistance to Sugarcane Mosaic Virus in maize. *Mol. Breed.* 30, 1593–1600. doi:10.1007/s11032-012-9741-8

Farkhari, M., Krivanek, A., Xu, Y., Rong, T., Naghavi, M. R., Samadi, B. Y., et al. (2013). Root-lodging Resistance in maize as an Example for High-Throughput Genetic Mapping via Single Nucleotide Polymorphism-Based Selective Genotyping. *Plant Breed* 132 (1), 90–98. doi:10.1111/pbr.12010

Fu, H., and Dooner, H. K. (2002). Intraspecific Violation of Genetic Colinearity and its Implications in maize. *Proc. Natl. Acad. Sci.* 99, 9573–9578. doi:10.1073/pnas.132259199

Gowda, M., Beyene, Y., Makumbi, D., Semagn, K., Olsen, M. S., Bright, J. M., et al. (2018). Discovery and Validation of Genomic Regions Associated with Resistance to maize Lethal Necrosis in Four Biparental Populations. *Mol. Breed.* 38 (5), 66. doi:10.1007/s11032-018-0829-7

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., et al. (2015). Genome-wide Association and Genomic Prediction of Resistance to maize Lethal Necrosis Disease in Tropical maize Germplasm. *Theor. Appl. Genet.* 128, 1957–1968. doi:10.1007/s00122-015-2559-0

Gregorio, A., Marco, L., Mateo, V., Pacheo, A., Rodriguez, F., Burgueno, J., et al. (2015). META-R (Multi Environment Trail Analysis with R for Windows). Version 6.04. https://hdl.handle.net/11529/10201, V23. CIMMYT Research Data and Software Repository Network.

Herbert, A., Gerry, N. P., McQueen, M. B., Heid, I. M., Pfeufer, A., Illig, T., et al. (2006). A Common Genetic Variant Is Associated with Adult and Childhood Obesity. *Science* 312 (5771), 279–283. doi:10.1126/science.1124779

Holland, J. (2007). Genetic Architecture of Complex Traits in Plants. *Curr. Opin. Plant Biol.* 10 (2), 156–161. doi:10.1016/j.pbi.2007.01.003

Hou, J., Xing, Y., Zhang, Y., Tao, Y., Tan, G., and Xu, M. (2013). Identification of Quantitative Trait Loci for Resistance to *Curvularia Leaf Spot* of maize. *Maydica* 58 (3), 266–273.

Jones, M. W., Penning, B. W., Jamann, T. M., Glaubitz, J. C., Romay, C., Buckler, E. S., et al. (2018). Diverse Chromosomal Locations of Quantitative Trait Loci for Tolerance to *Maize Chlorotic Mottle Virus* in Five maize Populations. *Phytopathology* 108 (6), 748–758. doi:10.1094/phyto-09-17-0321-r

Kaeppler, S. M., Galusha, D. D., Thomas-Compton, M., Doupnik, B., Jensen, S., and Compton, W. A. (1998). Registration of N211, N217, and N218 Parental Inbred Lines of Maize. *Crop Sci.* 38 (1), 289. doi:10.2135/cropsci1998.0011183x003800010069x

Kibe, M., Nyaga, C., Nair, S. K., Beyene, Y., Das, B., M, S. L., et al. (2020). Combination of Linkage Mapping, GWAS, and GP to Dissect the Genetic Basis of Common Rust Resistance in Tropical maize Germplasm. *Ijms* 21 (18), 6518. doi:10.3390/ijms21186518

Lin, J.-Z., and Ritland, K. (1996). The Effects of Selective Genotyping on Estimates of Proportion of Recombination between Linked Quantitative Trait Loci. *Theoret. Appl. Genet.* 93, 1261–1266. doi:10.1007/bf00223458

Madden, L. V., Hughes, G., and van den Bosch, F. (2007). *The Study of Plant Disease Epidemics*. Saint Paul, MN: The American Phytopathological Society.

Mahuku, G., Lockhart, B. E., Wanjala, B., Jones, M. W., Kimunye, J. N., Stewart, L. R., et al. (2015). Maize Lethal Necrosis (MLN), an Emerging Threat to maize-based Food Security in Sub-saharan Africa. *Phytopathology* 105, 956–965. doi:10.1016/j.virusres.2020.19794310.1094/phyto-12-14-0367-fi

Marenya, P. P., Erenstein, O., Prasanna, B., Makumbi, D., Jumbo, M., and Beyene, Y. (2018). Maize Lethal Necrosis Disease: Evaluating Agronomic and Genetic Control Strategies for Ethiopia and Kenya. *Agric. Syst.* 162, 220–228. doi:10.1016/j.agsy.2018.01.016

McCord, P., Glynn, N., and Comstock, J. (2019). Identifying Markers for Resistance to Sugarcane orange Rust (*Puccinia Kuehnii*) via Selective Genotyping and Capture Sequencing. *Euphytica* 215, 150. doi:10.1007/s10681-019-2340-6

McDonald, B. A., and Stukenbrock, E. H. (2016). Rapid Emergence of Pathogens in Agro-Ecosystems: Global Threats to Agricultural Sustainability and Food Security. *Phil. Trans. R. Soc. B* 371 (1709), 20160026. doi:10.1098/rstb.2016.0026

Nyaga, C., Gowda, M., Beyene, Y., Muriithi, W. T., Makumbi, D., Olsen, M. S., et al. (2020). Genome-Wide Analyses and Prediction of Resistance to MLN in Large Tropical Maize Germplasm. *Genes* 11 (1), 16. doi:10.3390/genes11010016

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal Components Analysis Corrects for Stratification in Genome-wide Association Studies. *Nat. Genet.* 38, 904–909. doi:10.1038/ng1847

Redinbaugh, M. G., Lübberstedt, T., Leng, P., and Xu, M. (2018). "The Genetics and Genomics of Virus Resistance in maize," in *The Maize Genome. Compendium of Plant Genomes*. Editors J. Bennetzen, S. Flint Garcia, C. Hirsch, and R. Tuberosa (Cham: Springer), 185–200. doi:10.1007/978-3-319-97427-9_12)

Redinbaugh, M. G., and Zambrano, J. L. (2014). "Control of Virus Diseases in maize," in *Control of Plant Virus Disease: Seed-Propagated Crops*. Editors G. Lobenstein and N. Katis (Waltham, MA USA: Academic Press), Vol. 90, 391–429. Advances in Virus Research. doi:10.1016/b978-0-12-801246-8.00008-1

Saunak, S., Johannes, F., and Broman, K. W. (2009). Selective Genotyping and Phenotyping Strategies in a Complex Trait Context. *Genetics* 181, 1613–1626. doi:10.1534/genetics.108.094607

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., and Nelson, A. (2019). The Global burden of Pathogens and Pests on Major Food Crops. *Nat. Ecol. Evol.* 3, 430–439. doi:10.1038/s41559-018-0793-y

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2013). Single Nucleotide Polymorphism Genotyping Using Kompetitive Allele Specific PCR (KASP): Overview of the Technology and its Application in Crop Improvement. *Mol. Breed.* 33, 1–14. doi:10.1007/s11032-013-9917-x

Semagn, K., Beyene, Y., Babu, R., Nair, S., Gowda, M., Das, B., et al. (2015). Quantitative Trait Loci Mapping and Molecular Breeding for Developing Stress Resilient maize for Sub-saharan Africa. *Crop Sci.* 55, 1–11. doi:10.2135/cropsci2014.09.0646

Sitonik, C. a., Suresh, L. M., Beyene, Y., and Olsen, M. S. (2019). Simulation of Population Size and Genome Saturation Level for Genetic Mapping of Recombinant Inbred Lines (RILs). *Genet. Mol. Biol.* 30, 1101–1108. doi:10.1590/s1415-47572007000600013

Song, R., and Messing, J. (2003). Gene Expression of a Gene Family in maize Based on Noncollinear Haplotypes. *Proc. Natl. Acad. Sci.* 100, 9055–9060. doi:10.1073/pnas.1032999100

Springer, N. M., Ying, K., Fu, Y., Ji, T., Yeh, C.-T., Jia, Y., et al. (2009). Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/absence Variation (PAV) in Genome Content. *Plos Genet.* 5, e1000734. doi:10.1371/journal.pgen.1000734

Sun, Y., Wang, J., Crouch, J. H., and Xu, Y. (2010). Efficiency of Selective Genotyping for Genetic Analysis of Complex Traits and Potential Applications in Crop Improvement. *Mol. Breed.* 26, 493–511. doi:10.1007/s11032-010-9390-8

Tao, Y., Jiang, L., Liu, Q., Zhang, Y., Zhang, R., Ingvardsen, C. R., et al. (2013). Combined Linkage and Association Mapping Reveals Candidates for *Scmv1*, a Major Locus Involved in Resistance to Sugarcane Mosaic Virus (SCMV) in maize. *BMC Plant Biol.* 13, 162. doi:10.1186/1471-2229-13-162

Uemoto, Y., Abe, T., Tameoka, N., Hasebe, H., Inoue, K., Nakajima, H., et al. (2011). Whole-genome Association Study for Fatty Acid Composition of Oleic Acid in Japanese Black Cattle. *Anim. Genet.* 42, 141–148. doi:10.1111/j.1365-2052.2010.02088.x

Wangai, A. W., Redinbaugh, M. G., Kinyua, Z. M., Miano, D. W., Leley, P. K., Kasina, M., et al. (2012). First Report of Maize Chlorotic Mottle Virus and Maize Lethal Necrosis in Kenya. *Plant Dis.* 96 (10), 1582. doi:10.1094/pdis-06-12-0576-pdn

Xia, X., Melchinger, A. E., Kuntze, L., and Lübberstedt, T. (1999). Quantitative Trait Loci Mapping of Resistance to *Sugarcane Mosaic Virus* in maize. *Phytopathology* 89, 660–667. doi:10.1094/phyto.1999.89.8.660

Zambrano, J. L., Jones, M. W., Brenner, E., Francis, D. M., Tomas, A., and Redinbaugh, M. G. (2014). Genetic Analysis of Resistance to Six Virus Diseases in a Multiple Virus-Resistant maize Inbred Line. *Theor. Appl. Genet.* 127, 867–880. doi:10.1007/s00122-014-2263-5

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 24 November 2021
doi: 10.3389/fgene.2021.790789

DCET1 Controls Male Sterility Through Callose Regulation, Exine Formation, and Tapetal Programmed Cell Death in Rice

Riaz Muhammad Khan[1], *Ping Yu*[1], *Lianping Sun*[1], *Adil Abbas*[1], *Liaqat Shah*[2], *Xiaojiao Xiang*[1], *Dongfei Wang*[1], *Amir Sohail*[1], *Yingxin Zhang*[1], *Qunen Liu*[1], *Shihua Cheng*[1] *and Liyong Cao*[1]*

[1]*Key Laboratory for Zhejiang Super Rice Research and State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou, China,* [2]*Department of Botany, Mir Chakar Khan Rind University, Sibi, Pakistan*

OPEN ACCESS

Edited by:
Awais Rasheed,
Quaid-i-Azam University, Pakistan

Reviewed by:
Rajeev Ranjan,
Purdue University, United States
Zhanjie Li,
Fujian Agriculture and Forestry University, China

***Correspondence:**
Liyong Cao
caoliyong@caas.cn

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *07 October 2021*
Accepted: *04 November 2021*
Published: *24 November 2021*

Citation:
Khan RM, Yu P, Sun L, Abbas A, Shah L, Xiang X, Wang D, Sohail A, Zhang Y, Liu Q, Cheng S and Cao L (2021) DCET1 *Controls Male Sterility Through Callose Regulation, Exine Formation, and Tapetal Programmed Cell Death in Rice. Front. Genet. 12:790789. doi: 10.3389/fgene.2021.790789*

In angiosperms, anther development comprises of various complex and interrelated biological processes, critically needed for pollen viability. The transitory callose layer serves to separate the meiocytes. It helps in primexine formation, while the timely degradation of tapetal cells is essential for the timely callose wall dissolution and pollen wall formation by providing nutrients for pollen growth. In rice, many genes have been reported and functionally characterized that are involved in callose regulation and pollen wall patterning, including timely programmed cell death (PCD) of the tapetum, but the mechanism of pollen development largely remains ambiguous. We identified and functionally characterized a rice mutant *dcet1*, having a complete male-sterile phenotype caused by defects in anther callose wall, exine patterning, and tapetal PCD. *DCET1* belongs to the RNA recognition motif (RRM)-containing family also called as the ribonucleoprotein (RNP) domain or RNA-binding domain (RBD) protein, having single-nucleotide polymorphism (SNP) substitution from G (threonine-192) to A (isoleucine-192) located at the fifth exon of LOC_Os08g02330, was responsible for the male sterile phenotype in mutant *dcet1*. Our cytological analysis suggested that *DCET1* regulates callose biosynthesis and degradation, pollen exine formation by affecting exine wall patterning, including abnormal nexine, collapsed bacula, and irregular tectum, and timely PCD by delaying the tapetal cell degeneration. As a result, the microspore of *dcet1* was swollen and abnormally bursted and even collapsed within the anther locule characterizing complete male sterility. GUS and qRT-PCR analysis indicated that *DCET1* is specifically expressed in the anther till the developmental stage 9, consistent with the observed phenotype. The characterization of *DCET1* in callose regulation, pollen wall patterning, and tapetal cell PCD strengthens our knowledge for knowing the regulatory pathways involved in rice male reproductive development and has future prospects in hybrid rice breeding.

Keywords: male sterility, callose, pollen exine, tapetum, PCD, *DCET1*

INTRODUCTION

Rice (*Oryza sativa* L.) is a staple food crop worldwide, feeding around three billion people, nearly half of the global population (Cheng et al., 2007). The global population is intensively increasing, and water resources and agricultural land for rice production are shrinking because of urban expansion and climatic changes. Therefore, new breeding strategies are critically needed to vertically enhance the rice production. Several measures can help overcome food shortage, such as eradicating soil problems, improving cultural practices, proper control of pests and diseases, use of fertilizers, proper utilization of water resources, and use of elite varieties and hybrids (Li et al., 2009). Hybrid rice production is one of the key technologies, specifically among the genetic options, to overcome the growing population's food shortage. Rice is a self-pollinated crop; therefore, the male sterility technique is used to develop commercial hybrid parental lines (Chang et al., 2016). Male sterility produces infertile pollens, so that rice spikelets are incapable of setting seeds through self-pollination (Chang et al., 2016). Thus, pollen viability is an important character for improved rice production (Zou et al., 2017b). Pollen is the key source for improved rice grain production; therefore, understanding the mechanism of pollen development is extremely important for rice breeding (Zhang et al., 2011).

At the end of meiosis-II, the callose is degraded by the tapetum-secreted callases/glucanases, which initiate pollen wall formation (Shi et al., 2015). The first fabricated complex structure of the pollen wall is cellulose-formed primexine, which is the functional site for sporopollenin precursors (Jiang et al., 2013; Lou et al., 2014b; Zhang and Li, 2014). In higher plants, the pollen wall is a bilayer, and consists of an outer layer exine and an inner layer intine (Hu et al., 2014). The exine is the outer protective wall of pollen grains, which plays an important role in the selective flow of fluids across the pollen and attracting pollinators (Scott, 1994; Scott et al., 2004; Ariizumi and Toriyama, 2011). The pollen wall also protects the pollen from microbial attacks and various environmental stresses (Lou et al., 2014a).

In rice, pollen exine is majorly made up of sporopollenin. It consists of an outer layer of sexine (tectum), a foot layer (nexine), the middle bacula, and the tryphine in the cavities (Blackmore et al., 2007). The sexine usually identifies the species, while the nexine is used as a frame for exine formation (Shi et al., 2015). The exine (cell wall polymer), is involved in sperm protection, consisting of three major developmental processes, such as primexine, callose wall, and sporopollenin formation. The callose wall is paramount for pollen development, consists of β-1,3-linked glucose residues, and is first synthesized by pollen mother cells (PMCs) during meiosis initiation (Xie et al., 2012; Shi et al., 2016). The perimetric callose wall controls the integration of microspores and tetrad rupture during microsporogenesis (Richmond and Somerville, 2001; Verma and Hong, 2001; Nishikawa et al., 2005). In Arabidopsis, several callose synthase (CalS) or glucan synthase–like (GSL) genes, namely, *CalS5* (*GSL2*), *CalS7* (*GSL7*), *CalS9* (*GSL10*), *CalS10* (*GSL8*), *CalS11* (*GSL1*), and *CalS12* (*GSL5*) are reported for the regulation of different biological processes and adoption of stressed conditions (Xie et al., 2012). Among them, *CalS5* is specifically involved in the synthesis and functioning of the surrounding callose wall of the pollen (Dong et al., 2005; Nishikawa et al., 2005).

The four lobes of rice anther are attached with the same median of vascular and connective tissues. Each lobe has central microsporocytes and is surrounded by four layers of somatic cells, such as the outer epidermis, endothecium, middle layer, and innermost tapetum (Sun et al., 2018). The tapetal cell layer is adjacent to the anther locule, and provides essential nutrients for the microspores and secretes enzymes, such as β-1,3-glucanase (callase), which disintegrate the callose wall to release the young haploid microspores (Piffanelli et al., 1998; Li et al., 2006). Tapetum development and degeneration are well-organized processes, and disturbance usually leads to male sterility (Parish and Li, 2010). Timely tapetal cell programmed cell death (PCD) is thought to be the main stimulant for tapetum degeneration. Tapetal PCD is also associated with the degradation of callose and the formation of primexine on the microspore's surface, which is the first step in pollen cell wall formation (Ariizumi and Toriyama, 2011). In contrast to animals, during the plant reproduction process, PCD is relatively more beneficial for cell breakage and release of un-useful cellular constituents into useful functional components of the cell (Wu and Cheung, 2000). In microsporogenesis, proper tapetum development and differentiation is essential for pollen fertility, while in meiosis and megagametogenesis, proper tapetal PCD is vital for pollen fertility. Several genes are identified for tapetal PCD and pollen development, including persistent tapetal cell 2 (*PTC2*), helix–loop–helix (*bHLH*) transcription factors, tapetum degeneration retardation (*TDR*, *bHLH5*), undeveloped tapetum 1 (*UDT1*, *bHLH164*), eternal tapetum 1/delayed tapetum degeneration (*EAT1*/*DTD*, *bHLH141*), TDR-interacting protein 2 (*TIP2*, *bHLH142*), MYB family transcription factor *GAMYB*, PHD-finger protein persistent tapetal cell 1 (*PTC1*), TGA transcription factor *OsTGA10*, glycerol-3-phosphate acyltransferase 3 (*OsGPAT3*), and degenerated panicle and partial sterility 1 (*DPS1*) (Sun et al., 2018; Uzair et al., 2020; Zafar et al., 2020). Specifically, *TDR*, *TIP2*, *EAT1*, *PTC2*, and *OSGPAT3-2* exhibit delayed PCD, leading to an abnormal pollen wall and eventually sterile phenotype (Sun et al., 2018; Yang et al., 2019a; Uzair et al., 2020).

This study reported novel RNA recognition motifs containing protein, named *DCET1* (defective in callose, exine and tapetum1). *Dcet1* mutant showed small whitish anthers with irregular outer and inner surfaces having sterile pollens due to defective callose biosynthesis and degeneration, improper pollen exine patterning, and delayed tapetal PCD. Additionally, *DCET1* showed higher expression in anthers, specifically at stages 6–8 of anther development, confirming its involvement in callose and pollen wall development. Our results provide the first evidence of the RRM/RBD family protein *DCET1* in rice male sterility through defective callose, exine wall, and tapetal cell PCD, which justifies its role in hybrid rice breeding. Also, being the closest homolog of *Arabidopsis* CID-like proteins of mainly unknown functions, *DCET1* receives key importance for predicting and characterizing RBDs in general and CIDs in specific in all eukaryotes.

MATERIALS AND METHODS

Plant Materials

The *dcet1* mutant was identified from an ethyl methanesulfonate (EMS)–induced mutant library of an indica rice cultivar Zhonghui 8015. The complete male-sterile mutant *dcet1* was crossed with the wild-type (WT) and 02428 (sp. japonica), respectively. F_1 heterozygous plants were self-pollinated to obtain BC_1F_2 and F_2 populations for genetic analysis and mapping. All the plants were grown in paddy fields of the China National Rice Research Institute Hangzhou, Zhejiang Province, and Lingshui, Hainan Province, China.

Phenotypic Observation of WT and *dcet1* Mutant

The phenotypes of the whole WT and *dcet1* mutant lines were photographed using Nikon HB-40, Japan. The reproductive organs were captured with a Carl Zeiss Stereo Lumar V12 stereo fluorescence stereomicroscope (Markku Saari, Jena, Germany). To assess the pollen viability, mature anthers were collected from the WT and *dcet1* mutant and stained with 1.2% I_2-KI solution. The samples were observed and imaged using a Leica DM2500 microscope.

Cytological Observation

For semi-thin sections, spikelets of both WT and *dcet1* at different developmental stages were collected and fixed in a formalin–acetic acid–alcohol (FAA) solution (1:1:18) followed by dehydration using a graded series of 50–100% ethanol as described previously (Sun et al., 2018). After dehydration, the samples were embedded using Technovit glycol methacrylate 7100 resin (Heraeus, Kulzer, Germany), and polymerized at 50°C. Semi-thin sections of 2 μm were cut using an RM2265 rotary microtome (Leica, Germany) and stained with 0.1% (w/v) toluidine blue. The stained sections were observed and imaged using the Leica DM2000 microscope.

For transmission electron microscopy (TEM) analysis, the spikelets of WT and *dcet1* were pre-fixed in 2.5% glutaraldehyde in phosphate buffer (0.1M, pH7) overnight at 4°C. The samples were then rinsed thrice using phosphate-buffered saline (PBS; 0.1M, pH = 7.2) for 15 min at each step. The specimens were then post-fixed for 1.5 h with 1% OsO_4 in phosphate buffer and rinsed three more times as before, followed by dehydration using ethanol series (30–100%). After dehydration, the samples were infiltrated in a 1:1 mixture of absolute acetone and the final Spurr's resin for 1 h at room temperature, followed by a 1:3 mixture for 3 h, and the final Spurr's resin overnight. Following infiltration, the specimens were placed in Eppendorf contained Spurr's resin and heated at 70°C for more than 9 h. Ultra-thin sections were cut in a LEICA EM UC7 ultratome (Germany) and double-stained with 2% uranyl acetate and 2.6% alkaline lead citrate aqueous solution for 5–10 min, respectively. The stained sections were pictured with a Hitachi Model H-7650 transmission electron microscope (Japan) at the Center of Electron Microscopy, Zhejiang University (Hangzhou, China).

Scanning electron microscopy (SEM) analysis was performed by fixing the mature anthers of WT and mutant with 2.5% glutaraldehyde in 0.1 M sodium phosphate buffer at 4°C. The samples were then dehydrated with the same graded ethanol series (30–100%) as performed in TEM, and exchanged three times with isoamyl acetate. The fixed samples were then critical point-dried, gold-coated, and mounted. The samples were observed and photographed using a scanning electron microscope (Hitachi TM-1000) at the Center of Electron Microscopy, Zhejiang University (Hangzhou, China) with an accelerating voltage of 10 or 15 kV.

TUNEL Analysis

The terminal deoxynucleotidyl transferase-mediated dUTP nick-end labeling (TUNEL) assay was carried out to scrutinize the tapetal cell PCD. Paraffin sections of the WT and *dcet1* anthers at different developmental stages were prepared and processed as described previously (Chang et al., 2014). The suitable paraffin sections were dewaxed in xylene and rehydrated in a graded ethanol series. The TUNEL assay was performed using the TUNEL Bright Green Apoptosis Detection Kit A-112 (Vazyme Biotech Co., Ltd.) according to the supplier's instructions with little modifications. The TUNEL (green fluorescence, at a wavelength of 488 nm) and DAPI signals (blue fluorescence, at a wavelength of 405 nm) were observed and imaged using a fluorescence confocal scanner microscope (ZEISS LSM 700, Jena, Germany).

Meiotic Chromosome Preparation

For meiotic observation, 4′,6-diamidino-phenylindole (DAPI) staining was carried out by sampling the young meiosis stage spikelets of both WT and *OSRRMS1*. The samples were fixed in Carnoy's solution (ethanol: glacial acetic, 3:1) and processed according to the previous prescription (Wu et al., 2015).

Aniline Blue Staining for Callose

For callose staining of WT and *dcet1*, anthers at different developmental stages were collected and stained through 0.1% aniline blue for 15 min at 4°C. The samples were prepared and processed as described previously (Lu et al., 2014). The presence of callose (blue-yellow fluorescence) was visualized under an excitation wavelength of 420 nm and emission wavelength of 530 nm using a fluorescence confocal microscope. The same setting was used across all samples.

DCET1-Promoter-GUS Assay

GUS staining was observed by sampling fresh tissues and anthers at different developmental stages from transgenic plants expressing the *DCET1*-promoter-GUS construct and stained as described previously (Zou et al., 2017a). The samples were washed in ethanol and pictured by using a light microscope.

Gene Mapping of *DCET1*

For genetic mapping, the *dcet1* mutant (sp. indica) was crossed with 02428 (sp. japonica) to get F_1, which was further grown to derive the F_2 population. A total of 1,050 F_2 male-sterile plants were used for genetic mapping. For fine mapping of *DCET1*, SSR

(simple sequence repeat) and InDel (insertion–deletion) markers were developed based on polymorphism between parents. After fine mapping and cloning of the targeted *DCET1*, 30 homozygous sterile plant DNA of the *dcet1* F_2 population were used for MutMap to verify the mapping results further. The MutMap technique was employed according to the previous prescription (Abe et al., 2012).

Sequence and Phylogenetic Analysis

The amino acid full-length sequence of *DCET1* and selected *Arabidopsis* CIDs and other 17 gene sequences were retrieved with BLASTP (http://www.ncbi.nlm.nih.gov/), and conserved domains were searched in CD-search (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Because *DCET1* belongs to a CID-like protein, first *Arabidopsis* CID group-D genes and second 17 other amino acid sequences were selected for multiple sequence alignments and evolutionary analysis. Sequence alignments were carried out using Clustal Omega (https://www.ebi.ac.uk/Tools/), and the Neighbor-Joining method (Saitou and Nei, 1987). Resultant alignments were used to construct the maximum likelihood trees by using Mega X with 1000 bootstrap replicates (Kumar et al., 2018).

Vector Construction

For the complementary vector, an entire 10,249 bp region was amplified from the WT (ZH8015) genomic DNA; consisting of the entire ORF of LOC_Os08g02330, 2,943 bp upstream sequence and 1,043 bp downstream sequence by using the primers *DCET1*-CF and *DCET1*-CR (**Supplementary Table S1**). The fragment was then cloned into binary vector pCAMBIA1300 using the In-Fusion HD Cloning Kit (Takara Bio USA Inc., Mountain View, CA, United States).

To gain further insights into the role of LOC_Os08g02330 in rice male sterility, we designed a 23 bp target in the 2nd exon within this gene (**Supplementary Table S1**). Vector construction was carried out by using the CRISPR/Cas9 genome editing system in ZH8015 indica background to investigate its function as previously described (Zou et al., 2017b), with some modifications. We also observed the CRISPR/Cas9 transgenic plants for the target site mutation through direct or cloned sequencing of the PCR products by using site-specific primers (**Supplementary Table S1**).

For the GUS vector, a 3,074 bp promoter of *DCET1* was amplified from the WT genomic DNA, using the primers *DCET1*-GF and *DCET1*-GR (**Supplementary Table S1**). The fragment was cloned into the binary vector pCAMBIA1305 with BamH1 and Nco1 restriction sites using an In-Fusion HD Cloning Kit (Takara Bio USA Inc., Mountain View, CA, United States).

RNA Isolation, cDNA Synthesis, and Real-Time Quantitative Reverse Transcription PCR

Total RNA was isolated from different rice tissues using the RNAprep Pure Plant Kit (Tiangen, Beijing, China), according to the manufacturer's instructions. RNA concentration was adjusted and reverse-transcribed into cDNA using the ReverTra Ace® qPRT-PCR Master Mix with gDNA Remover (Toyobo, Japan). For the reaction and qRT-PCR, LightCycler 480 (Roche, Germany) using LightCycler® 480 SYBR® Green I Master Mix (Roche, United States) was used according to the standard manufacturer's recommended program and guidelines. In all samples, ubiquitin is used as an internal control, and each reaction was repeated three times (Zhang et al., 2017).

Analysis of Anther Wax and Cutin Monomers

For anther cuticular lipids, WT and *dcet1* anthers at stage 12 were collected and immediately frozen in liquid nitrogen. The wax and cutin contents were determined and analyzed as previously described (Shi et al., 2011; Zhu et al., 2013). Statistical analysis was carried out by using Student's *t*-test.

RESULTS

Characterization of the *DCET1* Mutant

Mutant *dcet1* was screened from EMS treatment of an *Indica* background cultivar Zhonghui8015 (ZH8015). All vegetative and agronomic traits of the *dcet1* mutant, including plant height and general floral morphology, were the same as that of wild-type ZH8015 (WT), except the anthers (**Figures 1A–C**). The *dcet1* anthers were weaker, smaller and whitish-yellow as than the WT having dark yellow and normal-sized anthers (**Figures 1B–D**). After staining with I_2-KI, the mutant *dcet1* was found to be completely sterile with whitish-yellow and hill-shaped pollen grains, but fertile WT pollen grains were dark-black and round-shaped (**Figures 1E,F**). Additionally, the extrusion of stigma was also observed in the *dcet1* mutant plants (**Figure 1G**). The BC_1F_1 plants of the mutant cross with WT were completely fertile (**Figure 1H**), and F_2 plants showed a fertile/sterile segregating ratio of 3:1 (168 fertile and 58 sterile plants), confirming the recessive mutant nature of *dcet1*. These results also ensure that the female part is fertile in the *dcet1* mutant.

Male Reproductive Defects of the *dcet1* Mutant

To explore the reproductive system defects of the *dcet1* mutant, we investigated the cytological mechanism of WT and *dcet1*. The anther development was divided into 14 stages based on the length as prescribed previously (Zhang et al., 2011). The semi-thin transverse sections of the different developmental stages from the anthers of *dcet1* were compared with those of WT (**Figures 2A,P**). No obvious defects were traced till the formation of microspore mother cells (MMCs) from the secondary sporogenous cells (**Figures 2A,I**). Typically at this stage of anther development, the anther locule is encircled by the four-layered anther wall (Zhang et al., 2011). Prominent abnormalities were observed with the initiation of meiotic division stage 7, when the WT meiocytes were darkly stained, well-developed, and

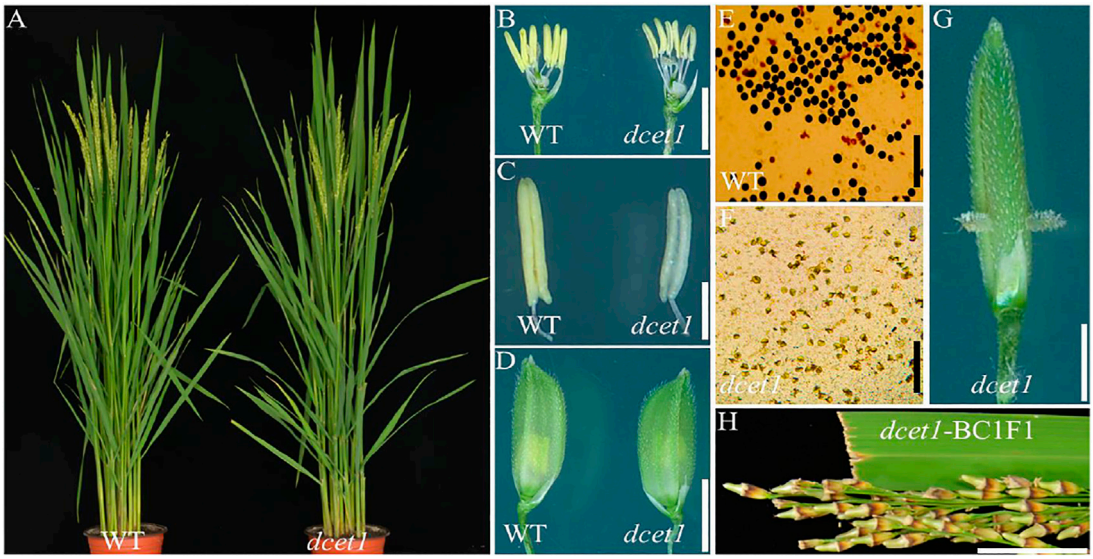


FIGURE 1 | Phenotypic comparison of wild-type (WT) and mutant (*dcet1*). Plants after heading **(A)**; Spikelets after the removal of palea and lemma **(B)**; Anthers at stage 13 **(C)**; Spikelets at the heading stage **(D)**; Pollen grains after staining with 1.2% I2-KI staining solution **(E,F)**; Spikelets of *dcet1* with stigma extrusion **(G)**; Panicle of *dcet1*-BC1F1 **(H)**. Scale bars = 20 mm in **(A)**, 1 mm in **(B,C)**, 2 mm in **(D,G)**, 100 μm in **(E,F)**, and 1 cm in **(H)**.

spherical-shaped, but *dcet1* mutant meiocytes were less stained, hollow, and variable-shaped, as described in **Figures 2B,J**. However, at the corresponding stage, the *dcet1* anthers produced morphologically normal four somatic layers of epidermis, endothecium, middle layer, and tapetum, like those of WT (**Figures 2B,J**). At stage 8a of anther development, the meiocytes undergo meiosis, and the anther initiates the tapetal cell programmed cell death (PCD). Meanwhile, the WT microspores were divided into dyads separated by the cell plate, and the tapetum became more vacuolated and condensed (**Figure 2C**). In contrast to the WT, in the *dcet1* mutant, the formation of dyads was completely irregular, the microspores changed into a conical and bursted shape, and the tapetum was less vacuolated and had not expanded properly (**Figures 2C,K**). At stage 8b, the WT meiocytes developed into dark-stained tetrads and the tapetum was further vacuolated, but the *dcet1* microspore burst out and became less stained with variable vacuoles inside (**Figures 2D,L**). In addition, the middle layer was nearly invisible in WT anthers, while in *dcet1*, it was still clearly apparent with negligible signs of degradation (**Figures 2D,L**). During stage 9, the WT microspores were released from the tetrad and evenly spread inside the anther locule (**Figure 2E**). However, the *dcet1* microspores were degenerated, shrinked, or over-broadened and looked unable to produce viable pollen grains (**Figure 2M**). Here the *dcet1* tapetum was lightly stained and not properly vacuolated as compared to that of WT, indicating abnormal PCD (**Figures 2E,M**). Furthermore, at stage 10, the WT-vacuolated and teeth-like microspores were almost attached with the tapetum leaving a circle-like space at the middle of locule; but the *dcet1* microspores were asymmetrical, non-vacuolated, and grass-like residues spread inside the anther locule (**Figures 2F,N**). Correspondingly, the WT tapetum was degenerated and hill-like, whereas the *dcet1* tapetum was still properly intact and exhibited the delayed nature of *dcet1* tapetal cell PCD (**Figures 2F,N**). Eventually, at mature stages of anther development, when the WT released the normal and viable pollen grains, the *dcet1* mutant produced shriveled and debrised pollen grains inside the anther locule with complete sterility (**Figures 2G,H,O,P**).

To further confirm the male reproductive and morphological defects of the *dcet1* mutant, we performed the SEM analysis of mature anthers (stage 13) for both WT and *dcet1*. In conformity with the abovementioned phenotypic results, the anthers were smaller in *dcet1*, having collapsed nanoridges on the anther outer surface than WT, typically having prominent and organized nanoridges on the anther cuticle outer surface (**Figures 3A,B,F,G**). More obvious differences were found on the inner surface of the anther between the WT and *dcet1* mutant. The inner surface of the WT anther was covered by evenly distributed Ubisch bodies, while the *dcet1* anther inner surface was almost missing in the Ubisch bodies (**Figures 3C,H**). The WT plant pollen grain was round globe-shaped with an organized lobed exine surface; in contrast to WT, the pollen of *dcet1* was shrunken, with irregular and softened outer exine (**Figures 3D,E,I,J**). These results indicated the abnormalities of anther and pollen and their surfaces in the *dcet1* mutant.

To further elaborate the differences between WT and *dcet1*, we examined the different stages of anther development of both WT and *dcet1* mutant through transmission electron microscopy (TEM). No obvious changes were observed until the sporogenous cell divides into secondary sporogenous cells and until the formation of anther somatic layers (endothecium, middle layer, and and tapetum) between the WT and *dcet1*. However, a stronger callose layer was seen in WT, to that of mutant *dcet1* during this period (**Figures 4A,B**). But the most distinguished abnormalities were observed at stage 7 and onward,

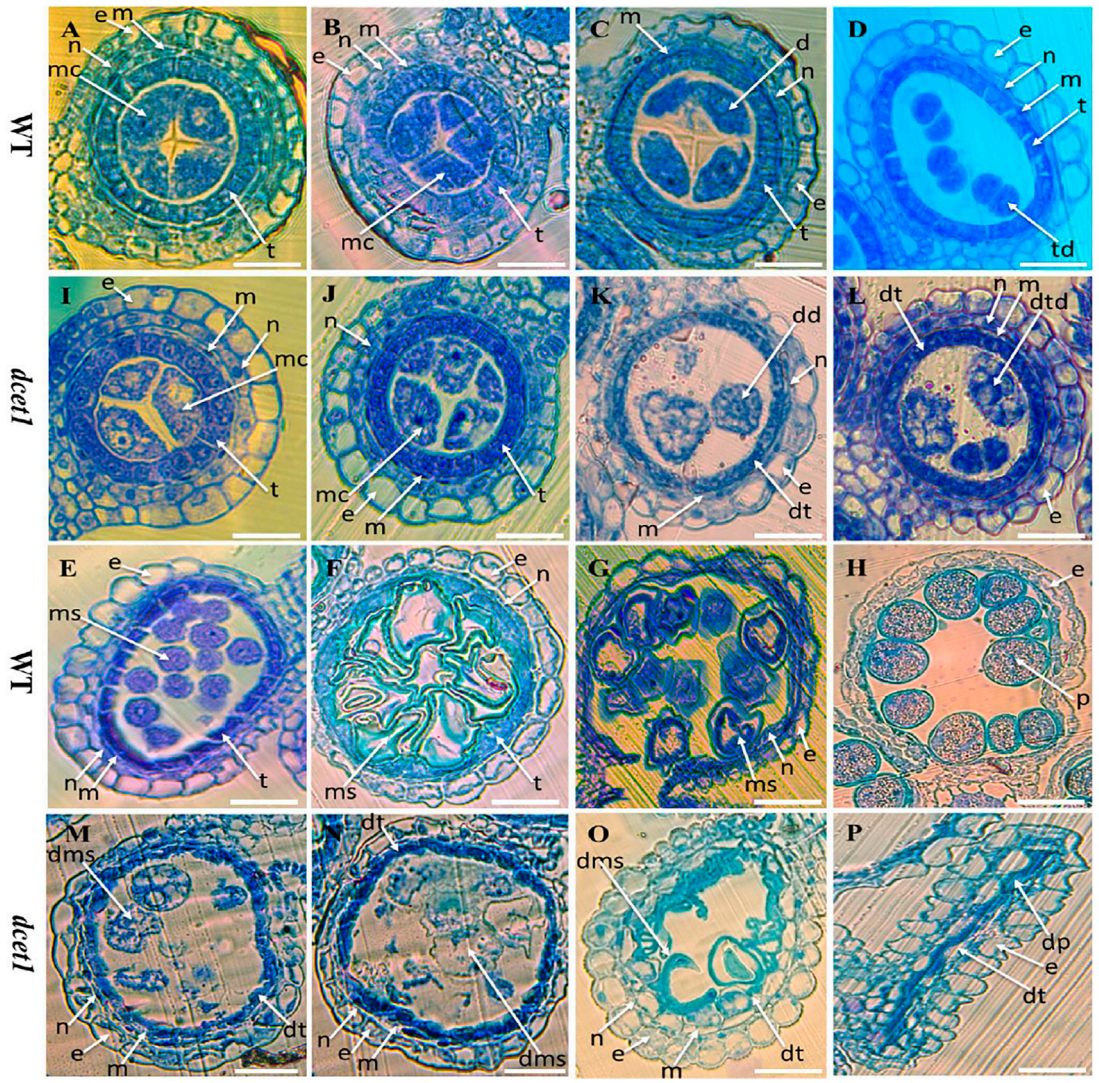


FIGURE 2 | Semi-section analysis of wild-type (WT) and mutant (*dcet1*). Anther sections of WT **(A–H)** and *dcet1* **(I–P)** from stages 6–10 and 12 and 13 of anther development, respectively. e, *epidermis*; n, endothecium; m, middle layer; t, tapetum; mc, meiocyte; dt, defective tapetum; d, dyad; dd, defective dyad; td, tetrad; dtd, defective tetrad; ms, microspore; dms, defective microspore; p, pollen; dp, defective pollen. Scale bars = 15 μm.

when the microspores start differentiating from each other through callose wall formation. At this stage, an off-white–colored callose layer could be observed in the WT anther locule differentiating the microspores, while in the *dcet1* mutant; this callose layer was not clearly visible, showing the male reproductive defective nature of *dcet1* before proper initiation of meiosis (**Figures 4C,D**). In addition to callose layer defects between WT and mutant *dcet1* at the corresponding stage, the WT meiocytes were electron-densely stained and compact in shape; in contrast, the *dcet1* meiocytes were thinly stained having variable vacuoles and hollow cavities inside (**Figures 4C,D**). In semi-thin transverse section studies, the same defects were observed at this stage in *dcet1* anthers. At early stage 8a, the callose wall surrounded the meiocytes in WT, while in *dcet1*, the callose wall was missing or degenerated, favoring the abnormalities (**Figures 4E,F**). Similarly the surface was irregularly stained in *dcet1* as compared to WT (**Figures 4E,F**). At later stage 8b, the WT tetrads were generated and separated by the thick callose layer, while in *dcet1* no organized callose wall was observed; only an irregularly curled thin line could be seen surrounding the defective dyads (**Figures 4G,H**). TEM analysis also confirmed the delayed tapetum degeneration and programmed cell death (PCD) in the *dcet1* mutant. The *dcet1* anthers did not differ from those of WT for tapetum development and other somatic layers till the dyad formation stage 8a (**Figures 4I,J**). At the tetrad stage 8b, the WT tapetum started to degrade through proper PCD and underwent extended vacuolation (**Figure 4K**). In contrast to WT, the *dcet1* mutant tapetum was weakly stained, covered with a thick cell wall, and produced decreased vacuolation (**Figures 4K,L**). At stage 9,

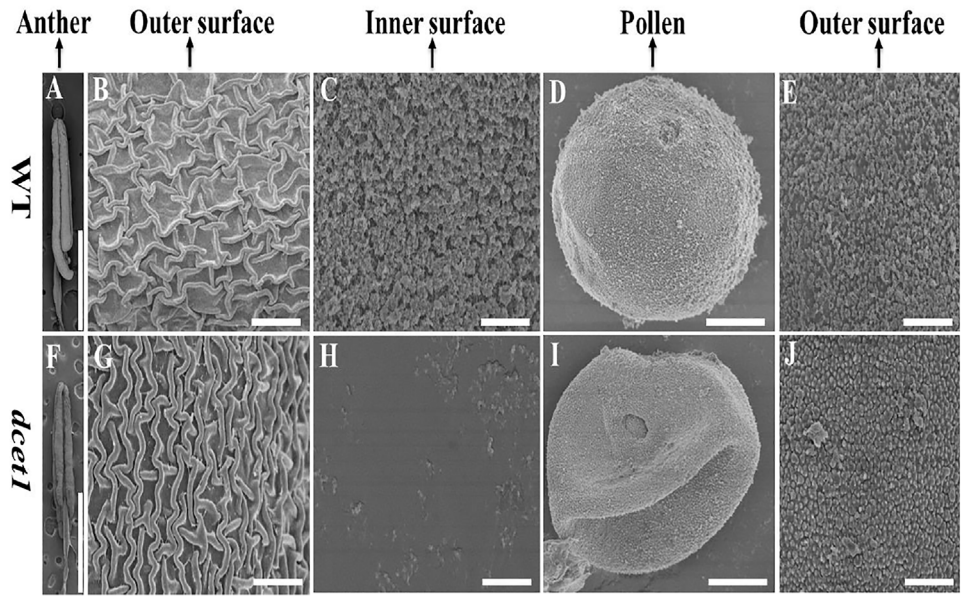


FIGURE 3 | Scanning electron microscopy (SEM) analysis of the surfaces of anthers and pollen grains in the wild-type (WT) and mutant (*dcet1*) **(A–E)** Represents WT and **(F–J)** represents *dcet1*. Scale bars = 100 μm in **(A,F)**, 2 μm in **(B,C,G,H,E,J)**, and 1 μm in **(D,I)**.

the WT tapetum continued to degrade and looked thinner, while the *dcet1* mutant tapetum was devacuolated and formed a bubble-like structures spread over the tapetum surface (**Figures 4M,N**). At stage 10, both the WT and *dcet1* mutant formed normal Ubisch bodies; however, the *dcet1* tapetum was wider and darkly stained with negligible vacuolation than that of WT (**Figures 4O,P**). At stage 11, when the WT tapetum was completely dissolved, the *dcet1* tapetum could be easily observed, and even at the mature stage, the thinner mutant tapetum could be seen (**Figures 4Q–T**).

Furthermore, we had also observed the pollen exine patterning in both WT and *dcet1* mutant through TEM analysis. The results showed obvious changes between WT and *dcet1* mutant at the tetrad stage (8b). At this stage, we had observed the deposition of sporopollenin precursors on the primexine surface in the shape of black dots, while no such deposition was seen in the *dcet1* mutant (**Figures 5A,B**). At stage 9 in WT, a well-organized protectum was observed in the form of darkly stained protrusions on the microspore outer surface; in contrast, the *dcet1* microspore surface was having a detached and irregularly distributed protectum (**Figures 5E,F**). At stage 10, the exine of the WT microspore wall was thickly stained, having a regular two-layered surface with typical exine patterning, while the exine of the *dcet1* microspore was thinly stained with deformed bacula, defective tectum, and irregular nexine (**Figures 5C,D**). Furthermore, at the mature stage (stage 13), in contrast to the spherical pollen grains of WT which were having normal exine with well-formed bacula, protruded tectum, and normal nexine, the exine of *dcet1* pollen was malformed with disintegrated bacula, defective tectum, and misshaped nexine (**Figures 5G,H**). TEM results demonstrated the abnormal and abortive nature of *dcet1* pollen grains through defective callose regulation, improper exine patterning, and delayed tapetal cell PCD.

Callose Wall During Microsporogenesis Was Defective in *dcet1*

During TEM analysis, the abnormalities in callose metabolism were observed in the *dcet1* mutant. Thus we performed the aniline blue staining of both WT and *dcet1* anther sections during various stages of anther development. Both WT and *dcet1* mutant meiocytes were nearly normal for pre-callose wall synthesis; however, the weakly signals were observed in *dcet1* anthers compared to WT at stage 5 of anther development (**Figures 6A,D**). This specified that callose synthesis was prominently affected in the *dcet1* mutant at the start (**Figures 6A,D**). Moreover, no other obvious changes were observed in the *dcet1* mutant than WT until stage 6 of anther development, except the weakest callose layer signals and its irregular distribution over the *dcet1* mutant microsporocyte surface within the anther locule (**Figures 6B,E**). At anther stage 7, WT microspores were separated by a thickened and well-formed callose wall (**Figure 6C**). Meanwhile in the *dcet1* anther, some amount of callose was accumulated on the surface of the four somatic layers of the anther wall, and the meiocytes had such weaker signals for callose wall accumulation that only a few meiocytes could be hardly recognized (**Figure 6F**). At stage 8a, when the WT meiocytes developed dyads, they were separately well-surrounded by the callose wall, but *dcet1* mutant anthers almost had no callose wall, and a thinner callose whitish layer was observed spread over the whole surface of the meiocytes and anther wall (**Figures 6G,J**). At stage 8b of anther development, WT meiocytes were divided into tetrads through meiosis, properly encircled from the outside as well as within the tetrads separately,

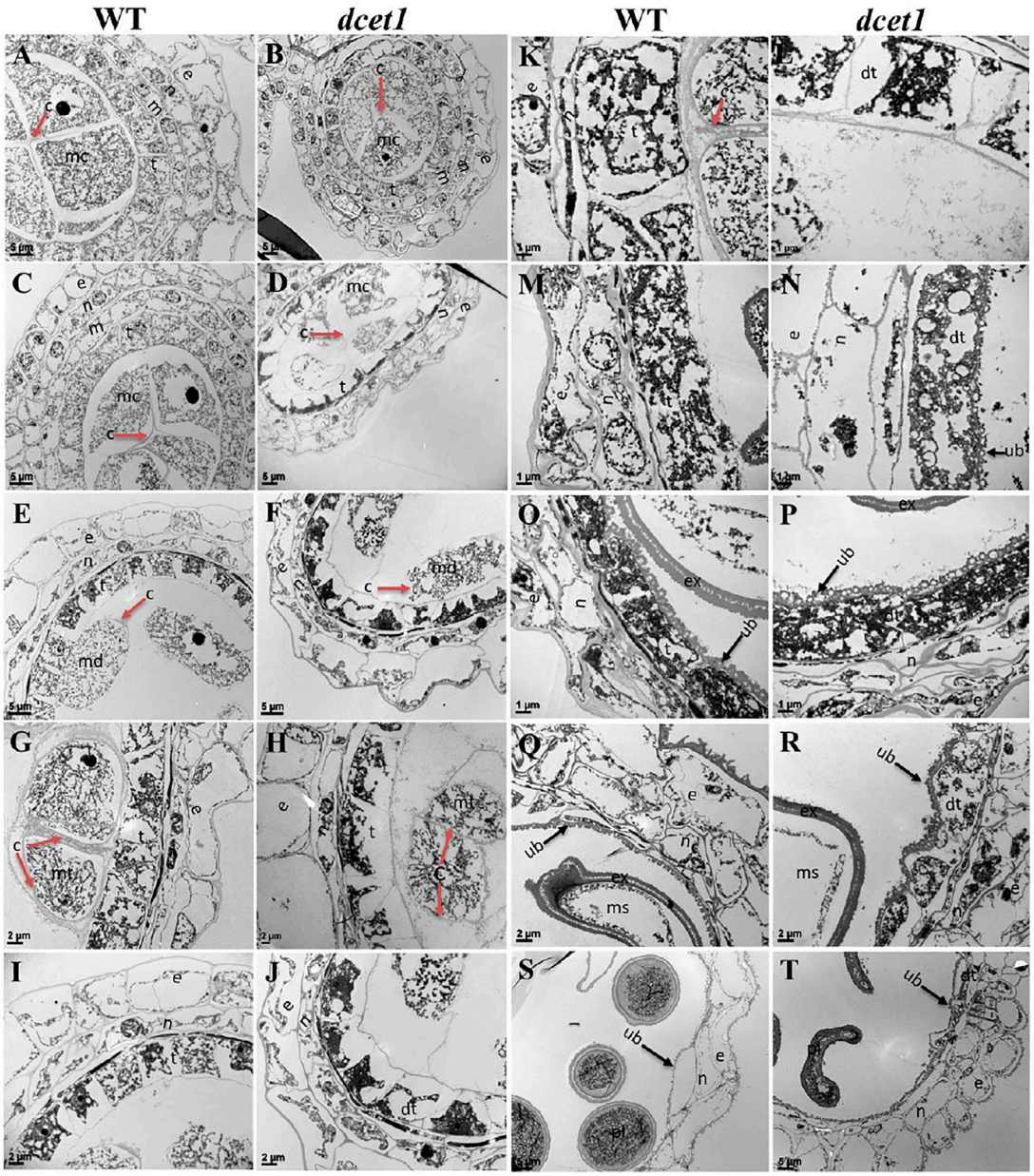


FIGURE 4 | TEM analysis of the anther callose wall and tapetum in the wild-type (WT) and mutant (*dcet1*). Comparison of WT and *dcet1* for the callose wall **(A–H)**; Comparison of WT and *dcet1* for the tapetum **(I–T)**. Red and black arrows are used for describing the callose layer and tapetum Ubisch bodies, respectively. c, callose; mc, mother cell; md, dyad; mt, tetrad; ms, microspore, dms, defective microspore; pl, pollen; dpl, defective pollen; e, *epidermis*; n, endothecium; m, middle layer; t, tapetum; dt, defective tapetum; ub, Ubisch bodies. Scale bars = 5 µm **(A–F,S,T)**, 2 µm **(G–J,Q,R)**, and 1 µm **(K–P)**.

but the *dcet1* mutant meiocytes underwent no such division due to impaired and degenerated callose wall synthesis as presented in **Figures 6H,K**. After meiosis at stage 9, when the WT microspores were properly released due to timely degeneration of the callose wall through callases and initiation of the pollen wall, the *dcet1* microspores still had a thinner and defective callose wall, and hence abnormal pollen grains were produced (**Figures 6I,L**). These results suggested that the synthesis and degradation of the callose wall was a continuous process in the WT. However, in *dcet1*, a thinner or defective callose wall staining was observed, and even after the tetrad stage, some amount of callose was seen spread over the anther locule, resulting in abnormal pollen grains and causing fusion among sibling microspores. These findings indicated that callose biosynthesis and dissolution were defective and abnormal in the *dcet1* mutant.

Meiosis in the *dcet1* Mutant

To check whether meiosis in the *dcet1* mutant was defective or not with the defects in callose accumulation, we carried out

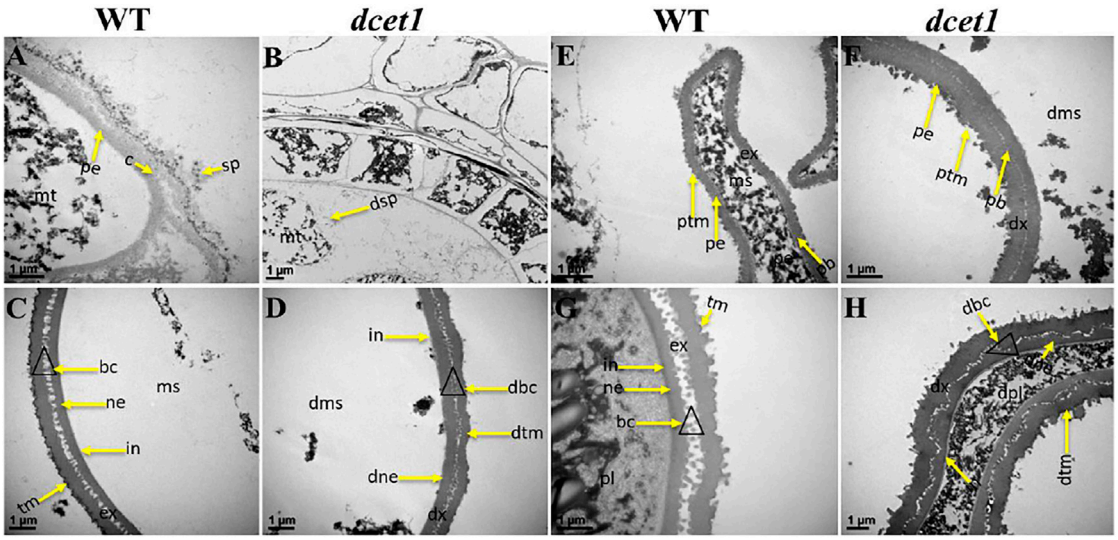


FIGURE 5 | TEM analysis of pollen wall development in the wild-type (WT) and mutant (*dcet1*). Comparison of WT **(A,C,E,G)** and *dcet1* **(B,D,F,H)**. mt, tetrad; ms, microspore, dms, defective microspore; pl, pollen; dpl, defective pollen; pe, primexine; sp, sporopollenin; dsp, defective sporopollenin; ex, exine; dx, defective exine; ne, nexine; dne, defective nexine; pb, probocula; bc, bacula; dbc, defective bacula; tm, tectum; ptm, protectum; dtm, defective tectum; tr, tryphine; dtr, defective tryphine; in, intine. Scale bars = 1 µm.

aniline blue and DAPI stainings for callose and meiosis at different anther development stages. In the WT and mutant *dcet1*, no changes were observed in meiosis from the start till the production of tetrads. At diakinesis, both WT and *dcet1* condensed normally and produced paired chromosomes into 12 bivalents and were aligned on the equatorial plate during metaphase I (**Supplementary Figures S1A–C,E–G**). As in WT, anaphase I was also perceived to be standard for *dcet1* by migrating the homologous chromosomes into opposite poles, and finally dyads were formed at the completion of meiosis I (**Supplementary Figures S1D,H,I,M**). During meiosis II, parallel to the WT, in the *dcet1* chromosome, the sister chromatids were separated normally and produced normal tetrads (**Supplementary Figures S1L,P**). Although both WT and *dcet1* microsporocytes appeared normal for meiosis, the callose accumulation was not normal, and a thinner and scattered callose layer was observed in the *dcet1* mutant than WT (**Figures 6A–L**; **Supplementary Figures S1A–P**). At meiosis stage 8a and b, the microspores could be observed separately and showed a normal nuclear division of the mutant male gametophyte, although the surrounding callose wall is invisible in the *dcet1* mutant (**Figures 6G,H,J,K**; **Supplementary Figures S1A–P**). However, most of the *dcet1* young pollens were observed to be swollen and bursted due to the abnormal callose wall (**Figures 6I,L**), irrespective of normal meiosis, confirming the key role of callose for pollen maturation and viability.

Delayed DNA Fragmentation in *dcet1* Tapetal Cells

The *dcet1* mutation showed the defective degradation of somatic walls during cytological analysis. Thus, we assumed that the abortive anthers of the *dcet1* mutant would have abnormal PCD. Therefore, we performed the terminal deoxynucleotidyl transferase-mediated dUTP nick-end labeling (TUNEL) assay in WT and *dcet1* mutant at different stages of anther development as previously described (Uzair et al., 2020). Before stage 8 of anther development, both the WT and *dcet1* mutant had no DNA fragmentation signals (**Figures 7A,E**). DNA fragmentation signals were first observed at meiotic stage 8 in the WT tapetal cells, but no such signals were visualized during this stage in the *dcet1* mutant (**Figures 7B,F**). At stage 9, when the microspores were released, signals for DNA fragmentation appeared in the *dcet1* mutant and reached its peak till the later stages of anther development (**Figures 7G,H**). On the other hand, the DNA fragmentation in the WT occurred in a typical manner and ended at stage 11 (**Figures 7C,D**). These results further confirmed the delayed PCD of the tapetal cells in the *dcet1* mutant.

Analysis of Anther Cuticular Lipids

TEM analysis had revealed that the *dcet1* pollen wall was comparatively thick at the later stages of anther development (**Figures 5G,H**). Simultaneously, SEM results indicated a subtle change in the anther outer surface between *dcet1* and WT (**Figures 3B,G**), encouraging the comprehensive analysis of anther cuticular lipids. The cuticular wax constituents and cutin monomers in both WT and *dcet1* anthers at stage 12 were measured by gas chromatography-mass spectrometry (GC-MS) and gas chromatography-flame ionization detection (GC-FID). The amounts per unit area were determined by plotting the calculated anther surface area against the dry weight of each corresponding sample (**Figure 8A**). The results showed that the total amount of cutin monomers on the *dcet1*

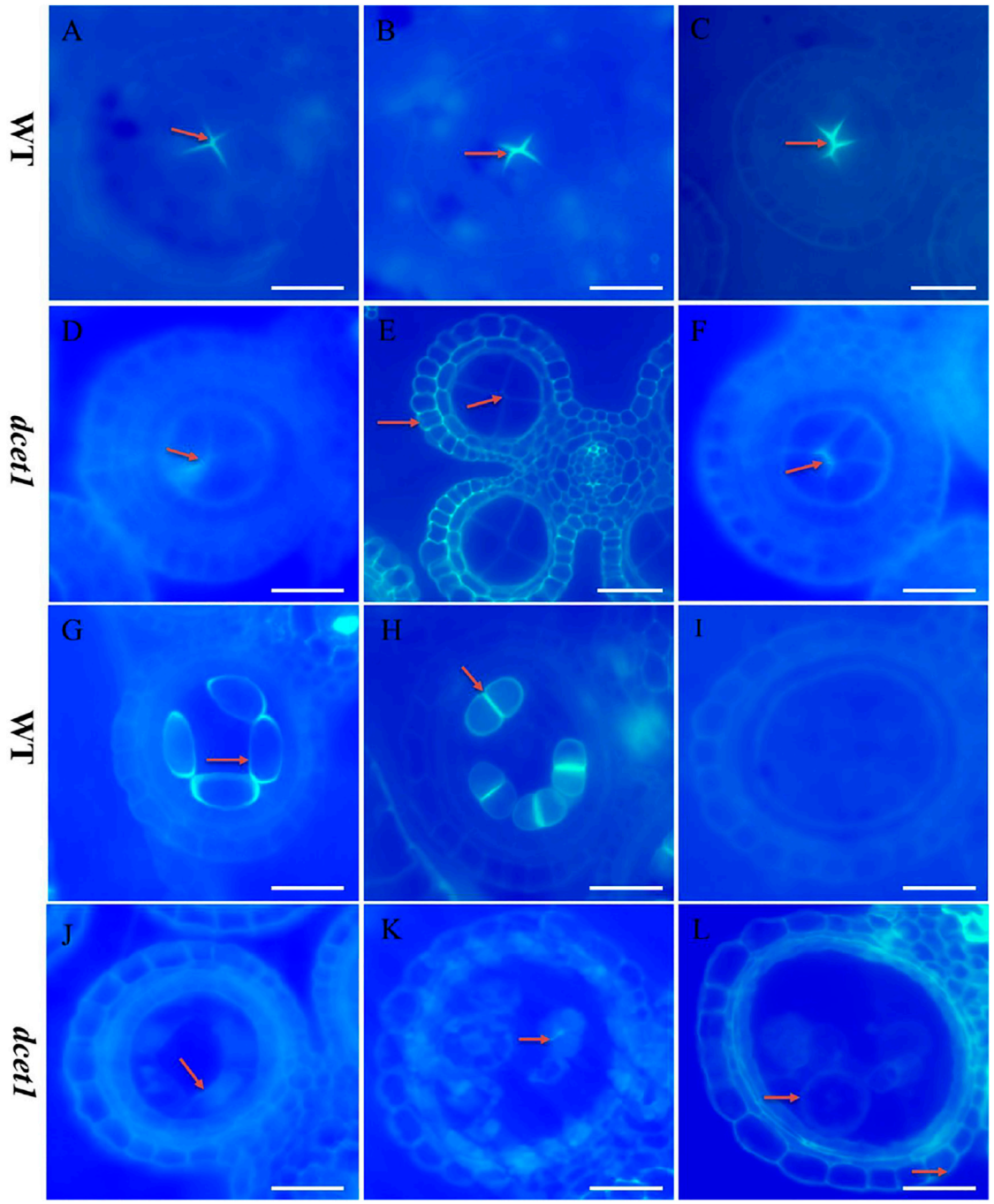


FIGURE 6 | Anther callose layer analysis of wild-type (WT), and mutant (*dcet1*). Aniline blue staining of WT **(A–C,G–I)** and *dcet1* **(D–F, J–L)** for anther development stages 5, 6, 7, 8a, 8b, and 9, respectively, Red arrows highlight the callose layer. Scale bars = 20 μm.

anther was 0.356 μg mm^{-2} compared to 0.499 μg mm^{-2} on the WT anther surface, which corresponds to a highly significant reduction in total cutin (**Figures 7B,C**). The cis and trans-ferulic acids, the hydroxy fatty acid C23:0 2HFA, the unhydroxylated fatty acid C18:2 FA, and ω-hydroxy fatty acids, such as C16:0 ω-HFA, C18:1 ω HFA, C18:2 ω-HFA (2), cis-9,10 epoxy C18:0 ω-HFA, and chlorohydrin of 9,10expoxy C18 ω-HFA, were dominant cutin monomers, and were found at significantly lower levels in *dcet1* anthers (**Figure 8C**). Most of the unhydroxylated fatty acids, such as C18:1 FA, C18:0 FA, C20:0FA, C16:0 FA, and C18:3 FA, and hydroxy fatty acids, such as C20:0 2HFA, C23:0 2HFA, C24:0 2HFA, C25:0 2HFA, and C26:0 2HFA, were found in very less or in non-significant amount (**Figure 8C**). The total amount of wax in the *dcet1* anther was

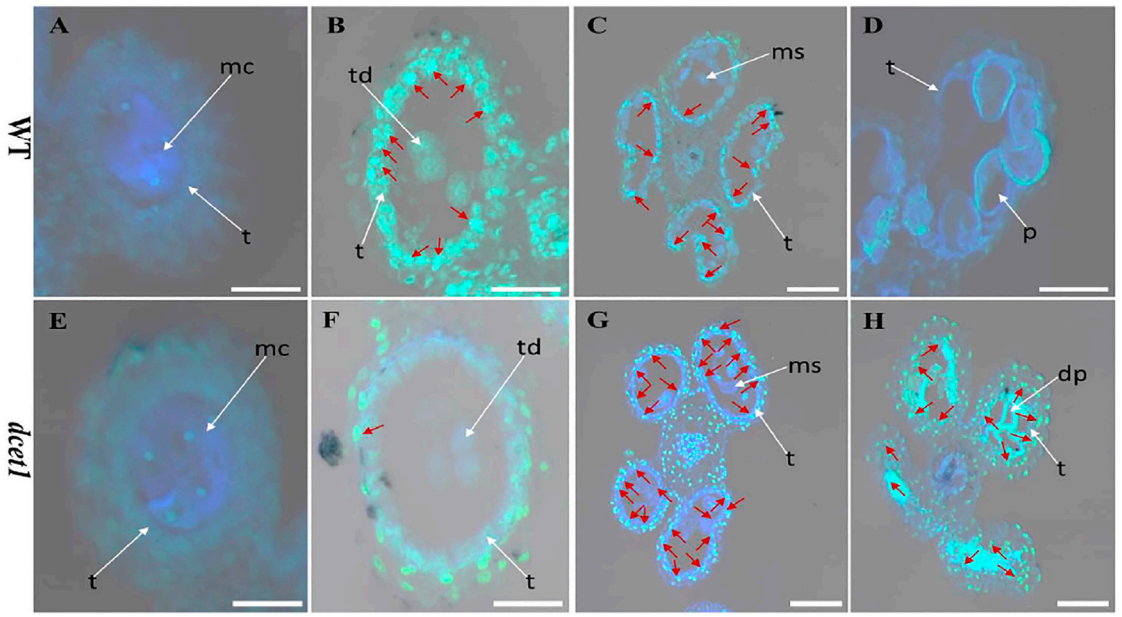


FIGURE 7 | TUNEL analysis of wild-type (WT) and mutant (*dcet1*). Anthers of WT **(A–D)**, and *dcet1* **(E–H)**, for pre-meiosis, meiosis, post-meiosis, and mature pollen stages of anther development, respectively. Red arrow highlights the strength of TUNEL signals. t, tapetum; mc, meiocyte; td, tetrad; ms, microspore; p, pollen; dp, defective pollen. Scale bars = 20 μm.

0.639 μg mm^{-2}, whereas the wild-type anther wax was 0.497 μg mm^{-2}, which corresponds to a highly significant increase in total wax content of the anther (*p-value* < 0.01) (**Figure 8D**). This increase of cuticular wax content in *dcet1* anthers was mainly attributed to the significant rise in alkene content, such as C27:1ALKENE, C29:1ALKENE, C31:1ALKENE, C33:1ALKENE, and a few alkanes such as C23:0ALK, C25:0ALK, and C35:1ALK (**Figure 8D**). The data also showed the decrease in almost all recorded fatty acids for *dcet1* anthers, while the other wax constituents are mainly non-significant (*p-value* < 0.01) (**Figure 8D**). The wax and cutin analysis of *dcet1* suggested that *DCET1* is involved in regulating the biosynthesis of lipidic compounds to form anther cuticle and pollen exine in rice, and the denser pollen exine at later stages might be due to the increase of wax constituents.

Cloning of *DCET1*

First, a map-based cloning technique was used to identify the gene responsible for the *dcet1* mutant phenotype. The F_2 population derived from the cross between the *dcet1* mutant and 02428 (sp. japonica) was used for genetic mapping. For linkage and primary mapping, we used 52 F_2 plants and eventually mapped the *DCET1* gene at 998-kb intervals between SSR markers named RM408 and RM3702 on chromosome 8 (**Figure 9A**). The region was further narrowed down to 84 kb by using 1,050 F_2 plants, between markers Indel8 and RM43 (**Figure 9B**), containing ten genes based on bioinformatics analysis (http://www.gramene.org/). All the ten potential candidate genes within this region were sequenced, and a single-bp substitution was identified from G (threonine-192) to A (isoleucine-192) in the fifth exon of the gene corresponding to LOC_Os08g02330 (http://rice.plantbiology.msu.edu/), also annotated as Os08t0116400-01 (http://rapdb.dna.affrc.go.jp/). This gene is predicted to be a CTC-interacting domain-11 (CID 11)–like gene**,** an RNA-binding protein having 10 exons, two domains named RRM1 and RRM2, and a C-terminal PAM2 conserved site (**Figure 9C**).

Second, the MutMap cloning approach was also used to re-confirm the LOC_Os08g02330 responsible for the *dcet1* mutant according to the previous description (Abe et al., 2012). For MutMap ,30 F_2 *dcet1* mutant plants were used, and the results identified four ORFs annotated as MH08t0020700-01, MH08t0401600-01, MH09t0093100-01, and MH01t0044600-01 (http://rice.hzau.edu.cn/rice/), among which MH08t0020700-01 was lying in the abovementioned 84-kb fine-mapped region (**Figure 9D**; **Supplementary Figures S3A,B**). All the four ORFs were sequenced and aligned with the reference sequence. The mutation was found only in MH08t0020700-01, which was the same fine-mapped gene also annotated as Os08t0116400-01 (http://rapdb.dna.affrc.go.jp/).

Phylogenetic Analysis of *DCET1*

Phylogenetic analysis showed that *DCET1* belongs to *Arabidopsis* CID11–like proteins having two RRM domains and a conserved motif at the C-terminus, designated as PAM2 (PABP-interacting motif 2), and are suggested to encode highly related RNA-binding proteins (RBPs). Previously in *Arabidopsis*, 13 CIDs were divided into four groups: group A consists of two genes CID1 and CID2, group B also contains two genes CID3 and CID4, group C consists of three genes CID5, CID6, and CID7, while group D includes six genes CID8, CID9, CID10, CID11, CID12, and CID13 (Bravo et al., 2005). Group D CIDs are the closest

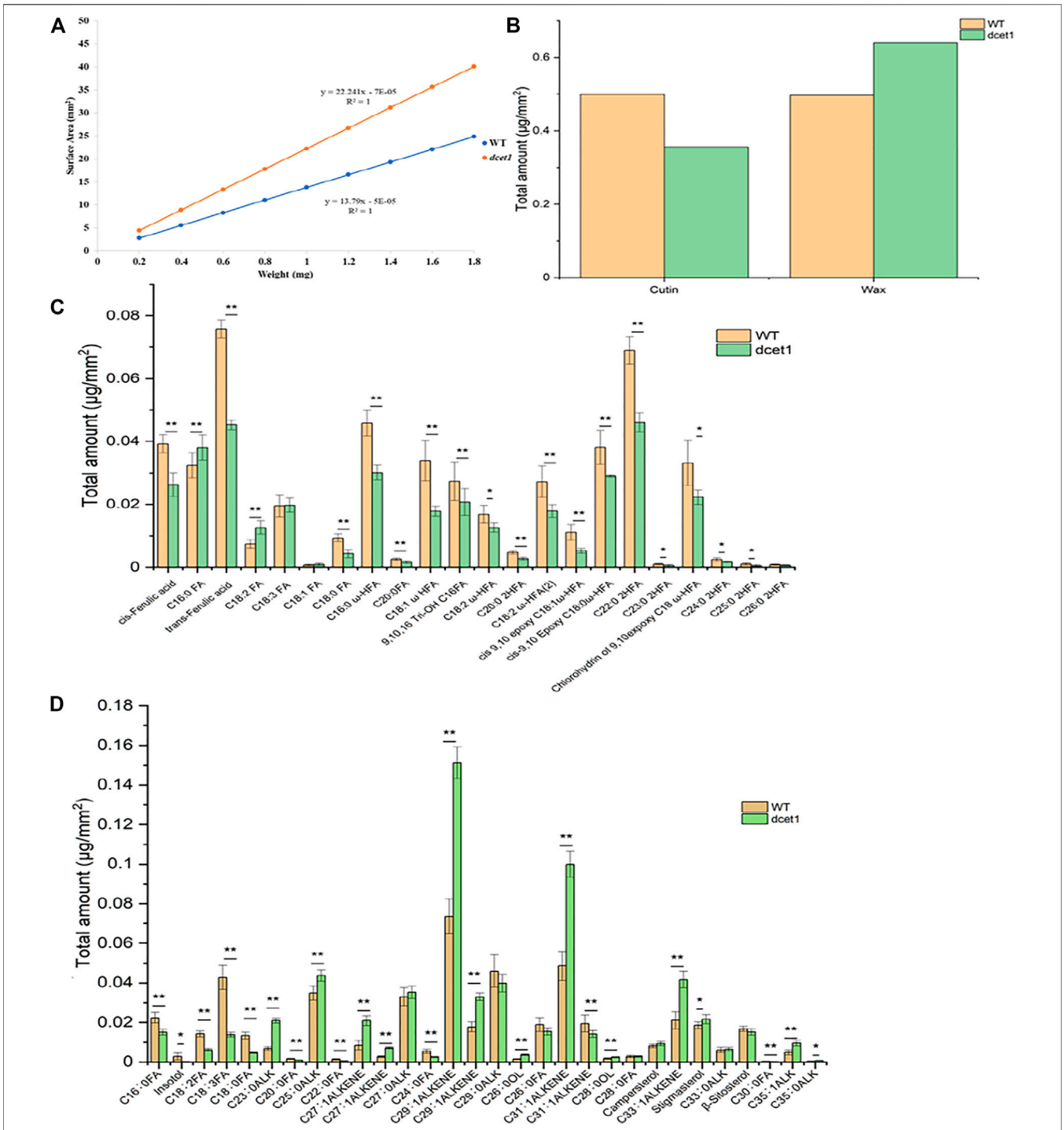
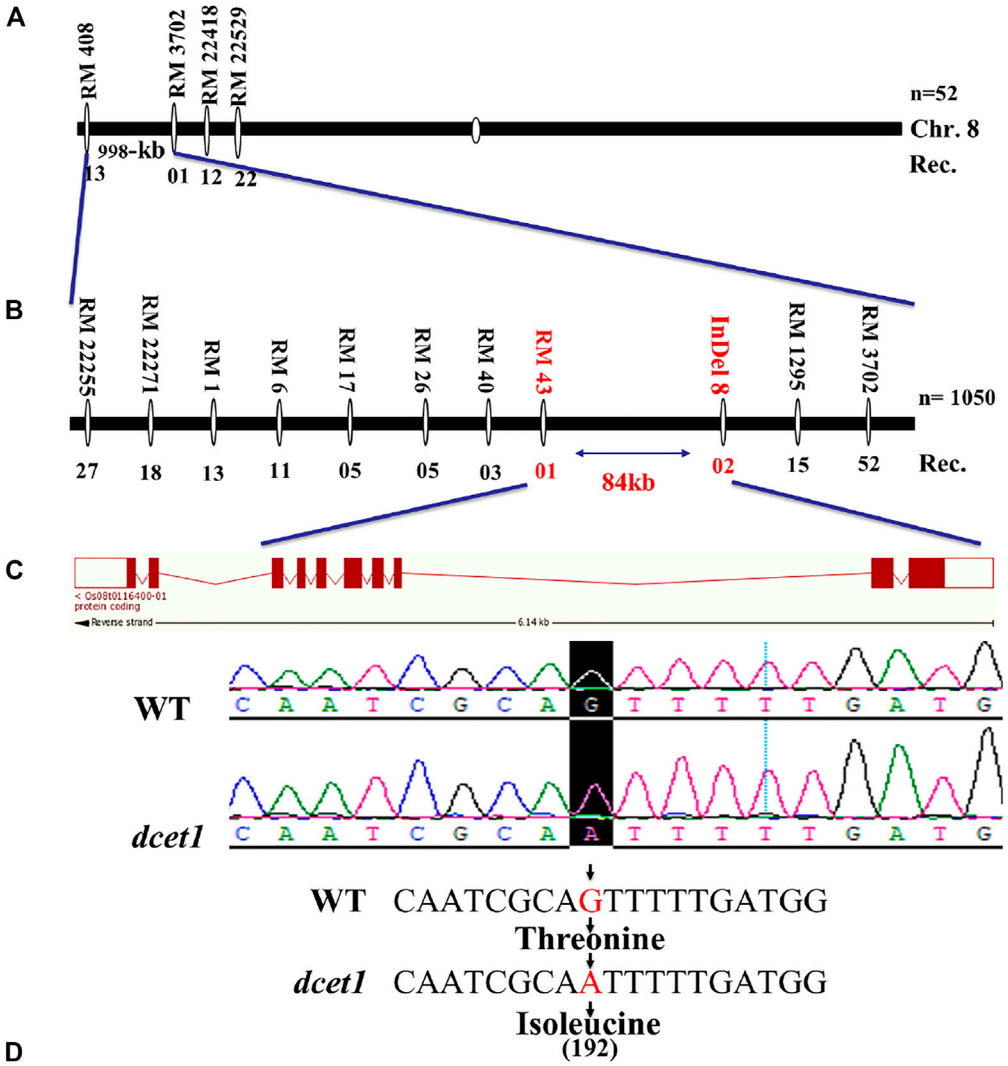


FIGURE 8 | Analysis of anther wax and cutin in the wild-type (WT) and mutant (*dcet1*). Weight/surface area ratio of the WT and *dcet1* anthers **(A)**. Total wax and cutin amounts per unit surface area in WT and *dcet1* anthers **(B)**. Cutin monomer amount per unit surface area in WT and *dcet1* anthers **(C)**. Wax constituent amount per unit surface area in WT and *dcet1* anthers **(D)**. R^2 represents regression values. Error bars indicate ± SD (n = 5). **p-value* < 0.05; ***p*-value< 0.01 by Student's *t*-test.

homologues of *DCET1*, holding the same two RRM domains and a basic C-terminal conserved region (**Figures 10A,B**), which can guide the protein interactions. The presence of RRMs and PAM2 in group D plus the previous study on CID12, one of the group D members, strongly suggests that in these proteins, there may be a dual-type of interaction taking place, that is, targeting mRNA and binding to PABP (Hecht et al., 1997). All the six members of this group are mainly uncharacterized proteins (with some exceptions for CID12, shown to express during early embryogenesis and growing organs) with unknown functions (Bravo et al., 2005). In



D

TransID	Variant	Chrom	Pos	Ref	Alt
MH08t0020700-01	nonsynonymous	Chr08	772504	G	A
MH08t0401600-01	nonsynonymous	Chr08	17652694	C	T
MH09t0093100-01	nonsynonymous	Chr09	3439883	T	A
MH01t0044600-01	splicing	Chr01	1395829	C	T

FIGURE 9 | Representation of map-based cloning of the *DCET1* gene and MutMap candidate all index SNP genes. Primary mapping **(A)**. Fine mapping **(B)**. Structure of the *DCET1* gene and single-nucleotide substitution in mutant *dcet1* from G (threonine) to A (isoleucine) on 192 amino acids **(C)**. MutMap candidate All index SNP genes **(D)**.

order to get further insight of the evolutionary and functional conservation among *DCET1* and its relatives in various plants species, additional 17 amino acid sequences were used. The peptide alignment and its phylogenetic tree showed that, including *DCET1*, all the 17 proteins had high conservation for the active sites (RRM1, RRM2, and PAM2), inferring that these functional positions are evolutionarily conserved in land plants (**Supplementary Figures S4A,B**). Our results showed severe defects in the *dcet1* anther phenotype during pollen development; we propose that like *DCET1*, its homologues, including CIDs, may have the same functional defects.

Expression Pattern of *DCET1*

Our early results suggested that mutation in *DCET1* led to abortive pollen in the *dcet1* mutant without affecting the vegetative growth, indicating that *DCET1* might be specifically expressed in anther. Thus, we performed the expression analysis through quantitative real-time-PCR (qRT-PCR) for different

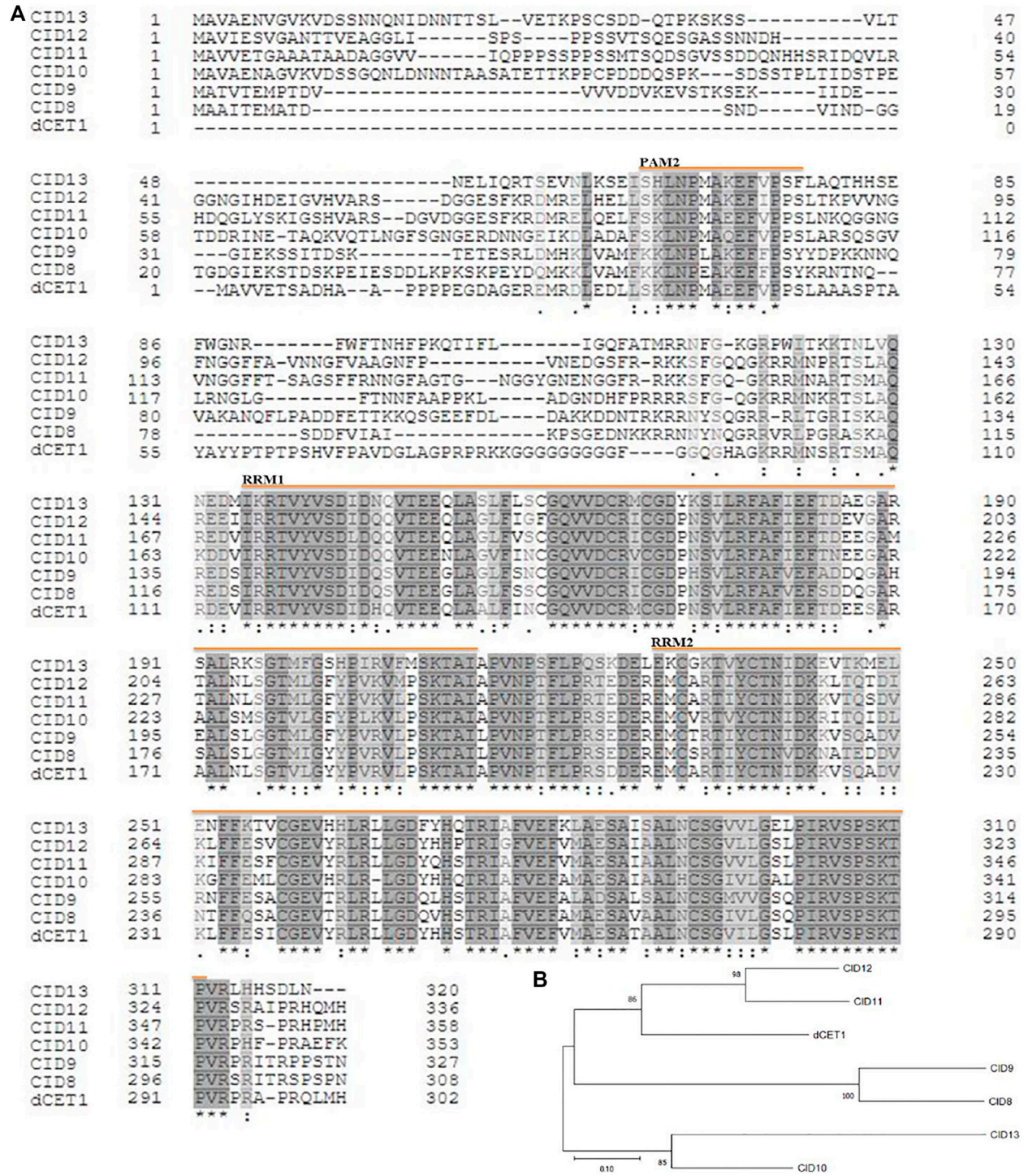


FIGURE 10 | Representation of sequence alignment and maximum likelihood tree of *DCET1* and CID-D proteins. Sequence alignment of *DCET1* and CID-D proteins **(A)**. Black and star indicate amino acid identity across all proteins; Grey and double dot indicate identity in most proteins; Light black and single dot indicate identity in many proteins; Yellow line with name indicates the region of the specific domain. Phylogenetic tree of *DCET1* and selected CID-D proteins **(B)**.

anther developmental stages and other WT plant tissues (**Figure 11A**). The level confirmed that although expressed in all tissues, the *DCET1* was ubiquitously expressed in anthers from stage 1 to 10 with the highest relative expression at stage 7 and 8 (**Figure 11A**). To further explicit the spatial and developmental nature of *DCET1* expression, we performed

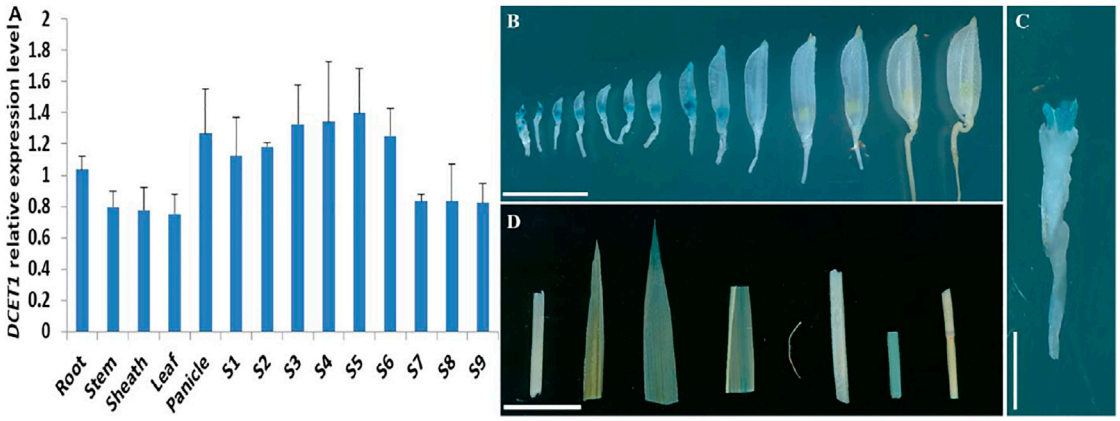


FIGURE 11 | Expression analysis of *DCET1*. QPCR relative expression level at different developmental stages of spikelets (on length basis) and other plant tissues **(A)**. S1–S10 anthers dissected from florets with 2–3 mm, 3–4 mm, 4–5 mm, 5–6 mm, 6–7 mm, 7–8 mm, 8–9 mm, 9–10 mm, >10 mm; GUS-staining of the *proDCET1*:*GUS* for transgenic line spikelets at different anther stages (1–14) **(B)**. GUS staining of the *proDCET1*:*GUS* for transgenic florets with palea and lemma removed **(C)**. GUS staining of the *proDCET1*:*GUS* leaf, root, sheath, and stem **(D)**. Scale bars **(B–D)** = 20 μm.

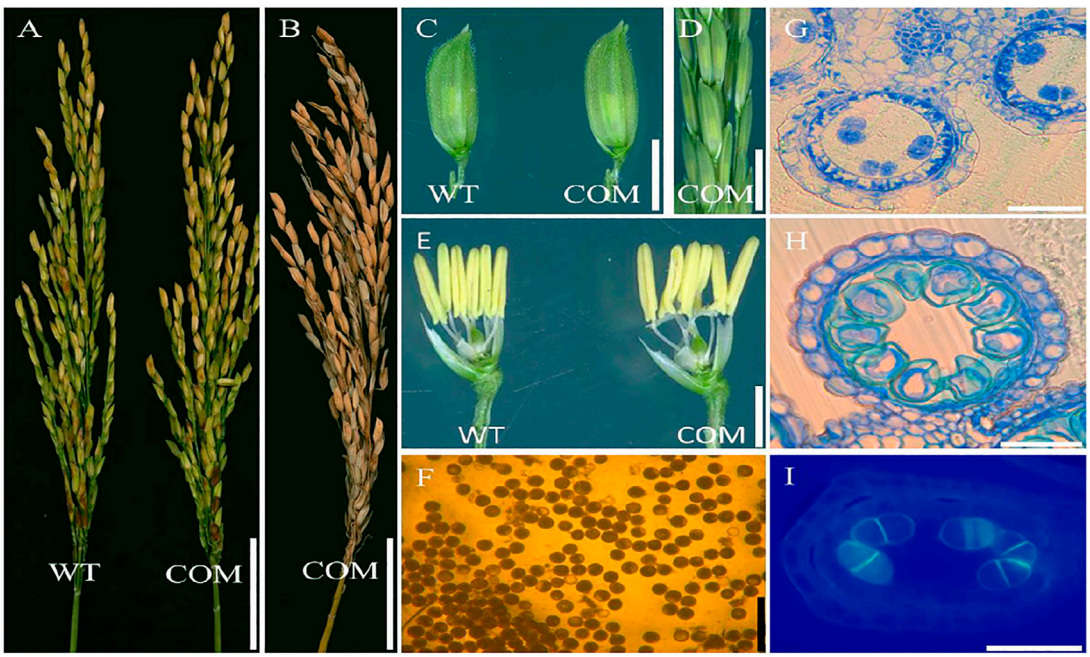


FIGURE 12 | Phenotypic characterization of the complemented line (COM) of *DCET1* and wild-type (WT). Panicles of WT and COM line at physiological maturity **(A)**. COM line panicle at harvest maturity **(B)**. Spikelets of the WT and COM line **(C)**. COM line panicle at stage 13 **(D)**. Spikelets of WT and COM line after the removal of palea and lemma **(E)**. Pollen grains of COM line after staining with 1.2% I2-KI staining solution **(F)**. Anther section of the COM line at stage 8b and 10, respectively, **(G,H)**. Aniline blue staining of the anther section at stage 8b **(I)**. Scale bars = 1 cm in **(A,B)**, 2 mm in **(C,D)**, 1 mm in **(E)**, 100 μm in **(F),** and 15 μm in **(G–I)**.

the native promoter-GUS analysis, which revealed that, although found to be expressed from stage 1 to 10 of anther development, the highest GUS staining expression was observed at stages 7 and 8 (**Figures 11B,C**). The other plant parts such as root, stem, sheath, and leaf showed negligible GUS staining (**Figure 11D**). These results were consistent with those of qRT-PCR, showing the highest relative expression levels in anthers from stages 7 to 8 (**Figure 11A**).

Functional Validation of the *DCET1* Gene

To confirm the mutation in the LOC_Os08g02330 caused male sterile phenotype in *dcet1*, a complementary vector of an entire 10,249-bp genomic sequence, consisting of 6,263 bp whole locus of *DCET1* (LOC_Os08g02330), 2,943 bp upstream sequence, and 1,043 bp downstream sequence was transformed into homozygous *dcet1* plants. The complementary transgenic plants produced normal seed setting (**Figures 12A,B**) with

dark yellow anthers (**Figures 12C–E**), and round black pollen grains in I_2-KI staining (**Figure 12F**). Furthermore, anther transverse section analysis was also carried out in complementary lines for the critical abnormal stages in anther development. Sections studies of stages 8b and 10 clearly showed the normal fertile phenotype in complementary plants (**Figures 12G,H**), like WT as described in **Figures 1B–E**. As the most critical abnormalities in the *dcet1* mutant for callose wall was observed at the anther development stage 8b, the aniline blue staining was carried out for this stage in the complementary line, which showed the normal callose layer (**Figure 12I**) resembling WT as described in **Figure 6H**. These results confirmed that the single bp substitution in LOC_Os08g02330 was responsible for the sterile phenotype in the *dcet1* mutant.

To gain further insights into the role of LOC_Os08g02330 in rice male sterility, we also generated a *dcet1-2* transgenic mutant by targeted knockout through the CRISPR/Cas9 system. We found two homozygous transgenic plants with similar single bp mutations in the targeted region (**Supplementary Figure S2E**). We used these mutated transgenic plants for phenotypic analysis, which showed the same male reproductive system defects as previously observed in *dcet1* (**Supplementary Figure S3A**). These findings confirmed that the mutation in the Os08t0116400-01 gene, named *DCET1*, was responsible for the defects in male fertility in the *dcet1* mutant phenotype.

DISCUSSION

Rice is a monocot and has many resources for understanding and exploring the male reproductive system development (Uzair et al., 2020). Abortive pollen is paramount for hybrid rice breeding; several genetic and environmental factors are responsible for pollen sterility. In *Arabidopsis* and rice, many genes are reported for pollen sterility through altered biosynthesis and biochemical degradation of the callose layer (Verma and Hong, 2001; Shi et al., 2016), abnormal PCD-induced tapetum degradation (Sun et al., 2018; Uzair et al., 2020), and defective pollen wall (Dong et al., 2005; Lu et al., 2014; Zou et al., 2017b). Here, we have cloned and functionally characterized a male sterility-related gene, *DCET1*, having normal vegetative growth but a defective male reproductive system through the abnormal callose wall, defective pollen wall, and delayed tapetum degradation and hence a complete male-sterile phenotype. *DCET1* belongs to *Arabidopsis* CID-D–like proteins of mainly unknown functions, suggested to perform the dual-type of interactions, targeting mRNA and binding to PABP (Hecht et al., 1997; Bravo et al., 2005). It is proposed that *DCET1* plays an important role in regulating and modifying male sterility through complex biological processes in rice.

DCET1 is Required for Callose Biosynthesis and Degradation

Callose (β-1,3-glucan) is an essential component in higher plants and is synthesized by CalS and deposited at various stages during reproductive system development to various cell organelles within the microsporocytes and megaspores (Shi et al., 2014; Shi et al., 2016). The callose wall is a provisional cell wall responsible for normal permeability and releasing and preventing the fusion of sibling microspores and the plasmic membranes (Lu et al., 2014). During microsporogenesis, the meiocytes till the division of meiotic dyads and tetrads need to be properly surrounded and separated by the callose wall. The degeneration and dissolution of callose through the callase enzyme secreted by both the tapetum and meiocytes itself after tetrad formation is the precursor for sporopollenin depositions on the primexine surface (Ariizumi and Toriyama, 2011). During the tetrad stage, the callose wall provides protection to the primexine by fusion with the plasma membrane and outside the locule (Ariizumi and Toriyama, 2011). Hence, proper time and amount of callose accumulation and dissolution is necessary for pollen wall and fertility (Dong et al., 2005). Glucan synthase–like (GSL) is a complex and important family that is involved in callose deposition in microspore developments. An *Arabidopsis* AtGSL10 (CalS9) knock-out line is involved in pollen development and aberrant asymmetric microspore formation at the mitotic division stage (Huang et al., 2009). Previous studies revealed that mutation in the callose synthase CalS5 leads to defective exine patterning; specifically, the microsporocytes of *CalS5-1* and *CalS5-2* mutants are not only missing with perceptible callose wall but also in exine patterning (Nishikawa et al., 2005; Shi et al., 2014). In the present study, we reported that the callose accumulation in *dcet1* mutant meiocytes is negligible, while during the tetrad stage, the callose wall was not formed around the microspore, although some amount of callose was seen to spread over the surface of the microspore and locule as shown in **Figure 6**. Our results are consistent with previously reported phenomena in the rice *GSL5* mutant, which showed the defects in callose accumulation, leading to pollen exine formation defects (Shi et al., 2014). These findings confirmed that *DCET1* is required for proper callose synthesis, not only for proper callose wall formation but also for pollen exine patterning.

DCET1 is Needed for Tapetal PCD

Tapetum is one of the four somatic layers in rice anther, lying to the innermost side, and is directly attached to the anther locule where the formation, division, and growth of microspores take place (Ariizumi and Toriyama, 2011). The previous study shows that the tapetum provides not only nutrients, metabolites, and sporopollenin precursors but also the materials for exine and tryphine formation; hence, any mutation in tapetum-related genes can lead to deformed tapetum and exine morphologies (Liu et al., 2017; Zhu et al., 2017). Likewise, the proper time and amount of tapetal cell PCD are considered necessary for callose degradation, exine wall patterning, and pollen viability through secretion of callases and sporopollenin precursors (Dong et al., 2005).

dcet1 exhibits delayed tapetal cell degradation and abnormal decay for other outer somatic layers validated by semi-thin section observation, TEM analysis, and TUNEL assay (**Figures 2**, **4**, **6**). *dcet1* showed the increase in somatic layer PCD specifically for tapetal cells till the pollen maturation, which

confirmed the delayed and abnormal nature of tapetal cell PCD and outer anther layer degradation in the *dcet1* mutant. Furthermore, the semi-thin studies and TEM analysis showed normal tapetum till the meiosis initiation stage. Interestingly, the tapetum was not degraded properly until the microspore maturation stage compared with the WT plants (**Figures 2**, **4**). These results are consistent with *OsGPAT3*, and *ptc2* mutants as previously characterized (Sun et al., 2018; Uzair et al., 2020). Our results indicate that *DCET1* plays a critical role in regulating PCD-induced tapetal cell degradation; however, the proper mechanism and pathway are still unclear and need further elucidation for male sterility functions in rice crops.

Proper Callose; Timely Tapetal PCD; or Both; Needed for Exine in *dcet1* Mutant

It is a well-known fact that both the tapetum and the developing microspores contribute to exine formation, and any interference in these processes can lead to the male sterility phenotype (Piffanelli et al., 1998; Yang et al., 2017). During microsporogenesis, the callose is synthesized and accumulated on microsporocytes, tetrads, and microspores (Dong et al., 2005). At the tetrad stage, the tapetum and meiocytes also secrete callases for callose degradation, but before the callases degrade the callose, primexine appears on the microspore surface. At the same time, the tapetum also starts to release sporopollenin precursors to be deposited on the primexine surface, responsible for the final exine wall and pollen fertility (Men et al., 2017). Furthermore, the normal tapetal PCD is considered essential for normal exine patterning (Lu et al., 2014). Here, the deposition of the sporopollenin precursor on the primexine surface needs the presence of a proper callose layer to guarantee the process by preventing cell fusion and cohesion. As mentioned, primexine is important for exine patterning, but for proper primexine formation, both the callose and tapetum are necessary. In rice and *Arabidopsis*, several genes and transcription factors have been identified for callose regulation, tapetal PCD, and pollen wall development and patterning. Based on their functional relations, we divided these genes and transcription factors into three groups. The callose-regulated genes which can affect the pollen wall are *CalS5* (Dong et al., 2005), *GSL5* (Shi et al., 2014), and *CDM1* (Lu et al., 2014). Those which affect tapetal PCD and also the pollen wall are *PTC1* (Li et al., 2011), *PTC2* (Uzair et al., 2020), *OsCOS12* (Yang et al., 2017), *TIP3* (Yang et al., 2019b), *OsGPAT3* (Sun et al., 2018), and *EAT1* (Niu et al., 2013). There are also several genes and transcriptional factors which are not only associated with the callose layer and tapetum PCD but also with the pollen wall such as *AtMYB103* (Zhang et al., 2007) and *OsDEX1* (Yu et al., 2016).

Several studies have confirmed that alteration in sporopollenin constituents are highly related to exine patterning, as both of the cutin and sporopollenin share common monomers (Liu et al., 2017). So, the changes in cutin monomers can lead to the sporopollenin precursor deviations and ultimately to pollen sterility. As reported for *DPW* to produce the critically needed C16:0 fatty alcohols for exine development, the *dpw* mutant has reduced sporopollenin deposition on the primexine surface of the pollen grains and hence having degenerated pollen grains with irregular exine and abnormal cuticle (Shi et al., 2011). Wax-defective anther1 (*WDA1*) is associated with the formation of cuticle and wax deposition by the biosynthesis of very long-chain fatty acids (VLCFA) in rice, and its mutant (*wda1*) shows abnormal anther wall and pollen exine formation (Jung et al., 2006). The *dcet1* mutant is deficient in cutin monomers, while the wax constituents were mainly over-accumulated like both of *ptc2* and *dpw* phenotypes in rice (Shi et al., 2011; Uzair et al., 2020) (**Figures 8B–D**). The maize *apv1* mutant has shows the same cuticular lipid trend as that of *dcet1*, and having a complete male sterile phenotype (Somaratne et al., 2017). Specifically, the delayed tapetal PCD phenotype of *dcet1* resembled that of the *ptc2* mutant; both expressed almost the same anther and pollen surface patterns in SEM and lipidic analysis (Uzair et al., 2020) (**Figures 3A–J**, **8A–D**). From the previous literature, it can also be assumed that the defects of pollen exine in *dcet1* may reflect that *DCET1* disturbs the mutual biosynthetic pathway for the two biopolymer's synthesis, the sporopollenin and cutin. Still for exploring the meticulous pathway, responsible for the dcet1 phenotype, further studies are needed with all related genes.

As previously reported in the *gls5* mutant (Shi et al., 2014) in rice, the timing of callose deposition and degradation is important for exine wall formation. Here, we add the proper tapetal PCD, with the quoted sentence important for exine wall formation in the *dcet1* mutant. As we can assume from our results that either callose or tapetum or both could be the causative agents for exine malformation, hence further research and investigations are needed on the *DCET1* gene to solve the complexity.

In short, we have identified RNA recognition motifs containing protein and *DCET1*, which is required for pollen development by regulating callose biosynthesis, tapetal PCD, and pollen wall patterning. In *dcet1*, both callose biosynthesis and its degradation and tapetal cell PCD are abnormal, and eventually, the exine with its patterning is improper.

CONCLUSION

The current study presents the functional characterization of the *DCET1* gene in rice. Our results showed that *DCET1*, belonging to RNA recognition motif (RRM)–containing proteins, is potentially involved in male sterility in rice. In brief, *DCET1* regulates callose biosynthesis and degradation and pollen exine formation by affecting exine wall patterning, including abnormal nexine, collapsed bacula, irregular tectum, and timely PCD by delaying the tapetal cell degeneration. As a result, the microspore of *dcet1* was swollen and abnormally bursted and even collapsed within the anther locule with complete sterility. Taken together, the disruption of *DCET1* function leads to the extension of male sterility and reproductive system in rice and will play an important role in hybrid rice breeding programs. CID-like family proteins are the closest homologs of *Arabidopsis*; however, many have unknown functions (except CID12 expressed during embryogenesis). Therefore, *DCET1* receives key importance and provides novel insights for the prediction

and characterization of RBDs, in general, and CIDs in specific in all eukaryotes.

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

RK performed most of the experiments, including phenotypic analysis, sample collection, cytological analysis, MutMap analysis, cloning, transformation, GUS staining, qRT-PCR, TUNEL, and aniline blue analysis. LSu, PY, SC, LC, and RK conducted the experimental design and data discussion. QL and YZ assisted in phenotypic identification and creation of transgenic lines. AA and XX helped with map-based cloning and cytological analysis. DW assisted in TUNEL analysis. LSh and AS provided help in writing the manuscript. LC supervised the overall research project.

FUNDING

This work was supported by the grants form Zhejiang Province Key Research and Development Program of China (2021C02056), China Agriculture Research System, and Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Science (CAAS-ASTIP-2013-CNRRI).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.790789/full#supplementary-material

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

Ariizumi, T., and Toriyama, K. (2011). Genetic Regulation of Sporopollenin Synthesis and Pollen Exine Development. *Annu. Rev. Plant Biol.* 62, 437–460. doi:10.1146/annurev-arplant-042809-112312

Blackmore, S., Wortley, A. H., Skvarla, J. J., and Rowley, J. R. (2007). Pollen wall Development in Flowering Plants. *New Phytol.* 174, 483–498. doi:10.1111/j.1469-8137.2007.02060.x

Bravo, J., Aguilar-Henonin, L., Olmedo, G., and Guzmán, P. (2005). Four Distinct Classes of Proteins as Interaction Partners of the PABC Domain of *Arabidopsis thaliana* Poly(A)-binding Proteins. *Mol. Genet. Genomics* 272, 651–665. doi:10.1007/s00438-004-1090-9

Chang, F., Zhang, Z., Jin, Y., and Ma, H. (2014). "Cell Biological Analyses of Anther Morphogenesis and Pollen Viability in Arabidopsis and Rice," in *Cell Biological Analyses of Anther Morphogenesis and Pollen Viability in Arabidopsis and RiceFlower Development: Methods and Protocols*. Editors J. L. Riechmann and F. Wellmer (New York, NY: Springer New York), 203–216. doi:10.1007/978-1-4614-9408-9_9

Chang, Z., Chen, Z., Wang, N., Xie, G., Lu, J., Yan, W., et al. (2016). Construction of a Male Sterility System for Hybrid rice Breeding and Seed Production Using a Nuclear Male Sterility Gene. *Proc. Natl. Acad. Sci. USA* 113, 14145–14150. doi:10.1073/pnas.1613792113

Cheng, S.-H., Cao, L.-Y., Zhuang, J.-Y., Chen, S.-G., Zhan, X.-D., Fan, Y.-Y., et al. (2007). Super Hybrid Rice Breeding in China: Achievements and Prospects. *J. Integr. Plant Biol.* 49, 805–810. doi:10.1111/j.1744-7909.2007.00514.x

Dong, X., Hong, Z., Sivaramakrishnan, M., Mahfouz, M., and Verma, D. P. S. (2005). Callose Synthase (CalS5) Is Required for Exine Formation during Microgametogenesis and for Pollen Viability in Arabidopsis. *Plant J.* 42, 315–328. doi:10.1111/j.1365-313x.2005.02379.x

Hecht, V., Stiefel, V., Delseny, M., and Gallois, P. (1997). A New Arabidopsis Nucleic-Acid-Binding Protein Gene Is Highly Expressed in Dividing Cells during Development. *Plant Mol. Biol.* 34, 119–124. doi:10.1023/a:1005834402536

Hu, J., Wang, Z., Zhang, L., and Sun, M. x. (2014). The Arabidopsis Exine Formation Defect (EFD) Gene Is Required for Primexine Patterning and Is Critical for Pollen Fertility. *New Phytol.* 203, 140–154. doi:10.1111/nph.12788

Jiang, J., Zhang, Z., and Cao, J. (2013). Pollen wall Development: the Associated Enzymes and Metabolic Pathways. *Plant Biol.* 15, 249–263. doi:10.1111/j.1438-8677.2012.00706.x

Jung, K.-H., Han, M.-J., Lee, D.-Y., Lee, Y.-S., Schreiber, L., Franke, R., et al. (2006). Wax-deficient Anther1 Is Involved in Cuticle and Wax Production in Rice Anther Walls and Is Required for Pollen Development. *Plant Cell* 18, 3015–3032. doi:10.1105/tpc.106.042044

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096

Li, H., Yuan, Z., Vizcay-Barrena, G., Yang, C., Liang, W., Zong, J., et al. (2011). PERSISTENT TAPETAL CELL1 Encodes a PHD-Finger Protein that Is Required for Tapetal Cell Death and Pollen Development in Rice. *Plant Physiol.* 156, 615–630. doi:10.1104/pp.111.175760

Li, J., Xin, Y., and Yuan, L. (2009). Hybrid rice Technology Development: Ensuring China Food Security. *Proven Successes in Agricultural Development: A Technical Compendium to Millions Fed*.

Li, N., Zhang, D.-S., Liu, H.-S., Yin, C.-S., Li, X.-X., Liang, W.-Q., et al. (2006). The RiceTapetum Degeneration RetardationGene Is Required for Tapetum Degradation and Anther Development. *The Plant cell* 18, 2999–3014. doi:10.1105/tpc.106.044107

Liu, Z., Lin, S., Shi, J., Yu, J., Zhu, L., Yang, X., et al. (2017). Rice No Pollen 1(NP1) Is Required for Anther Cuticle Formation and Pollen Exine Patterning. *Plant J.* 91, 263–277. doi:10.1111/tpj.13561

Lou, Y., Zhu, J., and Yang, Z. (2014a). Molecular Cell Biology of Pollen Walls. *Plant Cel Monogr.* 22, 179–205. doi:10.1007/978-3-642-41787-0_6

Lou, Y., Zhu, J., and Yang, Z. (2014b). "Molecular Cell Biology of Pollen Walls," in *Applied Plant Cell Biology: Cellular Tools and Approaches for Plant Biotechnology*. Editors P. Nick and Z. Opatrny (Berlin, Heidelberg: Springer Berlin Heidelberg), 179–205. doi:10.1007/978-3-642-41787-0_6

Lu, P., Chai, M., Yang, J., Ning, G., Wang, G., and Ma, H. (2014). The Arabidopsis CALLOSE DEFECTIVE MICROSPORE1 Gene Is Required for Male Fertility through Regulating Callose Metabolism during Microsporogenesis. *Plant Physiol.* 164, 1893–1904. doi:10.1104/pp.113.233387

Men, X., Shi, J., Liang, W., Zhang, Q., Lian, G., Quan, S., et al. (2017). Glycerol-3-Phosphate Acyltransferase 3 (OsGPAT3) Is Required for Anther Development and Male Fertility in rice. *J. Exp. Bot.* 68, 513–526. doi:10.1093/jxb/erw445

Nishikawa, S.-I., Zinkl, G. M., Swanson, R. J., Maruyama, D., and Preuss, D. (2005). Callose (β-1,3 Glucan) Is Essential for Arabidopsispollen wall Patterning, but Not Tube Growth. *BMC Plant Biol.* 5, 22. doi:10.1186/1471-2229-5-22

Niu, N., Liang, W., Yang, X., Jin, W., Wilson, Z. A., Hu, J., et al. (2013). EAT1 Promotes Tapetal Cell Death by Regulating Aspartic Proteases during Male Reproductive Development in rice. *Nat. Commun.* 4, 1445. doi:10.1038/ncomms2396

Parish, R. W., and Li, S. F. (2010). Death of a Tapetum: A Programme of Developmental Altruism. *Plant Sci.* 178, 73–89. doi:10.1016/j.plantsci.2009.11.001

Piffanelli, P., Ross, J. H. E., and Murphy, D. J. (1998). Biogenesis and Function of the Lipidic Structures of Pollen Grains. *Sex. Plant Reprod.* 11, 65–80. doi:10.1007/s004970050122

Richmond, T. A., and Somerville, C. R. (2001). "Integrative Approaches to Determining Csl Function," in *Plant Cell Walls*. Editors N. C. Carpita, M. Campbell, and M. Tierney (Dordrecht: Springer Netherlands), 131–143. doi:10.1007/978-94-010-0668-2_8

Saitou, N., and Nei, M. (1987). The Neighbor-Joining Method: a New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* 4, 406–425. doi:10.1093/oxfordjournals.molbev.a040454

Scott, R. J. (1994). "Pollen Exine - the Sporopollenin enigma and the Physics of Pattern," in *Molecular and Cellular Aspects of Plant Reproduction*. Editors A. D. Stead and R. J. Scott (Cambridge: Cambridge University Press), 49–82. doi:10.1017/cbo9780511752339.006

Scott, R. J., Spielman, M., and Dickinson, H. G. (2004). Stamen Structure and Function. *The Plant Cell Online* 16, S46–S60. doi:10.1105/tpc.017012

Shi, J., Cui, M., Yang, L., Kim, Y.-J., and Zhang, D. (2015). Genetic and Biochemical Mechanisms of Pollen Wall Development. *Trends Plant Sci.* 20, 741–753. doi:10.1016/j.tplants.2015.07.010

Shi, J., Tan, H., Yu, X.-H., Liu, Y., Liang, W., Ranathunge, K., et al. (2011). Defective Pollen Wall Is Required for Anther and Microspore Development in Rice and Encodes a Fatty Acyl Carrier Protein Reductase. *The Plant Cell* 23, 2225–2246. doi:10.1105/tpc.111.087528

Shi, X., Han, X., and Lu, T.-G. (2016). Callose Synthesis during Reproductive Development in Monocotyledonous and Dicotyledonous Plants. *Plant Signaling Behav.* 11, e1062196. doi:10.1080/15592324.2015.1062196

Shi, X., Sun, X., Zhang, Z., Feng, D., Zhang, Q., Han, L., et al. (2014). GLUCAN SYNTHASE-LIKE 5 (GSL5) Plays an Essential Role in Male Fertility by Regulating Callose Metabolism during Microsporogenesis in Rice. *Plant Cel Physiol.* 56, 497–509. doi:10.1093/pcp/pcu193

Somaratne, Y., Tian, Y., Zhang, H., Wang, M., Huo, Y., Cao, F., et al. (2017). ABNORMAL POLLEN VACUOLATION1 (APV1) Is Required for Male Fertility by Contributing to Anther Cuticle and Pollen Exine Formation in maize. *Plant J.* 90, 96–110. doi:10.1111/tpj.13476

Sun, L., Xiang, X., Yang, Z., Yu, P., Wen, X., Wang, H., et al. (2018). OsGPAT3 Plays a Critical Role in Anther Wall Programmed Cell Death and Pollen Development in Rice. *Ijms* 19, 4017. doi:10.3390/ijms19124017

Uzair, M., Xu, D., Schreiber, L., Shi, J., Liang, W., Jung, K.-H., et al. (2020). PERSISTENT TAPETAL CELL2 Is Required for Normal Tapetal Programmed Cell Death and Pollen Wall Patterning. *Plant Physiol.* 182, 962–976. doi:10.1104/pp.19.00688

Verma, D. P. S., and Hong, Z. (2001). Plant Callose Synthase Complexes. *Plant Mol. Biol.* 47, 693–701. doi:10.1023/a:1013679111111

Wu, H.-M., and Cheung, A. Y. (2000). Programmed Cell Death in Plant Reproduction. *Plant Mol. Biol.* 44, 267–281. doi:10.1023/a:1026536324081

Wu, Z., Ji, J., Tang, D., Wang, H., Shen, Y., Shi, W., et al. (2015). OsSDS Is Essential for DSB Formation in rice Meiosis. *Front. Plant Sci.* 6, 21. doi:10.3389/fpls.2015.00021

Xie, B., Deng, Y., Kanaoka, M. M., Okada, K., and Hong, Z. (2012). Expression of Arabidopsis Callose Synthase 5 Results in Callose Accumulation and Cell wall Permeability Alteration. *Plant Sci.* 183, 1–8. doi:10.1016/j.plantsci.2011.10.015

Yang, X., Liang, W., Chen, M., Zhang, D., Zhao, X., and Shi, J. (2017). Rice Fatty Acyl-CoA Synthetase OsACOS12 Is Required for Tapetum Programmed Cell Death and Male Fertility. *Planta* 246, 105–122. doi:10.1007/s00425-017-2691-y

Yang, Z., Liu, L., Sun, L., Yu, P., Zhang, P., Abbas, A., et al. (2019a). OsMS1 Functions as a Transcriptional Activator to Regulate Programmed Tapetum Development and Pollen Exine Formation in rice. *Plant Mol. Biol.* 99, 175–191. doi:10.1007/s11103-018-0811-0

Yang, Z., Sun, L., Zhang, P., Zhang, Y., Yu, P., Liu, L., et al. (2019b). TDR INTERACTING PROTEIN 3 , Encoding a PHD -finger Transcription Factor, Regulates Ubisch Bodies and Pollen wall Formation in rice. *Plant J.* 99, 844–861. doi:10.1111/tpj.14365

Yu, J., Meng, Z., Liang, W., Behera, S., Kudla, J., Tucker, M. R., et al. (2016). A Rice Ca2+ Binding Protein Is Required for Tapetum Function and Pollen Formation. *Plant Physiol.* 172, 1772–1786. doi:10.1104/pp.16.01261

Zafar, S. A., Patil, S. B., Uzair, M., Fang, J., Zhao, J., Guo, T., et al. (2020). DEGENERATED PANICLE and PARTIAL STERILITY 1 (DPS 1) Encodes a Cystathionine β - Synthase Domain Containing Protein Required for Anther Cuticle and Panicle Development in rice. *New Phytol.* 225, 356–375. doi:10.1111/nph.16133

Zhang, D., and Li, H. (2014). "Exine Export in Pollen," in *Plant ABC Transporters*. Editor M. Geisler (Cham: Springer International Publishing), 49–62. doi:10.1007/978-3-319-06511-3_4

Zhang, D., Luo, X., and Zhu, L. (2011). Cytological Analysis and Genetic Control of rice Anther Development. *J. Genet. Genomics* 38, 379–390. doi:10.1016/j.jgg.2011.08.001

Zhang, P., Zhang, Y., Sun, L., Sinumporn, S., Yang, Z., Sun, B., et al. (2017). The Rice AAA-ATPase OsFIGNL1 Is Essential for Male Meiosis. *Front. Plant Sci.* 8, 1639. doi:10.3389/fpls.2017.01639

Zhang, Z.-B., Zhu, J., Gao, J.-F., Wang, C., Li, H., Li, H., et al. (2007). Transcription Factor AtMYB103 Is Required for Anther Development by Regulating Tapetum Development, Callose Dissolution and Exine Formation in Arabidopsis. *Plant J.* 52, 528–538. doi:10.1111/j.1365-313x.2007.03254.x

Zhu, L., Shi, J., Zhao, G., Zhang, D., and Liang, W. (2013). Post-meiotic Deficient Anther1 (PDA1) Encodes an ABC Transporter Required for the Development of Anther Cuticle and Pollen Exine in rice. *J. Plant Biol.* 56, 59–68. doi:10.1007/s12374-013-0902-z

Zhu, X., Yu, J., Shi, J., Tohge, T., Fernie, A. R., Meir, S., et al. (2017). The Polyketide Synthase OsPKS2 Is Essential for Pollen Exine and Ubisch Body Patterning in rice. *J. Integr. Plant Biol.* 59, 612–628. doi:10.1111/jipb.12574

Zou, T., Li, S., Liu, M., Wang, T., Xiao, Q., Chen, D., et al. (2017a). An Atypical Strictosidine Synthase, OsSTRL2, Plays Key Roles in Anther Development and Pollen wall Formation in rice. *Sci. Rep.* 7, 6863. doi:10.1038/s41598-017-07064-4

Zou, T., Xiao, Q., Li, W., Luo, T., Yuan, G., He, Z., et al. (2017b). OsLAP6/OsPKS1, an Orthologue of Arabidopsis PKSA/LAP6, Is Critical for Proper Pollen Exine Formation. *Rice* 10, 53. doi:10.1186/s12284-017-0191-0

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fgene.2021.726152

Identification of a Diverse Core Set Panel of Rice From the East Coast Region of India Using SNP Markers

Debjani Roy Choudhury[1], *Ramesh Kumar*[1], *Vimala Devi S*[2], *Kuldeep Singh*[3], *N. K. Singh*[4] *and Rakesh Singh*[1]*

[1]Division of Genomic Resources, NBPGR, New Delhi, India, [2]Division of Germplasm Conservation, NBPGR, New Delhi, India, [3]NBPGR, New Delhi, India, [4]NIPB, New Delhi, India

OPEN ACCESS

Edited by:
Ahmed Sallam, Assiut University, Egypt

Reviewed by:
Bikram Pratap Singh, Indian Council of Agricultural Research, India
Yasser Shaaban Sayed Moursi, Fayoum University, Egypt
Shamseldeen Shehabeldin Eltaher, University of Sadat City, Egypt

***Correspondence:**
Rakesh Singh
rakesh.singh2@icar.gov.in

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *16 June 2021*
Accepted: *26 October 2021*
Published: *25 November 2021*

Citation:
Choudhury DR, Kumar R, S VD, Singh K, Singh NK and Singh R (2021) Identification of a Diverse Core Set Panel of Rice From the East Coast Region of India Using SNP Markers. Front. Genet. 12:726152. doi: 10.3389/fgene.2021.726152

In India, rice (*Oryza sativa* L.) is cultivated under a variety of climatic conditions. Due to the fragility of the coastal ecosystem, rice farming in these areas has lagged behind. Salinity coupled with floods has added to this trend. Hence, to prevent genetic erosion, conserving and characterizing the coastal rice, is the need of the hour. This work accessed the genetic variation and population structure among 2,242 rice accessions originating from India's east coast comprising Andhra Pradesh, Orissa, and Tamil Nadu, using 36 SNP markers, and have generated a core set (247 accessions) as well as a mini-core set (30 accessions) of rice germplasm. All the 36 SNP loci were biallelic and 72 alleles found with average two alleles per locus. The genetic relatedness of the total collection was inferred using the un-rooted neighbor-joining tree, which grouped all the genotypes (2,242) into three major clusters. Two groups were obtained with a core set and three groups obtained with a mini core set. The mean PIC value of total collection was 0.24, and those of the core collection and mini core collection were 0.27 and 0.32, respectively. The mean heterozygosity and gene diversity of the overall collection were 0.07 and 0.29, respectively, and the core set and mini core set revealed 0.12 and 0.34, 0.20 and 0.40 values, respectively, representing 99% of distinctiveness in the core and mini core sets. Population structure analysis showed maximum population at K = 4 for total collection and core collection. Accessions were distributed according to their population structure confirmed by PCoA and AMOVA analysis. The identified small and diverse core set panel will be useful in allele mining for biotic and abiotic traits and managing the genetic diversity of the coastal rice collection. Validation of the 36-plex SNP assay was done by comparing the genetic diversity parameters across two different rice core collections, i.e., east coast and northeast rice collection. The same set of SNP markers was found very effective in deciphering diversity at different genetic parameters in both the collections; hence, these marker sets can be utilized for core development and diversity analysis studies.

Keywords: rice, SNP markers, genetic diversity, genotyping, SNP, coastal rice

INTRODUCTION

Rice (*Oryza sativa* L.) is an important cereal crop which is a predominant food for over three billion people across the globe. Since it can adapt to an extensive spectrum of environmental conditions, it is therefore considered a varied crop species (Chang, 1976). In general, more genetically diverse crops will have an increased capacity to adapt with climatic conditions, while uniformity reduces the genetic diversity (Luan et al., 2006). There should be more significant variation in the breeding population so as to stimulate genetic gain from selection to improved yield, biotic and abiotic resistance to stress, and other traits. Genebanks with proper coordination can be explored to find genetic variations present in its accessions which can further help in finding advanced traits and selecting better parental combinations for developing lineage with maximum genetic variability (Barrett and Kidwell, 1998). Genetic diversity and population structure knowledge form the backbone in building core sets adequately representing variations found in the whole collection, and thereby making the collection small and condensed (Yan et al., 2007; Agrama et al., 2009; El Bakkali et al., 2013), thus creating standard data and calculating the potential loss of genetic diversity during conservation and management (Reif et al., 2005). Trait-specific germplasm identification from large collections is crucial to successful introduction of new diversity into the crop improvement programs (Upadhyaya et al., 2014). It is pivotal to comprehend the scope of genetic variation in crop germplasm and how to control it; results from molecular marker-based diversity studies should be used with caution, especially when it comes to germplasm conservation initiatives, the reason being adaptability, which plays a major role during the process of evolution and individuals' survival in populations. As a result, it becomes ambiguous, determining whether plant selection is based on markers directly or on linked loci responsible for adaptive traits (Raybould et al., 1996).

Single-nucleotide polymorphisms (SNPs) are supposed to be the most bountiful variation found across the genome and therefore are excellent for high-resolution genotyping, in turn useful for analysis of genetic diversity and association mapping, linkage mapping, and marker-assisted selection (MAS) (Gonzaga et al., 2015). Simple sequence repeats (SSRs) are being replaced by SNPs at a very high degree. They have become a choice for utilization in plant breeding and genetics (McCouch et al., 2010). Using second-generation sequencing methods, scientists have been able to detect millions of SNPs in rice genome (Huang et al., 2009, 2010; Xu et al., 2011). Array-based SNP detection is currently one of the most popular high-throughput marker detection methods, allowing data to be evaluated in real time. Because of their abundance, locus specificity, low error rates, and co-dominant inheritance SNP assays are popular markers (Rafalski, 2002; Schlotterer, 2004). There are numerous multiplex and uniplex genotyping platforms are available (Syvänen, 2001; Chen and Sullivan, 2003). SNP genotyping platforms including Illumina BeadXpress (Yamamoto et al., 2010; Thomson et al., 2012), Affymetrix (Singh et al., 2015; McCouch et al., 2016), and the KASP marker system (Cheon et al., 2018; Yang et al., 2019) have been developed recently and applied to rice molecular breeding. In the year 2020, Seo et al. (2020) developed two 96-plex *indica-japonica* SNP genotyping sets using the Fluidigm platform; however, genotyping platforms are very expensive.

Plant Genetic Resources of native rice needs to be conserved promptly. A major conservation strategy counts on developing core collections, thereby screening gene bank accessions for important agronomic traits. To create a core collection, one needs a good construction strategy. Frankel (1984) and Brown (1989) suggested the concept of core collection; it is a small set of accessions from the entire collection with least amount of repetition and maximal genetic diversity of a species. It is usually 5% to 10% representation of the total population. Van Hintum et al. (2000), in the year 2000, outlined a comprehensive process for assembling a core collection. These can be summarized as 1) determine the overall sampling ratio; 2) partition them into groups; 3) decide on the percentage of the group that will be sampled; and 4) choose from each group entry. Several other strategies have been proposed to create a core collection such as PowerCore (Kim et al., 2007), MStrat (Gouesnard et al., 2001), stepwise clustering (Hu et al., 2000), and least distance stepwise clustering (Wang et al., 2007). These various methodologies are dependent on several parameters such as species' genetic diversity, collection size, grouping, and data type (i.e., phenotypic or molecular data). However, if adequately characterized, a quality core collection can form a progressive collection for long-term conservation.

In the present study, we demonstrate the identification of a diverse core set panel from the east coast region of India comprising 247 core accessions and a mini core set of 30 accessions to enable elite gene mining for breeding and conservation and on a long-term perspective to form a panel for association studies. We have also done a comparative analysis of genetic diversity parameters across two different rice collections (northeast rice collection by Roy Choudhury et al., in 2014, and the coastal rice collection), thus validating the use of SNP markers for studying diversity parameters across any other rice collection.

MATERIALS AND METHODS

Germplasm resources

A set of 2,242 seed samples of the east coast region of India (Andhra Pradesh, Orissa, and Tamil Nadu) were procured from the Indian National Genebank, National Bureau of Plant Genetic Resources (NBPGR), New Delhi, with passport data (National ID, i.e., Indigenous Collection (IC) number) and the states to which they belong. This is given in **Supplementary Table S1**.

Genomic DNA isolation and molecular characterization using SNP markers

The seed was de-husked, and genomic DNA was isolated using the QIAGEN DNeasy Plant Mini Kit. A tissue lyser (Retsch, Haan, Germany) and a tissue lyser adapter set (Qiagen, Hilden,

Germany) was used to grind kernels into fine powder. Working stocks with 10 ng/µl of the genomic DNA samples was prepared, and 30 µl of the diluted sample was transferred to a 96-well plate to be run on the Sequenom MassARRAY which adopts the matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometer for most authentic detection of SNPs (www.sequenom.com). The information about the genetic location of the multiform assays designed for 36 SNPs having conserved single-copy rice genes was derived from Singh et al. (2007). Sequenom Corporation (San Diego, CA, USA) created and validated a 36-plex assay with three genes per chromosome. The assay Design 3.1 program was used to build pre-amplification primers and genotyping primers, purchased and utilized for SNP validation according to the Sequenom user manual's methodology. There were two polymerase chain reactions (PCR). The first was a normal PCR of 45 cycles with pre-amplification primers followed by removal of unincorporated dNTPs. After adjusting genotyping primers, the second PCR was an extension PCR. The extension rate was enhanced by increasing the number of cycles to 300, thus giving the highest call rates. Call rates and extension rates were also adjusted according to genotype calling algorithms. MassARRAY Typer Analyzer 3.4 Software was used to visualize SNP calling.

Genetic diversity and phylogenetic analyses

The results of SNP data were subjected to analysis using PowerMarker (V3.25) (Liu and Muse, 2005) to calculate major allele frequency, heterozygosity, gene diversity, and PIC (polymorphic information content). The genotypes' genetic distances (Nei et al., 1983) were computed, and a neighbor joining (NJ) tree was generated and viewed in FigTree v 1.4.3 (Rambaut, 2010). To infer historical origin, software STRUCTURE V2.3.1 was used which provides clusters of related genotypes (Pritchard et al., 2000). To infer the value of genetic cluster (K), each sample was run from K = 2 to K = 10 with the admixture model and correlated allele frequency. Each K run was replicated thrice with 100,000 burn-in period and 100,000 Monte Carlo Markov Chain replicates (Evanno et al., 2005). The analysis was carried out regardless of the accessions' geographical origin. The dataset optimal K value was obtained using program "structure harvester" (http://taylor0.biology.ucla.edu). Additional hierarchical structure analysis was done after observing additional peaks at two different K values to unmask the groups (Ambreen et al., 2018). Accessions with membership proportions >80% were considered as pure, whereas membership proportions less than 80% were judged as admixed. The software GenAlEx V6.5 (Peakall and Smouse, 2012) was used to perform principal coordinate analysis (PCoA) and analysis of molecular variance (AMOVA) between the STRUCTURE populations.

Development of core collection

PowerCore 1.0 software (Kim et al., 2007) was used to develop the core. The data set was in simple excel format, and the first column in the data set was given the name % Accessions and the subsequent columns were named NM1, NM2, and so on. After loading the data set in the software, settings were set to heuristic search with maximum possible entries. The diversity index was also mentioned in the same window of the software. There were a number of runs carried out to extract the best possible entries in the core set. Since the number of rice accessions from each state differed ranging from 1,133 accessions from Andhra Pradesh, 378 accessions from Orissa, and 731 accessions from Tamil Nadu, therefore, to avoid individual state collections taking precedence over the core, a separate core subset was developed using the rice collections from each state. Each state accession was analyzed for maximum genetic diversity parameters forming the basis of core subsets. Finally, the core accessions from all the three states were gathered together to develop a core of coastal Indian rice of 247 accessions (126 accessions from Andhra Pradesh, 45 accessions from Orissa, and 76 accessions from Tamil Nadu) (**Supplementary Table S2**). A mini core comprising 30 accessions was also generated using PowerCore (**Supplementary Table S3**). This core collection is of core collection type I (CC-I) giving a uniform representation of the original population. Here, each entry in the core set has one or more accessions that jointly make up the whole collection (Odong et al., 2013). Shannon's diversity index and Nei's gene diversity index were used to evaluate the diversity captured in the core and mini core collection relative to the entire collection using PowerCore. The statistical analysis of the core set and the mini core set was done using PowerMarker (V3.25) (Liu and Muse, 2005) to calculate major allele frequency, gene diversity, heterozygosity, and PIC. This has been done to give a complete depiction of diversification in the total collection and string out the best core set which can stand out as a benchmark collection.

Kinship analysis of the core collection

Kinship analysis was used to determine shared ancestry between individuals in the core collection. Tassel v 5 (Bradbury et al., 2007) was used to create a kinship matrix. All the negative values were considered as zero, and a simple bar graph was prepared using Microsoft Excel. An interactive heat map was prepared using the online available tool https://build.ngchm.net/NGCHM-web-builder, NG-CHM BUILDER: Interactive Heat Map (Ryan et al., 2020).

Validation of 36-Plex SNP assay in east coast rice collection and northeast rice collection

Genetic validation of 36-plex SNP assay (**Table1**) was done by comparing the genetic parameters (e.g., PIC, gene diversity, major allele frequency, and heterozygosity) in northeast rice (Roy Choudhury et al. in 2014) and east coast rice collection. A comparative analysis of these parameters across the two populations was summarized.

RESULTS

Genetic diversity of the total rice collection

Genotyping of the 2,242 rice accessions were performed using 36 SNP markers. The markers generated 72 alleles with a mean of

TABLE 1 | List of SNP primers used for genotyping of 2,242 rice accessions along with gene diversity, heterozygosity, PIC, and major allele frequency.

Chromosome no	Marker name	Physical	Amplification primer1	Amplification primer2	GeneDiversity	Heterozygosity	PIC	Major.Allele.Frquency
1	01-3916-1_C	25381654	ACGTTGGATGGG GTTTGCATGTTA ATAGGG	ACGTTGGATGCC GAATCTCTATCA AGGAAG	0.4700	0.0450	0.3596	0.6224
	01-608-4_C	3421011	ACGTTGGATGAG GACCATCTTCTT GCACTG	ACGTTGGATGCC ATTTGCAAGGCC CATTTC	0.4981	0.1173	0.3740	0.5312
	01-6351-1_C	40914292	ACGTTGGATGGT TGGAACACATGA TTTCAC	ACGTTGGATGAT CTCTTTGGACAG AGTCCC	0.2298	0.0358	0.2034	0.8676
2	02-267_C	1570149	ACGTTGGATGGT CAATCTTGCAGG AGTTGG	ACGTTGGATGTG GCTCCTCTTCTC CGGTCT	0.4060	0.0972	0.3236	0.7168
	02-3029-1_C	18821156	ACGTTGGATGTG TCTGCAATAACT TGTGCC	ACGTTGGATGAA ATCAGCTGCAGC ATTACC	0.4096	0.0864	0.3257	0.7126
	02-4333-1_C	28688819	ACGTTGGATGGG AATGTTTAGTTT TGAGG	ACGTTGGATGTG TAGGTGCTACTT GCTTCC	0.3282	0.0494	0.2743	0.7931
3	03-1691-1_C	10849512	ACGTTGGATGAA CAACGCCAGGAA CATCAC	ACGTTGGATGAA GCGGCTCAAGGT ACAATC	0.3208	0.0389	0.2693	0.7993
	03-3478-1_C	22815422	ACGTTGGATGCC TGCAGCAAACGC CAATTT	ACGTTGGATGTC AGGTAACCGATC GATTTG	0.4964	0.1558	0.3732	0.5425
	03-4660-1_C	31020366	ACGTTGGATGCT CCCATCCTAGTA TCCATC	ACGTTGGATGTG CCTTCTCTTACA GGTTCC	0.3987	0.0816	0.3192	0.7251
4	04-1801-20_C	11859836	ACGTTGGATGCC CTCAAAAAAAAGTTG TAAG	ACGTTGGATGCA GTAAATTTCCAG GGAGATA	0.4208	0.4270	0.3323	0.6990
	04-19-4_C	225838	ACGTTGGATGTC TACACATTAGCT CGCTGG	ACGTTGGATGAC AGTAACCACAAT ATGCCG	0.0198	0.0100	0.0196	0.9900
	04-3787-3_C	25211800	ACGTTGGATGTTATC TCTGCTTGC TCGCTC	ACGTTGGATGAA GTATCTGCCCCA AGTGAC	0.4387	0.0914	0.3425	0.6750
5	05-2692-1_C	18783426	ACGTTGGATGGA ACTTTACTCTCA GTACA	ACGTTGGATGTG GTTTGATGAGTC GTTTGC	0.1981	0.0388	0.1785	0.8885
	05-4192-1_C	28065769	ACGTTGGATGAGTTT GTTGACAGC AGAACC	ACGTTGGATGTA GCTTACTAGTTC ATGTG	0.4947	0.1062	0.3723	0.5515
	05-48-1_C	287362	ACGTTGGATGCA GAGATGTCTGTT GTTAGC	ACGTTGGATGCA ACCAGGGATACA ATATGAC	0.4584	0.1251	0.3533	0.6443
6	06-1256-1_C	7573979	ACGTTGGATGCA CGTGCCTATGAT TAGCAG	ACGTTGGATGGA TCGTTTACTTCT TTGCCC	0.0593	0.0112	0.0576	0.9694
	06-1776-1_C	11093772	ACGTTGGATGGG GCCAATTTGCTT AGTGC	ACGTTGGATGAG CATAAGGTATTA AAGTC	0.2320	0.0698	0.2051	0.8660
	06-2509-1_C	15737387	ACGTTGGATGCC TTCGCGCTTGCA ATTTGG	ACGTTGGATGAA ATCAGCACGCGT CAACAC	0.2046	0.0303	0.1837	0.8843
7	07-2904-39_C	19160255	ACGTTGGATGAA TGGTGGTGTATC TTGAGC	ACGTTGGATGGG TGTGACTTCTCA TGACAG	0.2541	0.0569	0.2218	0.8506
	07-293-12_C	1859603	ACGTTGGATGCA CTAATTCTTGGTATT ATGG	ACGTTGGATGTCAAT GTGTTCTCA CAGACC	0.1173	0.0260	0.1105	0.9374
	07-4304_C	2782410	ACGTTGGATGCA CGTGCCTATGAT TAGCA	ACGTTGGATGGA TCGTTTACTTCT TTGCC	0.4635	0.0592	0.3561	0.6351

(Continued on following page)

TABLE 1 | (*Continued*) List of SNP primers used for genotyping of 2,242 rice accessions along with gene diversity, heterozygosity, PIC, and major allele frequency.

Chromosome no	Marker name	Physical	Amplification primer1	Amplification primer2	GeneDiversity	Heterozygosity	PIC	Major.Allele.Frquency
8	08-2765-2_C	18084851	ACGTTGGATGTC CCTCCATGTTGT GAGTTC	ACGTTGGATGCT TGCAAGAGACAT CCAAGA	0.1636	0.0175	0.1502	0.9101
	08-4218-5_C	27692470	ACGTTGGATGGG TGGACAAAGATA AGGAAG	ACGTTGGATGGA CTGGAAATATAC TCCCTC	0.4658	0.1166	0.3573	0.6307
	08-847-6_C	5399913	ACGTTGGATGCC CAACGTATTAAT GGCAAC	ACGTTGGATGGC TGTGTAGTAATT TGCCTG	0.4754	0.1366	0.3624	0.6109
9	09-209_C	1297966	ACGTTGGATGGA GGCAAAAGGCAA ACCGAC	ACGTTGGATGGA CTTGAGCGAGTC GATGTC	0.2144	0.0419	0.1914	0.8779
	09-2107-5_C	13705487	ACGTTGGATGTG ACCACACCACAC AAACAC	ACGTTGGATGGG GATTTGCGGTTT TTGGAC	0.2831	0.0627	0.2430	0.8294
	09-2716-4_C	19541336	ACGTTGGATGTG AGCCACAGATTC CCTTTC	ACGTTGGATGCT CGAGTAATTCAA AACCAC	0.2056	0.0556	0.1845	0.8836
10	10-1192-7_C	8122635	ACGTTGGATGCT TTGCTACGGATA AAATG	ACGTTGGATGTC ATGCAAATACAG ACATGG	0.4980	0.1228	0.3740	0.5318
	10-188-1_C	1218215	ACGTTGGATGGC GCCAGTGTATGG AAAAAG	ACGTTGGATGGT CCATAACATCAT GGACTC	0.2492	0.0940	0.2182	0.8541
	10-2723_C	20696970	ACGTTGGATGCC CACAATGAGATG CAGATG	ACGTTGGATGAG ACAAAATGCAAC ACTCCG	0.0754	0.0538	0.0725	0.9608
11	11-1849_C	11974790	ACGTTGGATGCG CCACTCTTCCTG ATTTAG	ACGTTGGATGAC AGATACGGGAGG CATTTC	0.1747	0.0487	0.1594	0.9033
	11-3935_C	28434679	ACGTTGGATGAT CCCTGAGACTTT GGATGG	ACGTTGGATGCC AACTTGAATGTC CATTCC	0.1239	0.0273	0.1162	0.9337
	11-522-1_C	3033366	ACGTTGGATGCT ACATGGTATCAG ATACCG	ACGTTGGATGAG AAGCGAACGCGG AAAAAG	0.4596	0.0971	0.3540	0.6421
	12-1794_C	11215946	ACGTTGGATGGT GAGCCCCAAAAG TTGGTG	ACGTTGGATGTA AGGTCCAGTTTG CTTGGT	0.0287	0.0094	0.0283	0.9855
12	12-3200-2_C	21396181	ACGTTGGATGGC TCAAACCTAGCAATA ACTG	ACGTTGGATGCC TCCTTCCTACAA GTTTAA	0.0974	0.0314	0.0927	0.9487
	12-400_C	2160546	ACGTTGGATGCC AATAGAGTCCAT CTCAGC	ACGTTGGATGGC ACGAGGATTTAA GACAGC	0.2585	0.0990	0.2251	0.8475
	Mean				0.2970	0.0770	0.2412	0.7848

two alleles per locus for the entire 2,242 coastal collection (**Table 1**). The maximum PIC was 0.37 for markers 01-608-4_C and 03-3478-1_C, and the minimum was 0.01 for marker 04-19-4_C with an average value of 0.24. The maximum and minimum heterozygosity was 0.42 for marker 04-1801-20_C and 0.009 for marker 12-1794_C, respectively, with a mean value of 0.07. Likewise, maximum and minimum gene diversity was found to be 0.49 for marker 01-608-4_C and 0.01 for marker 04-19-4_C, respectively, with an average of 0.29. The maximum major allele frequency was 0.99 for marker 04-19-4_C, and the minimum major allele frequency was observed to be 0.53 for marker 01-608-4_C with a mean value of 0.78 (**Table 1**).

Phylogenetic analysis of the total rice collection

Cluster analysis of 2,242 rice accessions was performed and PowerMarker (v3.25) was used to determine the dissimilarity matrix which was used to construct an unrooted phylogenetic tree using FigTree v1.4.3 (**Supplementary Figure S1**). Total collection got grouped into three major groups. However, no differentiation

TABLE 2 | List of genetic diversity parameters estimated for three east coast states, core, and mini core set.

	Sample size	Major.Allele.Frquency	Gene diversity	Heterozygosity	PIC
Andhra Pradesh	1133	(0.55–0.98) 0.78	(0.02–0.49) 0.30	(0.008–0.68) 0.09	(0.02–0.37) 0.25
Andhra Pradesh Core	126	(0.52–0.92) 0.74	(0.13–0.49) 0.36	(0.04–0.69) 0.15	(0.12–0.37) 0.29
Orissa	378	(0.52–1.0) 0.80	(0.00–0.49) 0.26	(0–0.12) 0.05	(0–0.37) 0.21
Orissa core	45	(0.5–1.0) 0.77	(0.0–0.5) –0.32	(0.0–0.18) 0.08	(0.0–0.37) 0.26
Tamil Nadu	731	(0.55–0.98) 0.82	(0.02–0.49) 0.24	(0.008–0.68) 0.07	(0.02–0.37) 0.20
Tamil Nadu core	76	(0.50–0.97) 0.79	(0.05–0.49) 0.29	(0.020–0.33) 0.10	(0.05–0.37) 0.23
core	247	0.7547	0.3417	0.1231	0.2759
Mini core	30	0.68	0.4	0.2	0.32

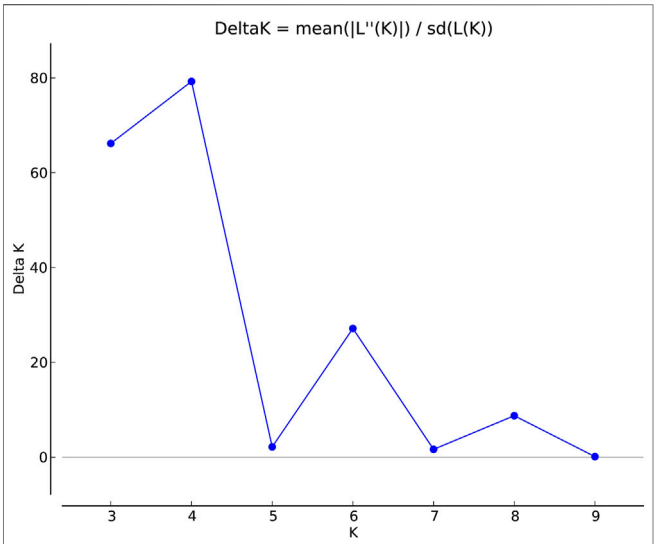


FIGURE 1 | Estimation of population using LnP(D) derived Δk for k from 2 to 10 of a total 2,242 east coast rice collection.

could be made between these groups based on their geographical origin with each group displaying a heterogeneous clustering of individuals.

Genetic diversity of the rice collection of the three coastal states (Andhra Pradesh, Orissa, and Tamil Nadu)

The rice collections of coastal states belonging to the east coast of India were analyzed for the genetic diversity and population structure study. The lowest PIC recorded was 0.0 for marker 04-19-4_C from the state of Orissa, and the highest was 0.37 for marker v03-3478-1_C from the state of Andhra Pradesh. Heterozygosity was observed to be 0.0 for three markers 04-19-4_C, 11-3935_C, and 12-3200-2_C from Orissa state, and the highest value observed was 0.69 for marker 04-1801-20_C from the state of Andhra Pradesh. The lowest value of genetic diversity was found to be 0.0 for marker 04-19-4_C, and the highest was 0.49 for marker 02-267_C both from Orissa. The lowest and highest major allele frequencies recorded were 0.52 for 02-267_C and 1.0 for 04-19-4_C, respectively, from Orissa state (**Table 2**). For the collections from three states, phylogenetic analysis was done using the neighbor joining (NJ) method, and an unrooted NJ tree was created. Rice collection from the state of Andhra Pradesh got clustered into two major groups. Group 1 had 220 accessions, and group 2 had 913 accessions (**Supplementary Figure S2**). There were two major groups each in case of rice collection from Orissa (**Supplementary Figure S3**) and Tamil Nadu (**Supplementary Figure S4**). The NJ tree of rice collection of Orissa exhibited 7 accessions in group 1, and 371 accessions got clustered in group 2. The NJ tree of rice collection of Tamil Nadu exhibited 5 accessions in group 1 and 726 accessions in group 2.

Population structure of the total rice collection

The genetic link between individual rice accessions was determined using STRUCTURE, a model-based tool. Each accessions' membership was run from K = 2 to K = 10 (**Figure 1**). The ultimate number of populations was determined using Structure Harvester (http://taylor0. biology. ucla.edu). The number of populations was found to be four for the entire 2,242 east coast rice collection (**Figure 2**). Population 1 had 128 pure and 51 admix accessions. Population 2 showed 821 pure and 107 admixed accessions, population 3 showed 157 pure and 78 admixed accessions, and population 4 showed 700 pure and 200 admixed accessions. Most of the aromatic rice accessions got grouped in population 3. Such grouping of aromatic accessions was not seen in the NJ tree. An overview of the state-wise distribution in population shows that population 1 had 60% accessions from Orissa, while in populations 2 and 3, 72% and 93% accessions were from Andhra Pradesh, respectively, fairly exhibiting the dominance of states over population. Population 4 showed around 50% accessions from Tamil Nadu (**Figure 2**). Likewise, the population mean value of alpha, Fst1, Fst2, Fst3, and Fst4 generated from the model-based approach and allele-freq. divergence among populations (net nucleotide distance), computed using point estimates of population obtained using the model-based approach, is given in **Supplementary Table S4** and **Supplementary Table S5**. Fst values showed good genetic differentiation and acceptable population structure. Venn diagrams between NJ tree and population structure was constructed to find co-linearity between them. There were 21 accessions (0.93%) overlapping between population 1 of STRUCTURE and group 2 of the NJ tree. A total of 254 accessions (11%) were found to be overlapping in population 2 of STRUCTURE and group 2 of the NJ tree. Similarly, 225

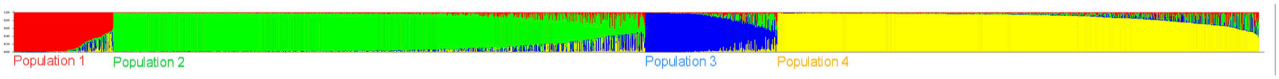


FIGURE 2 | Population structure bar plot of 2,242 east coastal rice accessions.

accessions (10%) were common in population 3 of STRUCTURE and group 3 of NJ tree, and 754 accessions (33%) were common in population 4 of STRUCTURE and group 3 of the NJ tree. This shows less similarity being observed between groups of the NJ tree and populations of the model-based approach (**Supplementary Figures S5–S8**).

Clustering of accessions in STRUCTURE was diverse. Apart from obtaining the highest peak at K = 4, additional smaller peaks were obtained at K = 6 and K = 8 (**Figure 1**), implying that there are subgroups within the four major groups. As a result, an independent STRUCTURE was run with K values from 2 to 10 for all four populations obtained above. Subclustering of population 1 identified the highest peak at K = 4, giving subpopulations named subpopulation 1a, subpopulation 1b, subpopulation 1c, and subpopulation 1d (**Supplementary Figures S9, S10**). Subclustering of population 2 identified the highest peak at K = 6, giving subpopulations named subpopulation 2a, subpopulation 2b, subpopulation 2c, subpopulation 2d, subpopulation 2e, and subpopulation 2f (**Supplementary Figures S11, S12**). Subclustering of population 3 also identified the highest peak at K = 6, giving subpopulations named subpopulation 3a, subpopulation 3b, subpopulation 3c, subpopulation 3d, subpopulation 3e, and subpopulation 3f (**Supplementary Figures S13, S14**). Subclustering of population 4 identified the highest peak at K = 3, giving subpopulations named subpopulation 4a, subpopulation 4b, and subpopulation 4c (**Supplementary Figures S15, S16**). Grouping in Structure was diverse, and no dominance of states was observed in any of the population. The allocation of accessions from different states and the expected heterozygosity of the 19 subpopulations are shown in **Supplementary Table S6**. In this table, there are slight variations in the values of the expected heterozygosity ranging between 0.06 and 0.40 with a mean value of 0.22, indicating good genetic diversity in the subpopulation (Luo et al., 2019).

Population structure of the rice collection of three coastal states (Andhra Pradesh, Orissa, and Tamil Nadu)

Another aspect of population structure was studied to see the clustering of accessions state-wise. Population structure analysis grouped rice collection from Andhra Pradesh into four different populations (**Supplementary Figure S17**) Population 1 has 150 pure and 40 admix accessions, population 2 has 260 pure and 145 admix accessions, population 3 has 150 pure and 107 admix accessions, and population 4 has 179 pure and 97 admix accessions (**Supplementary Figure S18**). All aromatic accessions from Andhra Pradesh got grouped in population 3. Rice collection from Orissa got grouped into three populations (**Supplementary Figure S19**) with population 1 having 62 pure and 24 admix accessions, population 2 having 96 pure and 6 admix accessions, and population 3 having 156 pure and 34 admix accessions (**Supplementary figure S20**). Rice collection from Tamil Nadu got grouped into five populations (**Supplementary Figure S21**), population 1 having 44 pure and 20 admix accessions, 73 pure and 41 admix accessions in population 2, 168 pure and 42 admix accessions in population 3, 95 pure and 67 admix accessions in population 4, and 104 pure and 77 admix accessions in population 5 (**Supplementary Figure S22**). The mean values of alpha, Fst1, Fst2, Fst3, Fst4, Fst5, and Allele-freq. divergence among populations (net nucleotide distance), computed using point estimates of population generated from the model-based approach, are shown in **Supplementary Table S7** and **Supplementary Table S8**, respectively. The values of Fst showed standardized genetic differentiation, suggesting a good population structure. Allele frequency divergence among populations computed using the point estimates of the population also gives about the genetic variation among populations. The values are indicative of the accessions being diverged in the population structure (Luo et al., 2019).

AMOVA and PCoA of total 2,242 east coast rice collection

The distribution of genetic diversity between and within the populations obtained following STRUCTURE analysis was investigated using AMOVA for total rice accessions (2242). In the first case, four populations were assumed and AMOVA analyses revealed that 29% diversity exists among populations, 20% within individuals, and 51% among individuals of the total east coast rice collection (**Supplementary Table S9**), while the PCoA plot showed that out of four populations obtained population 3 was getting distinctly separated from the others (**Supplementary Table S10**; **Supplementary Figure S23**). In the second case, assuming 19 subpopulations when AMOVA was done there was 30% diversity existing among populations, 21% within individuals, and 49% among individuals of the total east coast rice collection (**Supplementary Table S11**), while the PCoA plot showed plots with overlapping populations (**Supplementary Table S12**; **Supplementary Figure S24**). The AMOVA and PCoA studies confirmed that the actual population number in the case of the east coast collection is only four because assumption of the population of 19 did not show any advantage.

AMOVA and PCoA of the rice collection coastal states

The AMOVA study of Andhra Pradesh revealed 23% variance within individuals and 30% and 47% variance among populations

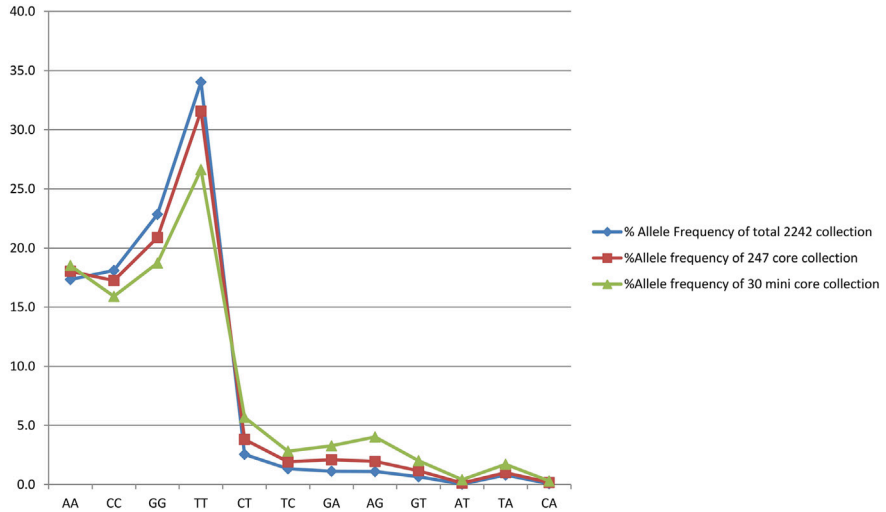


FIGURE 3 | Graph showing the allele frequency of total east coast collection, core collection, and mini core collection.

and among individuals, respectively (**Supplementary Table S13**; **Supplementary figure S25**). PCoA plot analysis revealed that population 3 is distinctly isolated from the rest of the populations, and all the aromatic samples from Andhra Pradesh got grouped in to this **Supplementary figure S26**. Likewise, the AMOVA study of Orissa showed 19%, 68%, and 14% among populations, among individuals and within individuals, respectively (**Supplementary Table S13**; **Supplementary figure S27**). The PCoA plot showed mixing of three populations (**Supplementary figure S28**). The AMOVA study of Tamil Nadu showed 24%, 53%, and 23% variance among populations, among individuals, and within individuals, respectively (**Supplementary Table S13**; **Supplementary figure S29**). The PCoA plot revealed slightly isolated population 3 (**Supplementary figure S30**); however, there was intermixing between populations 1, 2, 4, and 5. Principal coordinate analyses of rice collection of coastal states, with percentage of variation explained by the first three axes, are summarized in **Supplementary Table S14**.

Generation of the core set

Out of 2,242 rice accessions studied, a core set of 247 accessions (i.e., 126 accessions from Andhra Pradesh, 45 accessions from Orissa, and 76 accessions from Tamil Nadu) and a mini core set of 30 accessions were selected using POWERCORE (**Supplementary Tables S2, S3**). Thirty-six SNP markers produced nine allele types, four of which were homozygous and six were heterozygous (three transitions and two transversions). We found no C/G- or G/C-type substitutions in our research. Allele frequency was determined for all three state collections and their core sets. The study of allele frequency revealed that no alleles were lost in the resulting core set and they were 99.9% similar. The same has been plotted in line plots for the total 2,242 in the east coast collection as well as for the core sets (**Figure 3**). These results were also in concordance with the results obtained by PowerCore where the Shannon's diversity index and Nei's gene diversity showed an increasing trend starting from the entire 2,242 collection to 247 core to 30 mini core collections (**Figure 4**). The 247 core accessions represent 11.01% of the entire collection, and the 30 mini core set represents 12.14% of the core collection (**Supplementary Figure S31**). Thus, we have been able to fulfill the recommended size required for a perfect core collection, which is between 5% and 20% of the original size (Bisht et al., 1998).

Genetic diversity of the core set and the mini core set

The degree of genetic diversity of the core set was studied to find the degree of genetic diversity captured from the overall coastal collection. Major allele frequency, gene diversity, heterozygosity, and PIC were observed to be 0.75, 0.34, 0.12, and 0.27 respectively (**Table 2**). The values of genetic diversity of the mini core set were approximately similar to the core, as shown in **Table 2**. A greater value genetic diversity has been observed in case of core/mini core than that of the total east coast rice collection. Hence, definitely the east coast core collection developed has rich and diverse representatives having good diversity parameters. The NJ tree showed two distinct groups in the core set and three groups in the mini core set (**Figures 5**, **6**).

Population structure of core set

The population structure grouped the core set accessions into four populations (**Figures 7**, **8**). The STRUCTURE bar plot showed population 1 having 27 pure and 20 admixed accessions, population 2 having 55 pure and 22 admixed accessions, population 3 having 20 pure and 7 admixed accessions, and population 4 having 60 pure and 36 admixed accessions (**Figure 8**). The mean value of alpha (0.10) for the core collection is greater than the mean value of alpha for the total collection (0.06). An alpha value close to zero means that individuals are essentially from different populations (Evanno et al., 2005). In our case, a 0.10 value of alpha in the core collection as compared to the 0.06 value of alpha in the total

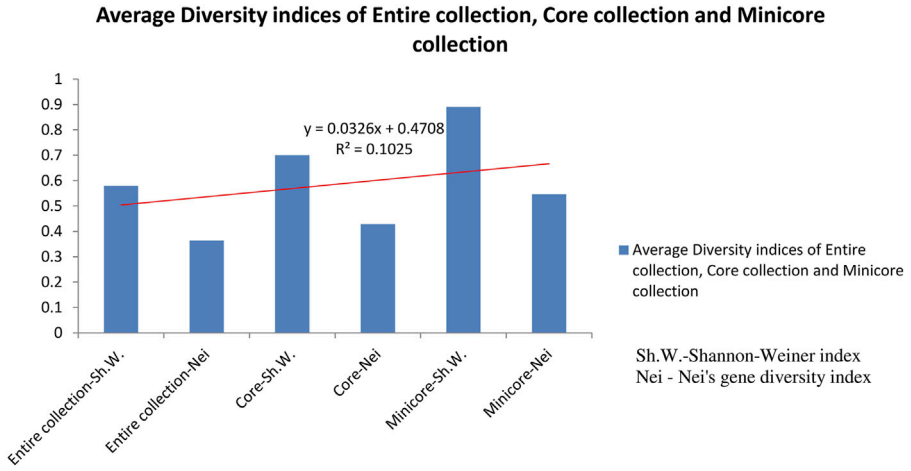


FIGURE 4 | Graph showing average diversity indices of entire collection, core collection, and mini core collection.

collection signifies more admixed individuals in the total collection. Allele-freq. divergence among populations (net nucleotide distance), computed using point estimates of population (core collection), are given in **Supplementary Table S15** and **Supplementary Table S16**.

AMOVA and PCoA of the core set

The AMOVA study of the core collection revealed 29% variance within individuals as well as among population and 42% variance among individuals (**Supplementary Table S17**; **Supplementary Figure S32**). The PCoA plot showed populations being scattered in different quadrants (**Supplementary Figure S33**; **Supplementary Table S18**).

Kinship analysis of the core set collection

Kinship analysis of the core set showed that more than 50% of the samples had a kinship value less than zero, and less than 10% of the accessions had kinship values between 0.5 and 0.75 (**Supplementary Figure S34**). The kinship index and clustered heat map showed more diversity in the core collection because clustering based on this map was more heterogeneous, which indicates that maximum unique genotypes have been selected in the core collection (**Figure 9**).

Validation of SNP markers in coastal rice collection and northeast rice collection

A comparative analysis of genetic diversity parameters, i.e., PIC and gene diversity between the coastal rice and northeast rice collection (Roy Choudhury et al., 2014), was done and are shown in **Figure 10**. On comparing the values, in both collections it was observed that in the case of the PIC, the same markers had given the maximum and minimum values in both collections (i.e., coastal rice collection and northeast rice collection). This means that markers 01-608-4_C and 03-3478-1_C had given the maximum PIC value of 0.37 in both the collections and marker 04-19-4_C had given the minimum PIC value of 0.01 in both collections. Similarly, markers 01-608-4_C and 03-3478-1_C had displayed the highest gene diversity with a value of 0.49 across both collections and a minimum value of gene diversity 0.02 with marker 04-19-4_C across both collections. Generally, line graphs for PIC and gene diversity were overlapping for both collections except at certain points where deviations were observed. For example, marker 03-1691-1_C gave a gene diversity value of 0.32 in the coastal rice collection and 0.4 in the northeast rice collection. The same marker gave a PIC value of 0.26 in the coastal rice collection and 0.32 in the northeast rice collection. Marker 11-522-1_C gave a gene diversity value of 0.45 in the coastal collection and 0.19 across the northeast collection; also, this marker 11-522-1_C gave PIC values of 0.35 and 0.17 across the coastal rice and northeast rice collections. A subsequent analysis between major allele frequency and heterozygosity from the current study and with the northeast rice collection was also evaluated (**Supplementary figure S35**). The values for major allele frequency and heterozygosity were overlapping except at few points where deviation was observed. For example, marker 04-1801-20_C gave heterozygosity values of 0.42 and 0.28 in coastal collection and northeast collection, respectively depicting a small amount of deviation. Similarly, marker 11-522-1_C gave major allele frequency values of 0.64 and 0.89 in the coastal collection and northeast collection, respectively, again depicting slight deviations. The validation of the same set of SNP markers (36-plex assay) in different collections (northeast and east coast collection) shows that they are very effective in deciphering the genetic diversity parameters in both the collections.

DISCUSSION

The east coast collection of rice germplasm available at the National Genebank, NBPGR, New Delhi, is a valuable collection of rice accessions for the assessment of genetic diversity and other important traits. Rice accessions from different collections have served as and continued to act as sources of genes for desired qualities, contributing to the variety developments that have been reported (Choudhury et al., 2013; Das et al., 2013). Regular floods hit coastal areas

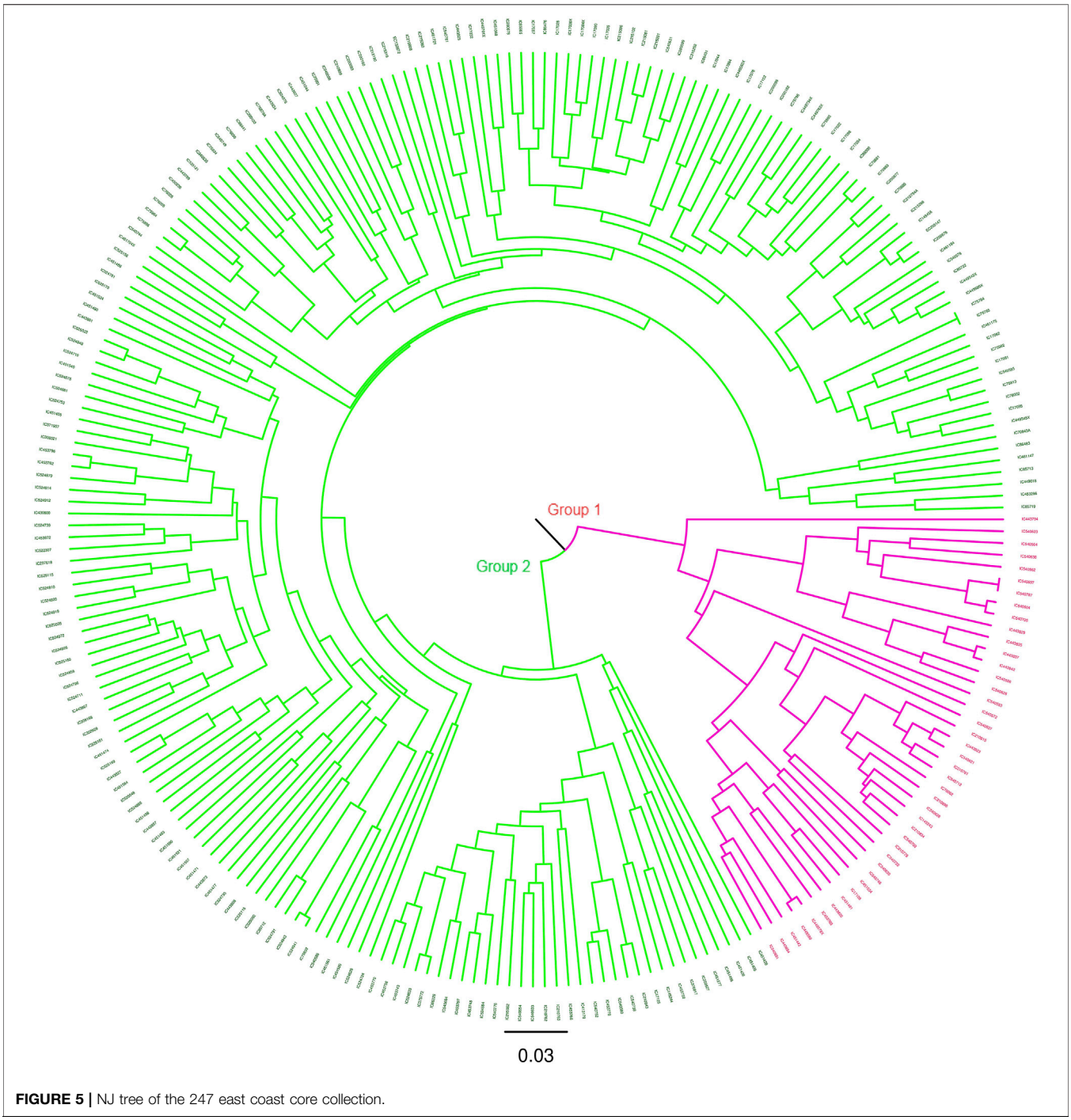


FIGURE 5 | NJ tree of the 247 east coast core collection.

and have saline soils, and various other constraints make them a fragile ecosystem with lower productivity and slow trend in growth rate (Amanullah et al., 2007). Therefore, a coastal core collection would be an appropriate measure to conserve rice in these areas for better management studies. Despite advances in genomics, the Indian rice collection has remained uncharacterized at the molecular level in terms of genetic diversity and population structure. This has been a key stumbling block in their ability to use and develop superior cultivars. In this study, effort has been made to characterize the east coast rice collection available at the National Genebank, NBPGR, New Delhi, using 36 SNP markers to enhance genome wide studies in rice. These unlinked SNP markers, which were generated and used in diverse studies, are located on the short arm, centromeric region, and long arm of all 12 rice chromosomes. As discussed, a state-wise study of the east coast collection of rice showed interesting results. A total of 72 alleles were amplified with 2 alleles per locus. The average PIC

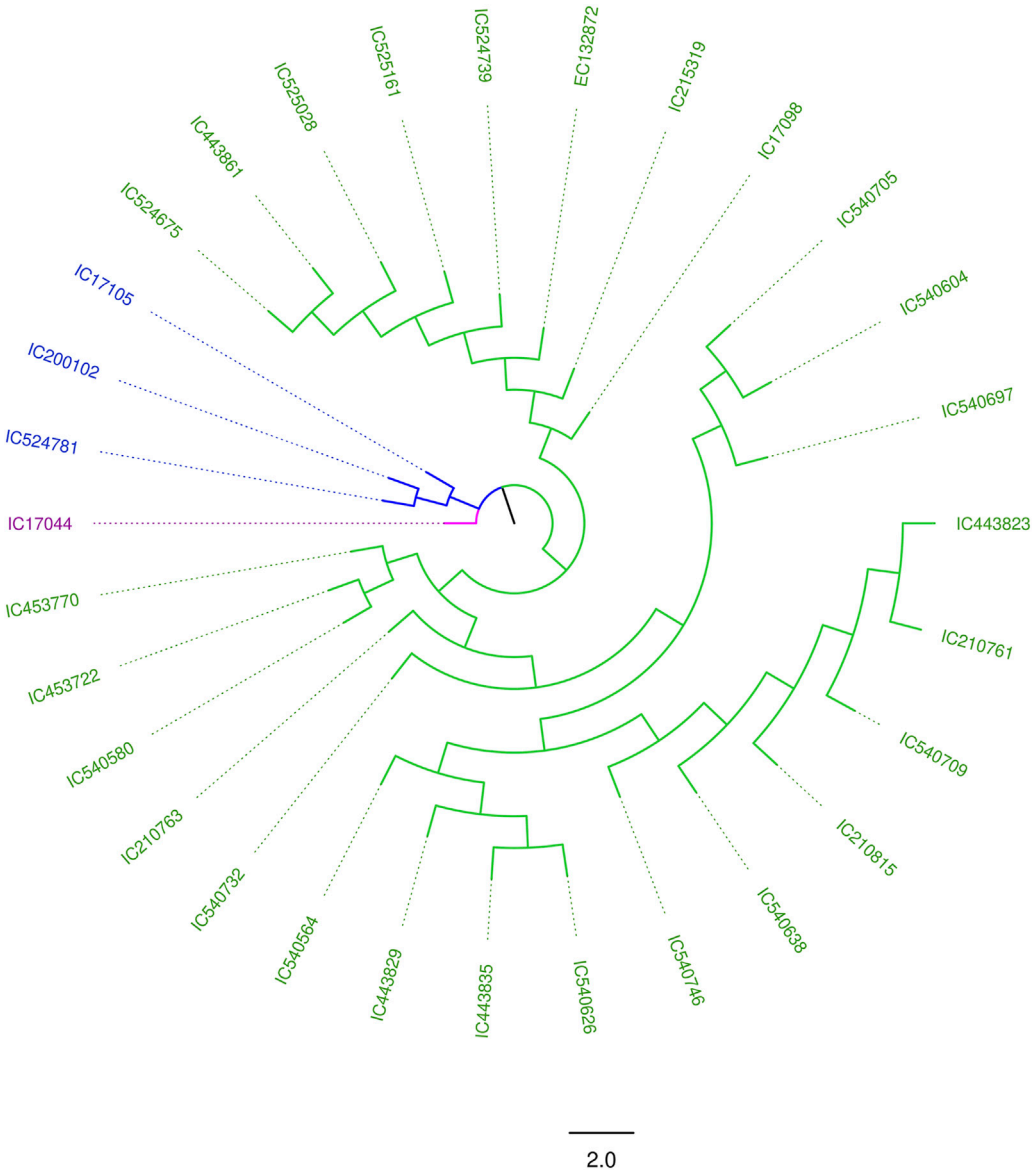


FIGURE 6 | NJ tree of the east coast mini core set.

values ranged from 0.20 for Tamil Nadu, 0.21 for Orissa, and 0.25 for Andhra Pradesh. The values observed are concurrent with those observed by Singh et al. (2013) on 375 rice varieties (0.25) and Roy Choudhury et al. (2014) on the northeast rice collection (0.23) using SNP markers. Chen et al. (2017) observed PIC values of 0.27 on *Ziziphus jujuba* Mill, China's most important fruit species, and 0.29 reported by Luo et al. (2019) on *Camelina sativa* using SNP markers. A PIC value of 0.4 was reported by Mourad et al. (2020) while they were studying the genetic diversity, population structure, and linkage disequilibrium in the spring wheat core collection, which is higher than that reported in the present study. The PIC value is generally high when SSR markers are used as observed by Rashmi et al. (2017) in 65 rice accessions (0.38) characterized using SSR markers. Pathaichindachote et al. (2019) reported a PIC value of 0.56 in 167 Thai and exotic rice varieties using 49 SSR markers and a PIC value of 0.63 reported by Jasim Aljumaili et al. (2018) on 50 aromatic rice accessions with 32 SSR markers. A mean PIC value of 0.61 has been reported by Suvi et al. (2020) while they were accessing the genetic diversity and population structure of 54 rice accessions using 14 SSR markers. Tarang et al. (2020) reported a PIC value of 0.92 with 60 microsatellite markers of 63 rice genotypes in Central and West Asia. PIC values and expected heterozygosity (He, also called gene diversity) are both indices of genetic diversity among genotypes in breeding populations. This also reveals the evolutionary pressure on the alleles as well as the mutation rate a locus may have experienced over time (Botstein et al., 1980; Shete et al., 2000; Luo et al., 2019). In our study, the average

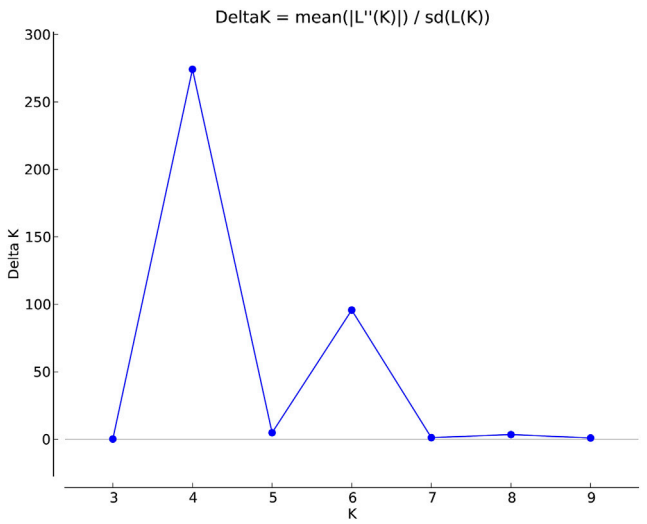


FIGURE 7 | Estimation of population using LnP(D) derived Δk for k from 2 to 10 of the east coast core collection.

gene diversity was observed to be 0.24, 0.26, and 0.30 for Tamil Nadu, Orissa, and Andhra Pradesh, respectively (**Table 2**). As a result, the overall gene diversity value was slightly higher than the PIC value, which was expected. The PIC values will always be lower than gene diversity and will become closer to gene diversity as more alleles are added and with rising evenness of allele frequencies (Shete et al., 2000). PIC values are limited to 0.5 due to the biallelic nature of the SNPs (where the two alleles have identical frequencies) (Eltaher et al., 2018) and could possibly be attributable to low mutation rates in SNPs (Coates et al., 2009; Eltaher et al., 2018; Luo et al., 2019). There are some accessions in the total list of accessions which have the same IC numbers followed by X and P; these are accessions which were collected at two different periods from the same areas, hence denoted by X or P. It has been distinctly noticed that these accessions and their original counterparts did not give the same results with SNP markers; this could be due to the high evolutionary drive during collection at different periods. (Kasso and BalaKrishnan, 2013).

The genetic distances were estimated, and the dissimilarity matrix was used to build the NJ tree. The NJ tree of the 2,242 coastal samples showed three major groups. However, nothing very captivating was observed in the clusters formed. Such widely overlapped groups in the NJ tree has also been reported by Xu et al. (2016) on *indica* rice.

Initially, the population structure of the overall east coast collections revealed four populations. A subsequent population structure analysis gave 19 populations altogether. A weak population structure and low relatedness as revealed in kinship analysis between the east coast rice accessions and the core accessions support the statement by Nachimuthu et al. (2015) that these are critical factors to circumventing spurious data hindering the downstream study (). In the present study, the NJ tree, Bayesian-based STRUCTURE, and AMOVA and PCoA did not show any consensus clustering that could be highlighted. Similar results were also observed by Ambreen et al. in 2018. Also, a population structure study on the rice collection of the east coast states revealed four, three, and five populations for Andhra Pradesh, Orissa, and Tamil Nadu, respectively. In case of Andhra Pradesh, population structure showed a conspicuous grouping of aromatic samples in population 3. Similar type of grouping was reported in basmati rice by Civáň et al. (2019). Admixtures were observed, which suggests that besides pure lines there are samples which are heterogeneous in nature. The mean values of alpha ranged from 0.04 for Orissa, 0.05 for Tamil Nadu, and 0.10 for Arunachal Pradesh (**Supplementary Table S4**). When the alpha value approaches zero, it means that the majority of individuals are from distinct populations (Li et al., 2014). The values of Fst correspond to a standardized genetic differentiation, suggesting an acceptable population structure. The STRUCTURE analysis indicated good genetic diversity among the rice accessions of the east coast collection. The presence four populations were confirmed by model-based analysis. This method has been used extensively by Edae et al. (2014) to explore association mapping.

The advanced M strategy with minimum redundancy and heuristic approach was used for east coast core collection. The minimum redundancy is required for increasing allelic richness in the core collection; hence, accessions need to be of unique allelic combinations and an unstructured population (Ambreen et al., 2018). A kinship analysis study of the east

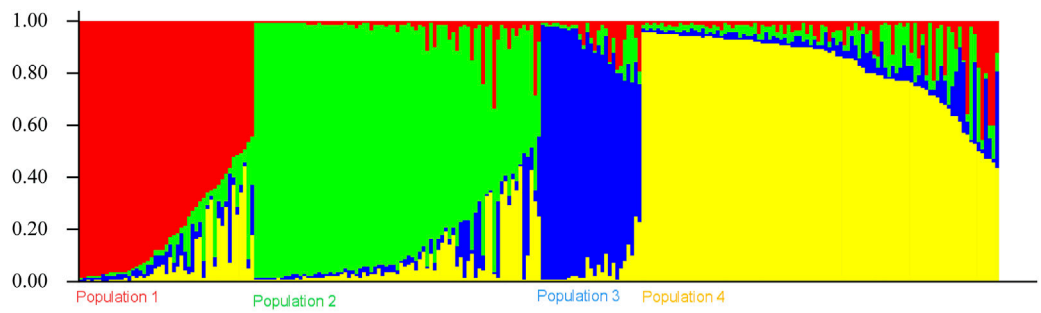


FIGURE 8 | Population structure bar plot of the east coast core collection.

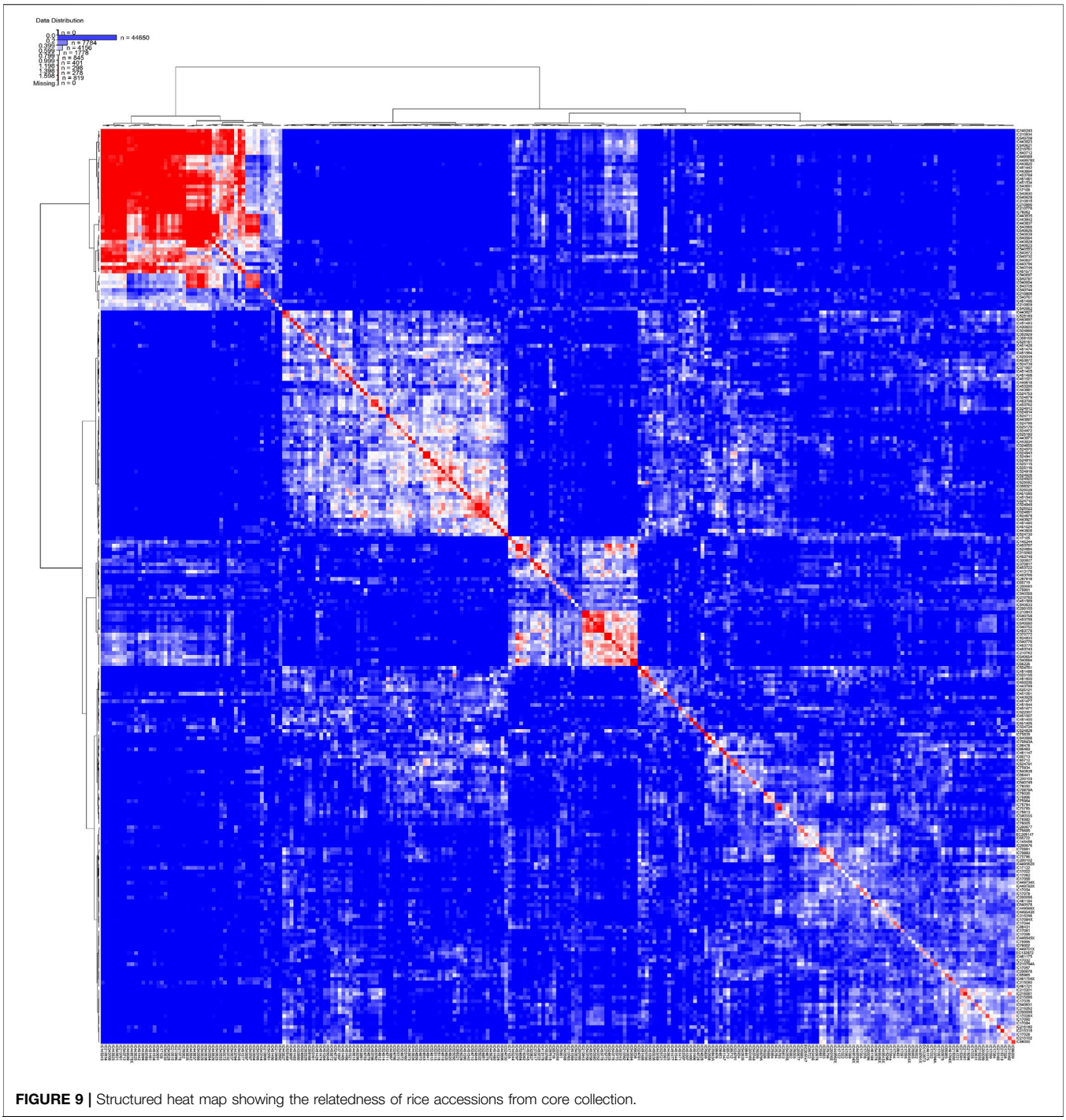


FIGURE 9 | Structured heat map showing the relatedness of rice accessions from core collection.

coast core collection demonstrated a low amount of genetic relatedness, meeting the key condition of an ideal core collection as well as an idealistic association panel (Kumar et al., 2020). Thus, this coastal core set qualifies all the benchmarks of a standard core set.

While validating our results with the northeast rice collection (Roy Choudhury et al., 2014), markers 01-608-4_C and 03-3478-1_C were found to give the highest PIC value of 0.37, while marker 04-19-4_C was least informative giving the lowest PIC of 0.01 in both collections. Marker 01-608-4_C is a locus of evolutionary conserved genes from the saponin family of proteins, which is involved in the sphingolipid metabolic process and active in extracellular space (Bruhn, 2005); in the present study, this marker is highly conserved yet highly polymorphic. However, marker 03-3478-1_C, which is a locus of the GRP (gibberellin-regulated protein) family and an evolutionary conserved gene involved in plants' defense mechanism as well as in growth (Inomata, 2020), has not

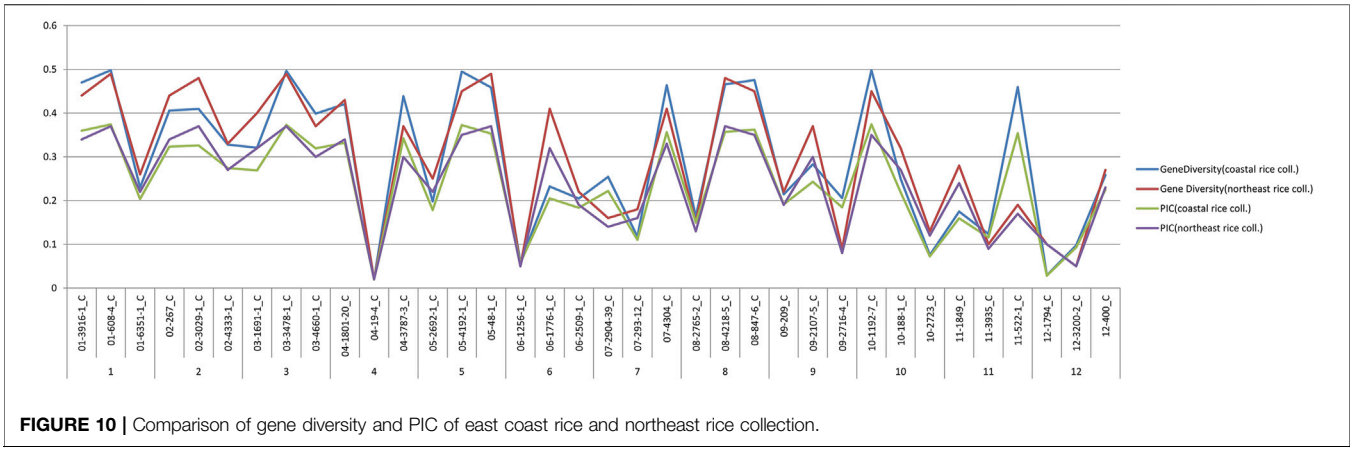


FIGURE 10 | Comparison of gene diversity and PIC of east coast rice and northeast rice collection.

been observed polymorphic in both collections, which is contradictory to earlier an report by Mukesh Jain et al. (2014) in rice.

Validation of the same set of SNP markers on two collections (northeast rice and east coast rice collection) has established that the 36-plex SNP assay is sufficient and efficient for initial diversity analysis and core development. Hence, this 36-plex SNP assay can be exploited by researchers for the genetic diversity study and development of core based on their own collections, thus accelerating their breeding program.

CONCLUSION

This is the first study where India's east coast rice collections were characterized using SNP markers. The genetic diversity and population structure were studied, and core and mini core collections with maximum diversity and minimum redundancy were developed. A total of 2,242 east coast rice accessions from three different states of India, i.e., Andhra Pradesh, Orissa, and Tamil Nadu, have been characterized, and a wide range of gene diversity and PIC was observed. A phylogenetic analysis of the total east coast rice collection revealed three groups, and a population structure analysis revealed four populations. The 36-SNP assay used in this study was validated by comparing the genetic diversity parameters (gene diversity, PIC, major allele frequency, and heterozygosity) across two different rice collections, i.e., east coastal rice and northeast rice collection, and it was observed the these markers were sufficient to decipher all genetic parameters very efficiently; hence, they can be effectively utilized for core development and diversity study of different rice genotypes.

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

RS conceived and designed the experiments; DC and RK performed the experiments; RS and DC analyzed the data; VS and RS contributed reagents/materials/analysis tools; DC and RS contributed to the writing of the manuscript; and KS and NS edited the manuscript.

FUNDING

This work was supported by ICAR grant for the project Network Project on Functional Genomics and Genetic Modification in crops.

ACKNOWLEDGMENTS

We are thankful to Avantika Maurya and Shantanu Das for their help in kinship analysis. We are grateful and thankful to the Director, NBPGR, New Delhi, who provided facilities for this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.726152/full#supplementary-material

Supplementary Figure S1 | Neighbor joining tree of the total 2242 east coast rice collection

Supplementary Figure S2 | Neighbor joining tree of 1133 rice collection of Andhra Pradesh

Supplementary Figure S3 | Neighbor joining tree of 378 rice collection of Orissa

Supplementary Figure S4 | Neighbor joining tree of 731 rice collection of Tamil Nadu

Supplementary Figure S5 | Venn diagram showing co-linearity between all three groups of neighbor joining tree and population 1 of population structure of total east coast rice collection

Supplementary Figure S6 | Venn diagram showing co-linearity between all three groups of neighbor joining tree and population 2 of population structure of total east coast rice collection

Supplementary Figure S7 | Venn diagram showing co-linearity between all three groups of neighbor joining tree and population 3 of population structure of total east coast rice collection

Supplementary Figure S8 | Venn diagram showing co- linearity between all three groups of neighbor joining tree and population 4 of population structure of total east coast rice collection

Supplementary Figure S9 | Hierarchical population structure analysis of subpopulation of population 1 total east coast rice collection

Supplementary Figure S10 | Bar plot of population 1 highlighting the sub populations in east coast rice collection

Supplementary Figure S11 | Hierarchical population structure analysis of subpopulation of population 2 total east coast rice collection

Supplementary Figure 12 | Bar plot of population 2 highlighting the sub populations in east coast rice collection

Supplementary Figure S13 | Hierarchical population structure analysis of subpopulation of population 3 total east coast rice collection

Supplementary Figure S14 | Bar plot of population 3 highlighting the sub populations in east coast rice collection

Supplementary Figure S15 | Hierarchical population structure analysis of subpopulation of population 4 total east coast rice collection

Supplementary Figure S16 | Bar plot of population 4 highlighting the sub populations in east coast rice collection

Supplementary Figure S17 | Estimation of population using LnP(D) derived Δk for k from 2 to10 of Andhra Pradesh rice collection

Supplementary Figure S18 | Bar plot of population structure of Andhra Pradesh rice collection

Supplementary Figure S19 | Estimation of population using LnP(D) derived Δk for k from 2 to10 of Orissa rice collection

Supplementary Figure S20 | Bar plot of population structure of Orissa rice collection

Supplementary Figure S21 | Estimation of population using LnP(D) derived Δk for k from 2 to10 of Tamil Nadu rice collection

Supplementary Figure S22 | Bar plot of population structure of Tamil Nadu rice collection

Supplementary Figure S23 | Scattered plot (PCoA) of total east coast rice accessions (2242) (among 4 populations)

Supplementary Figure S24 | Scattered plot (PCoA) of total east coast rice accessions (2242) (among 19 sub populations)

Supplementary Figure S25 | Pie chart showing percentage of molecular variance of rice collection of Andhra Pradesh

Supplementary Figure S26 | Scattered plot (PCoA) of rice collection of Andhra Pradesh

Supplementary Figure S27 | Pie chart showing percentage of molecular variance of rice collection of Orissa

Supplementary Figure S28 | Scattered plot (PCoA) of rice collection of Orissa

Supplementary figure S29 | Pie chart showing percentage of molecular variance of rice collection of Tamil Nadu

Supplementary Figure S30 | Scattered plot (PCoA) of rice collection of Tamil Nadu

Supplementary Figure S31 | Venn Diagram showing the distribution of accessions in the east coast core and the mini-core collection

Supplementary Figure S32 | Pie chart showing percentage of molecular variance of the east coast rice core set (247)

Supplementary Figure S33 | Scattered plot (PCoA) of the east coast rice core set (247)

Supplementary Figure S34 | Histogram showing the kinship status of rice accessions in the east coast core collection

Supplementary Figure S35 | Comparative analysis of heterozygosity and major allele frequency values across eat coast rice and north-east rice collection

REFERENCES

Agrama, H. A., Yan, W., Lee, F., Fjellstrom, R., Chen, M.-H., Jia, M., et al. (2009). Genetic Assessment of a Mini-Core Subset Developed from the USDA Rice Genebank. *Crop Sci.* 49, 1336–1346. doi:10.2135/cropsci2008.06.0551

Amanullah, M. M., Natarajan, S., Vanathi, D., Ramasamy, S., and Sathyamoorthi, K. (2007). Lowland Rice in Coastal Saline Soils - A Review. *Agric. Rev.* 28, 235–238.

Ambreen, H., Kumar, S., Kumar, A., Agarwal, M., Jagannath, A., and Goel, S. (2018). Association Mapping for Important Agronomic Traits in Safflower (Carthamus tinctorius L.) Core Collection Using Microsatellite Markers. *Front. Plant Sci.* 9, 402. doi:10.3389/fpls.2018.00402

Barrett, B. A., and Kidwell, K. K. (1998). AFLP-Based Genetic Diversity Assessment Among Wheat Cultivars from the Pacific Northwest. *Crop Sci.* 38, 1261–1271. doi:10.2135/cropsci1998.0011183X003800050025x

Bisht, I. S., Mahajan, R. K., Loknathan, T. R., and Agrawal, R. C. (1998). Diversity in Indian Sesame Collection and Stratification of Germplasm Accessions in Different Diversity Groups. *Genet. Resour. Crop Evol.* 45, 325–335. doi:10.1023/A:1008652420477

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.

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

Brown, A. H. D. (1989). "The Case for Core Collections," in *The Use of Plant Genetic Resources*. Editors A. H. D. Brown, O. H. Frankel, D. R. Marshall, and J. T. Williams (England, London: Cambridge University Press), 136–156.

Bruhn, H. (2005). A Short Guided Tour through Functional and Structural Features of Saposin-like Proteins. *Biochem. J.* 389, 249–257. doi:10.1042/BJ20050051

Chang, T.-T. (1976). The Origin, Evolution, Cultivation, Dissemination, and Diversification of Asian and African Rices. *Euphytica* 25, 425–441. doi:10.1007/bf00041576

Chen, W., Hou, L., Zhang, Z., Pang, X., and Li, Y. (2017). Genetic Diversity, Population Structure, and Linkage Disequilibrium of a Core Collection of Ziziphus Jujuba Assessed with Genome-wide SNPs Developed by Genotyping-By-Sequencing and SSR Markers. *Front. Plant Sci.* 8, 575–588. doi:10.3389/fpls.2017.00575

Chen, X., and Sullivan, P. F. (2003). Single Nucleotide Polymorphism Genotyping: Biochemistry, Protocol, Cost and Throughput. *Pharmacogenomics J.* 3, 77–96. doi:10.1038/sj.tpj.6500167

Cheon, K.-S., Baek, J., Cho, Y.-i., Jeong, Y.-M., Lee, Y.-Y., Oh, J., et al. (2018). Single Nucleotide Polymorphism (SNP) Discovery and Kompetitive Allele-specific PCR (KASP) Marker Development with Korean Japonica Rice Varieties. *Plant Breed. Biotech.* 6, 391–403. doi:10.9787/pbb.2018.6.4.391

Choudhury, B., Khan, M. L., and Dayanandan, S. (2013). Genetic Structure and Diversity of Indigenous rice (Oryza Sativa) Varieties in the Eastern Himalayan Region of Northeast India. *SpringerPlus* 2., Springer, 228–237. doi:10.1186/2193-1801-2-228

Civáň, P., Ali, S., Batista-Navarro, R., Drosou, K., Ihejieto, C., Chakraborty, D., et al. (2019). Origin of the Aromatic Group of Cultivated Rice (Oryza Sativa L.) Traced to the Indian Subcontinent. *Genome Biol. Evol.* 11, 832–843. doi:10.1093/gbe/evz039

Coates, B. S., Sumerford, D. V., Miller, N. J., Kim, K. S., Sappington, T. W., Siegfried, B. D., et al. (2009). Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. *J. Hered.* 100, 556–564. doi:10.1093/jhered/esp028

Das, B., Sengupta, S., Parida, S. K., Roy, B., Ghosh, M., Prasad, M., et al. (2013). Genetic Diversity and Population Structure of rice Landraces from Eastern and North Eastern States of India. *BMC Genet.* 14, 71–85. doi:10.1186/1471-2156-14-71

Edae, E. A., Byrne, P. F., Haley, S. D., Lopes, M. S., and Reynolds, M. P. (2014). Genome-wide Association Mapping of Yield and Yield Components of spring Wheat under Contrasting Moisture Regimes. *Theor. Appl. Genet.* 127, 791–807. doi:10.1007/s00122-013-2257-8

El Bakkali, A., Haouane, H., Moukhli, A., Costes, E., van Damme, P., and Khadari, B. (2013). Construction of Core Collections Suitable for Association Mapping to Optimize Use of Mediterranean Olive (Olea Europaea l.) Genetic Resources. *PLoS One* 8, e61265. doi:10.1371/journal.pone.0061265

Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., et al. (2018). Genetic Diversity and Population Structure of F3:6 nebraska winter Wheat Genotypes Using Genotyping-By-Sequencing. *Front. Genet.* 9, 76–84. doi:10.3389/fgene.2018.00076

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, 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Frankel, O. H. (1984). "Genetic Perspective of Germplasm Conservation," in *Genetic Manipulation: Impact on Man and Society*. Editors A WK. Llimensee, WL Peacock, and P Starlinger (England: Cambridge University Press), 161–170.

Gonzaga, Z. J., Aslam, K., Septiningsih, E. M., and Collard, B. C. Y. (2015). Evaluation of SSR and SNP Markers for Molecular Breeding in Rice. *Plant Breed. Biotech.* 3, 139–152. doi:10.9787/PBB.2015.3.2.139

Gouesnard, B., Bataillon, T. M., Decoux, G., Razole, C., Schoen, D. J., and david, J. L. (2001). MSTRAT: An Algorithm for Building Germ Plasm Core Collections by Maximizing Allelic or Phenotypic Richness. *J. Hered.* 92, 93–94. doi:10.1093/jhered/92.1.93

Hu, J., Zhu, J., and Xu, H. M. (2000). Methods of Constructing Core Collections by Stepwise Clustering with Three Sampling Strategies Based on the Genotypic Values of Crops. *Theor. Appl. Genet.* 101, 264–268. doi:10.1007/s001220051478

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide Association Studies of 14 Agronomic Traits in rice Landraces. *Nat. Genet.* 42, 961–967. doi:10.1038/ng.695

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., et al. (2012). Genome-wide Association Study of Flowering Time and Grain Yield Traits in a Worldwide Collection of rice Germplasm. *Nat. Genet.* 44, 32–39. doi:10.1038/ng.1018

Inomata, N. (2020). Gibberellin-regulated Protein Allergy: Clinical Features and Cross-Reactivity. *Allergol. Int.* 69, 11–18. doi:10.1016/j.alit.2019.10.007

Jain, M., Moharana, K. C., Shankar, R., Kumari, R., and Garg, R. (2014). Genomewide Discovery of DNA Polymorphisms in rice Cultivars with Contrasting Drought and Salinity Stress Response and Their Functional Relevance. *Plant Biotechnol. J.* 12, 253–264. doi:10.1111/pbi.12133

Jasim Aljumaili, S., Rafii, M. Y., Latif, M. A., Sakimin, S. Z., Arolu, I. W., and Miah, G. (2018). Genetic Diversity of Aromatic Rice Germplasm Revealed by SSR Markers. *Biomed. Res. Int.* 2018, 1–11. doi:10.1155/2018/7658032

Kasso, M., and Balakrishnan, M. (2013). Ex Situ *Conservation of Biodiversity with Particular Emphasis to Ethiopia*. Hindawi Publishing Corporation, ISRN Biodiversity, 1–11. Article ID 985037, 2013.

Kim, K.-W., Chung, H.-K., Cho, G.-T., Ma, K.-H., Chandrabalan, D., Gwag, J.-G., et al. (2007). PowerCore: a Program Applying the Advanced M Strategy with a Heuristic Search for Establishing Core Sets. *Bioinformatics* 23, 2155–2162. doi:10.1093/bioinformatics/btm313

Kumar, A., Kumar, S., Singh, K. B. M., Prasad, M., and Thakur, J. K. (2020). Designing a Mini-Core Collection Effectively Representing 3004 Diverse rice Accessions. *Plant Commun.* 1, 100049. doi:10.1016/j.xplc.2020.100049

Li, F. P., Lee, Y. S., Kwon, S. W., Li, G., and Park, Y. J. (2014). Analysis of Genetic Diversity and Trait Correlations Among Korean Landrace rice (*Oryza Sativa* L.). *Genet. Mol. Res.* 13, 6316–6331. doi:10.4238/2014.april.14.12

Liu, K., and Muse, S. V. (2005). PowerMarker: an Integrated Analysis Environment for Genetic Marker Analysis. *Bioinformatics* 21, 2128–2129. doi:10.1093/bioinformatics/bti282

Luan, S., Chiang, T.-Y., and Gong, X. (2006). High Genetic Diversity vs. Low Genetic Differentiation in Nouelia Insignis (Asteraceae), a Narrowly Distributed and Endemic Species in China, Revealed by ISSR Fingerprinting. *Ann. Bot.* 98, 583–589. doi:10.1093/aob/mcl129

Luo, Z., Brock, J., Dyer, J. M., Kutchan, T., Schachtman, D., Augustin, M., et al. (2019). Genetic Diversity and Population Structure of a Camelina Sativa Spring Panel. *Front. Plant Sci.* 10, 184–195. doi:10.3389/fpls.2019.00184

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, 1–13. doi:10.1038/ncomms10532

McCouch, S. R., Zhao, K., Wright, M., Tung, C.-W., Ebana, K., Thomson, M., et al. (2010). Development of Genome-wide SNP Assays for rice. *Breed. Sci.* 60, 524–535. doi:10.1270/jsbbs.60.524

Mourad, A. M. I., Belamkar, V., and Baenziger, P. S. (2020). Molecular Genetic Analysis of spring Wheat Core Collection Using Genetic Diversity, Population Structure, and Linkage Disequilibrium. *BMC Genomics* 21, 434. doi:10.1186/s12864-020-06835-0

Nachimuthu, V. V., Muthurajan, R., Duraialaguraja, S., Sivakami, R., Pandian, B. A., Ponniah, G., et al. (2015). Analysis of Population Structure and Genetic Diversity in rice Germplasm Using SSR Markers: An Initiative towards Association Mapping of Agronomic Traits in Oryza Sativa. *Rice* 8, 30–54. doi:10.1186/s12284-015-0062-5

Nei, M., Tajima, F., and Tateno, Y. (1983). Accuracy of Estimated Phylogenetic Trees from Molecular Data. *J. Mol. Evol.* 19, 153–170. doi:10.1007/bf02300753

Odong, T. L., Jansen, J., van Eeuwijk, F. A., and van Hintum, T. J. L. (2013). Quality of Core Collections for Effective Utilisation of Genetic Resources Review, Discussion and Interpretation. *Theor. Appl. Genet.* 126, 289–305. doi:10.1007/s00122-012-1971-y

Pathaichindachote, W., Panyawut, N., Sikaewtung, K., Patarapuwadol, S., and Muangprom, A. (2019). Genetic Diversity and Allelic Frequency of Selected Thai and Exotic Rice Germplasm Using SSR Markers. *Rice Sci.* 26, 393–403. doi:10.1016/j.rsci.2018.11.002

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research-Aan Update. *Bioinformatics* 28, 2537–2539. doi:10.1093/bioinformatics/bts460

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945–959. doi:10.1111/j.1471-8286.2007.01758.x10.1093/genetics/155.2.945

Rafalski, A. (2002). Applications of Single Nucleotide Polymorphisms in Crop Genetics. *Curr. Opin. Plant Biol.* 5, 94–100. doi:10.1016/S1369-5266(02)00240-6

Rambaut, A. (2010). FigTree v1.3.1. A Graphical Viewer of Phylogenetic Trees. Available at: http://tree.bio.ed.ac.uk/software/figtree/ (Accessed February 25, 2021).

Rashmi, D., Bisen, P., Saha, S., Loitongbam, B., Singh, S., Pallavi, P., et al. (2017). Genetic Diversity Analysis in Rice (Oryza Sativa L.) Accessions Using SSR Markers. *Intern. Jour. Agricul., Environ. Biotech.* 10, 457–467. doi:10.5958/2230-732X.2017.00057.2

Raybould, A. F., Mogg, R. J., and Clarke, R. T. (1996). The Genetic Structure of Beta Vulgaris Ssp. Maritima (Sea Beet) Populations: RFLPs and Isozymes Show Different Patterns of Gene Flow. *Heredity* 77, 245–250. doi:10.1038/hdy.1996.138

Reif, J. C., Zhang, P., Dreisigacker, S., Warburton, M. L., van Ginkel, M., Hoisington, D., et al. (2005). Wheat Genetic Diversity Trends during Domestication and Breeding. *Theor. Appl. Genet.* 110, 859–864. doi:10.1007/s00122-004-1881-8

Roy Choudhury, D., Singh, N., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., et al. (2014). Analysis of Genetic Diversity and Population Structure of Rice Germplasm from North-Eastern Region of India and Development of a Core Germplasm Set. *PLoS One* 9, e113094. doi:10.1371/journal.pone.0113094

Ryan, M. C., Stucky, M., Wakefield, C., Melott, J. M., Akbani, R., Weinstein, J. N., et al. (2020). Interactive Clustered Heat Map Builder: An Easy Web-Based Tool for Creating Sophisticated Clustered Heat Maps. *F1000Res* 8, 1750. doi:10.12688/f1000research.20590.2

Schlötterer, C. (2004). The Evolution of Molecular Markers - Just a Matter of Fashion. *Nat. Rev. Genet.* 5, 63–69. doi:10.1038/nrg1249

Seo, J., Lee, G., Jin, Z., Kim, B., ChinKoh, J. H., and Koh, H. J. (2020). Development and Application of Indica–Japonica SNP Assays Using the Fluidigm Platform for rice Genetic Analysis and Molecular Breeding. *Mol. Breed.* 40, 1–16. doi:10.1007/s11032-020-01123-x

Shete, S., Tiwari, H., and Elston, R. C. (2000). On Estimating the Heterozygosity and Polymorphism Information Content Value. *Theor. Popul. Biol.* 57, 265–271. doi:10.1006/tpbi.2000.1452

Singh, N., Choudhury, D. R., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., et al. (2013). Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties. *PLoS One* 8, e84136. doi:10.1371/journal.pone.0084136

Singh, N., Jayaswal, P. K., Panda, K., Mandal, P., Kumar, V., Singh, B., et al. (2015). Single-copy Gene Based 50 K SNP Chip for Genetic Studies and Molecular Breeding in rice. *Sci. Rep.* 5, 11600. doi:10.1038/srep11600

Singh, N. K., Dalal, V., Batra, K., Singh, B. K., Chitra, G., Singh, A., et al. (2006). Single-copy Genes Define a Conserved Order between rice and Wheat for Understanding Differences Caused by Duplication, Deletion, and Transposition of Genes. *Funct. Integr. Genomics.* 7, 17–35. doi:10.1007/s10142-006-0033-4

Suvi, W. T., Shimelis, H., Laing, M., Mathew, I., and Shayanowako, A. I. T. (2020). Assessment of the Genetic Diversity and Population Structure of rice Genotypes Using SSR Markers. *Acta Agriculturae Scand. Section B - Soil Plant Sci.* 70, 76–86. doi:10.1080/09064710.2019.1670859

Syvänen, A. C. (2001). Accessing Genetics Variation: Genotyping Single Nucleotide Polymorphisms. *Nat. Rev. Genet.* 2, 930–942.

Tarang, A., Kordrostami, M., Shahdi Kumleh, A., Hosseini Chaleshtori, M., Forghani Saravani, A., Ghanbarzadeh, M., et al. (2020). Study of Genetic Diversity in rice (*Oryza Sativa* L.) Cultivars of Central and Western Asia Using Microsatellite Markers Tightly Linked to Important Quality and Yield Related Traits. *Genet. Resour. Crop Evol.* 67, 1537–1550. doi:10.1007/s10722-020-00927-2

Thomson, M. J., Zhao, K., WrightMcNally, M. K. L., McNally, K. L., Rey, J., Tung, C.-W., et al. (2012). High-throughput Single Nucleotide Polymorphism Genotyping for Breeding Applications in rice Using the BeadXpress Platform. *Mol. Breed.* 29, 875–886. doi:10.1007/s11032-011-9663-x

Upadhyaya, H. D., Dwivedi, S. L., Sharma, S., Lalitha, N., Singh, S., Varshney, R. K., et al. (2014). Enhancement of the Use and Impact of Germplasm in Crop Improvement. *Plant Genet. Resour.* 12, S155–S159. doi:10.1017/S1479262114000458

van Hintum, Th. J. L., Brown, A. H. D., Spillane, C., and Hodgkin, T. (2000). *Core Collections of Plant Genetic Resources*. IPGRI Technical Bulletin No. 3. Rome, Italy: International Plant Genetic Resources Institute.

Wang, J. C., Hu, J., Xu, H. M., and Zhang, S. (2007). A Strategy on Constructing Core Collections by Least Distance Stepwise Sampling. *Theor. Appl. Genet.* 115, 1–8. doi:10.1007/s00122-007-0533-1

Xu, Q., Yuan, X., Wang, S., Feng, Y., Yu, H., Wang, Y., et al. (2016). The Genetic Diversity and Structure of *Indica* rice in China as Detected by Single Nucleotide Polymorphism Analysis. *BMC Genet.* 17, 53. doi:10.1186/s12863-016-0361-x

Xu, X., Liu, X., Ge, S., Jensen, J. D., Hu, F., Li, X., et al. (2011). Resequencing 50 Accessions of Cultivated and Wild rice Yields Markers for Identifying Agronomically Important Genes. *Nat. Biotechnol.* 30, 105–111. doi:10.1038/nbt.2050

Yamamoto, T., Nagasaki, H., Yonemaru, J.-i., Ebana, K., Nakajima, M., Shibaya, T., et al. (2010). Fine Definition of the Pedigree Haplotypes of Closely Related rice Cultivars by Means of Genome-wide Discovery of Single-Nucleotide Polymorphisms. *BMC Genomics* 11, 267. doi:10.1186/1471-2164-11-267

Yan, W., Rutger, J. N., Bryant, R. J., Bockelman, H. E., Fjellstrom, R. G., Chen, M.-H., et al. (2007). Development and Evaluation of a Core Subset of the USDA rice Germplasm Collection. *Crop Sci.* 47, 869–876. doi:10.2135/cropsci2006.07.0444

Yang, G., Chen, S., Chen, L., Sun, K., Huang, C., Zhou, D., et al. (2019). Development of a Core SNP Arrays Based on the KASP Method for Molecular Breeding of rice. *Rice (N Y)* 12, 21–18. doi:10.1186/s12284-019-0272-3

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 reviewer BPS declared a shared affiliation with one of the authors NKS to the handling editor at the time of the 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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 07 December 2021
doi: 10.3389/fgene.2021.758665

Genetic Factors Underlying Single Fiber Quality in A-Genome Donor Asian Cotton (*Gossypium arboreum*)

Muhammad Shahid Iqbal[1,2†], Shurong Tang[1†], Zareen Sarfraz[1†], Muhammad Sajid Iqbal[1,3], Hongge Li[1], Shoupu He[1], Yinhua Jia[1], Gaofei Sun[4], Zhaoe Pan[1], Geng Xiaoli[1], Abid Mahmood[2], Saghir Ahmad[2], Mian Faisal Nazir[1], Baojun Chen[1], Liru Wang[1], Baoyin Pang[1], Shoujun Wei[1] and Xiongming Du[1]**

[1]State Key Laboratory of Cotton Biology/Institute of Cotton Research, Chinese Academy of Agricultural Sciences (ICR, CAAS), Anyang, China, [2]Ayub Agricultural Research Institute Faisalabad, Cotton Research Institute, Multan, Pakistan, [3]Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan, [4]Anyang Institute of Technology, Anyang, China

OPEN ACCESS

Edited by:
Awais Rasheed, Quaid-i-Azam University, Pakistan

Reviewed by:
Mingzhou Song, New Mexico State University, United States
Pengcheng Li, Yangzhou University, China

***Correspondence:**
Shoujun Wei
13503728390@163.com
Xiongming Du
dujeffrey8848@hotmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *14 August 2021*
Accepted: *17 November 2021*
Published: *07 December 2021*

Citation:
Iqbal MS, Tang S, Sarfraz Z, Iqbal MS, Li H, He S, Jia Y, Sun G, Pan Z, Xiaoli G, Mahmood A, Ahmad S, Nazir MF, Chen B, Wang L, Pang B, Wei S and Du X (2021) Genetic Factors Underlying Single Fiber Quality in A-Genome Donor Asian Cotton (Gossypium arboreum). Front. Genet. 12:758665. doi: 10.3389/fgene.2021.758665

The study of A-genome Asian cotton as a potential fiber donor in Gossypium species may offer an enhanced understanding of complex genetics and novel players related to fiber quality traits. Assessment of individual fibers providing classified fiber quality information to the textile industry is Advanced Fiber Information System (AFIS) in the recent technological era. Keeping the scenario, a diverse collection of 215 Asiatic cotton accessions were evaluated across three agro-ecological zones of China. Genome-Wide Association Studies (GWAS) was performed to detect association signals related to 17 AFIS fiber quality traits grouped into four categories viz: NEPs, fiber length, maturity, and fineness. Significant correlations were found within as well as among different categories of various traits related to fiber quality. Fiber fineness has shown a strong correlation to all other categories, whereas these categories are shown interrelationships *via* fiber-fineness. A total of 7,429 SNPs were found in association with 17 investigated traits, of which 177 were selected as lead SNPs. In the vicinity of these lead SNPs, 56 differentially expressed genes in various tissues/development stages were identified as candidate genes. This compendium connecting trait-SNP-genes may allow further prioritization of genes in GWAS loci to enable mechanistic studies. These identified quantitative trait nucleotides (QTNs) may prove helpful in fiber quality improvement in Asian cotton through marker-assisted breeding as well as in reviving eroded genetic factors of *G. hirsutum via* introgression breeding.

Keywords: GWAS, AFIS, fiber quality, SNPs, Asian cotton, multi-environment

INTRODUCTION

Cotton has a prime position in the global natural textile fiber industry, making it a significant agricultural commodity. It remains an essential source of livelihood for a large percentage of the farming community. Cotton production has long been a crucial part of diverse farming systems, particularly those involving vegetables and cereals, helping farmers to maintain their incomes. The competition among major cotton-producing regions and the use of synthetic fibers are continuously increasing. It necessitates continual product quality improvements, measuring physical traits and germplasm resources (Meredith Jr, 2005). Almost 90% of the significance of the cotton crop relies on

its lint fiber. Generally, each fiber is an elongation (seed hair) originating from a cotton ovule, from protodermal cells in the ovule's seed coat integument (outer) layer. Single seed fibers are assumed to exhibit continua of shape, physical maturity, cell wall thickness, and length (Bradow et al., 1996). These physical traits determine the quality of the raw material, which underlie the quality of the finished product. However, determining cotton fiber quality is complex. As cotton fiber quality is a critical issue in cotton research, an accurate and precise measurement system is required for the various fiber quality traits (Berkley, 1948).

For the past two decades, industry and plant breeders have been utilizing High Volume Instrument (HVI) as the primary and sole source of measurement for selection and fiber quality improvement. However, the HVI system cannot assess many key fiber quality traits. Alternative systems for fiber quality evaluation, such as the Advanced Fiber Information System (AFIS), have been introduced. AFIS can obtain highly advanced and more accurate information about single fiber quality. Many studies have indicated that AFIS is an effective tool for predicting yarn quality along with spinning performance (Hequet et al., 2006). AFIS is the instrument of choice in the cotton industry, including cotton breeders, based on its ability to estimate mean fiber values and distributions. It accurately measures maturity and fineness through cross-sectional image analysis. Hence, AFIS is a powerful tool for the industry if appropriately linked with image analysis data (Thibodeaux et al., 2007).

Gossypium hirsutum (upland cotton), the most widely cultivated cotton species, is considered a natural allotetraploid cotton species with an AADD genome. It is thought to have resulted from natural interspecific hybridization involving the diploid species *Gossypium arboreum* (genome A2) and *Gossypium raimondii* (genome D5) (Paterson et al., 2012). It is challenging to explore the two co-resident genomes in *G. arboreum* that have unverified origins, and these tetraploid species are challenging to study. However, The A-genome donor of *G. hirsutum*, i.e., *G. arboreum*, harbors many putative genetic factors underlying fiber quality traits and stress resistance, and an in-depth study of *G. arboreum* might provide insights that could help to improve the *G. hirsutum* yield and fiber traits. Many cotton breeders and cotton geneticists are currently trying hard to understand the two-donor diploid genomes thoroughly. To understand mechanisms underlying fiber quality traits in the diploid species, it is essential to identify the genes controlling these traits. It may aid the introgression of genetic factors from diploid to tetraploid cotton.

In previous decades, *G. arboreum* cultivars rather than tetraploid cultivars were commercially grown in north-eastern Africa, the Middle East, and Asia (Guo et al., 2006). Being a diploid species, *G. arboreum* is highly adaptable to extreme environmental conditions (Maqbool et al., 2010) and can be cultivated using practical management approaches and fewer inputs (Iqbal et al., 2015). Its valuable traits include strengthened fiber, high seed index, and high oil content (Mehetre et al., 2003). Additionally, resistances against biotic stresses viz; reniform nematode (Erpelding and Stetina, 2013), tobacco budworm (Hedin et al., 1992), Cotton leaf curl virus (CLCuV) (Nazeer et al., 2014), and thrips (Stanton et al., 1992) can possibly be introgressed to *G. hirsutum*. However, specialized breeding techniques would be required to overcome the barriers during the hybridization process (Sacks and Robinson, 2009).

Information on the variability among *G. arboreum* genotypes and the complex interactions among valuable traits may allow improved cotton breeding programs to be developed. Predicting genotype variability can be accomplished *via* phenotypic assessments and characterization (Tahir and Noor, 2011). To develop a breeding program, the degree of potential within the genotypes in question and the extent of the associations among the target traits should be evaluated (Batool et al., 2010; Khan et al., 2010). There are four industrially important categories of fiber quality traits, namely, fineness, maturity, NEPS, and complete length. During the 20th and 21st centuries, many classical quantitative genetic studies calculated the variance and heritability of yield and fiber traits and their interactions with environmental factors, leading to yield and quality advances. However, the yield and quality have stagnated over the last decade, which may be due to the phenotypic selection pressure placed on commercial *G. hirsutum* cultivars. It may have ultimately reduced the genetic diversity in the primary cotton gene pool. It may explain the increase in the vulnerability of *G. hirsutum* cultivars to biotic and abiotic stresses (Maqbool et al., 2010). Classical cotton breeding efforts involving interspecific hybridizations for stable genetic transformation of novel allelic variation have encountered challenges. However, many related genomic tools and biological procedures have been developed. The significant advancements include polymorphic genetic markers, linkage maps, and divergent mapping populations.

Keeping the scenario, this study's primary goals include identifying genes and residing regulatory sequences controlling cellulose biosynthesis and cell development of fiber. For sequencing, assembling, and annotation of *G. arboreum* genome, contemporary genomic resources will be established. A genome-wide association study (GWAS) was conducted to correlate phenotype data of fiber quality traits taken from AFIS with genotypic data generated from Next-Generation Sequencing (NGS) technology. Using different algorithms while performing GWAS, key SNPs and fiber quality associated genes were identified and selected. A thorough analysis of functional annotation *via* bioinformatics tools was utilized to confirm the linkage between genes and traits in diploid Asian cotton genotypes.

MATERIALS AND METHODS

Plant Materials

Asian cotton (*G. arboreum*) collection having 215 accessions (**Supplementary Table S1**) were obtained from the Chinese National Germplasm Mid-term Genebank (Institute of Cotton Research, Anyang, China). They were grown in three diverse ecological regions of China in 2014, and their fiber quality was evaluated. The three regions were as follows: two major conventional cotton-growing regions, i.e., Anyang, Henan

(Yellow River Region), and Akesu, Xinjiang (Northwest Region), and a non-conventional potential cotton region, i.e., Sanya, Hainan (an island in South China). The experiment involved a triplicate randomized complete block design. Row spacing was 70 cm, and plant spacing was 30 cm in all locations. Planting was conducted during the regular cotton growing season (April) in Anyang and Akesu. However, in Sanya, planting was conducted during an extended cotton growing season (October–March). The agronomic and cultural practices were uniform across all locations to avoid biasness. Five guarded plants were randomly tagged for genotyping, along with phenotyping. At physical maturity, seed cotton was manually picked from each tagged plant separately. Muslin cloth bags were used for each sample to avoid mixing or any type of adulteration.

An AFIS PRO 2 (Zellweger-USTER) was used to assess the cotton fiber traits, including their distributions, which were presented using histograms of the distribution for the measured parameters (Shofner et al., 1990), including NEPS, fiber length, maturity as well as fiber fineness. The 17 traits assessed were as follows: Total nep count (TNN), Total nep mean Size (TNS), Fiber nep count (FNN), Seeds oat nep Count (SCN), Seed coat nep size (SCS), Mean length Weight (LW), Length Weight Variation (LWCV), Short Fiber Content Weight (SFCW), Upper Quartile Length Weight (UQLW), Fiber Length Variation (LNCV), Mean Length Number (Ln), Short Fiber Content (SFCn), The 5% Length Number (Ln5) and Short Fiber Content Number (SFCN), Maturity Ratio (MR), Immature Fiber Content (IMM), and Fiber Fineness (Mtex) detailed information about these traits is given in **Supplementary Table S2**.

The AFIS PRO 2 mechanically separates individual cotton fibers presented to an electro-optical sensor using high-velocity airflow. The AFIS Length & Maturity module optically determines the value for maturity ratio (MR) using Lord's equation while Immature Fiber Content (IMM) through the method described by Frydrych and Thibodeaux (2010). The AFIS PRO2 (which is the latest version) can complete one test in 2.5–3 min (Frydrych and Thibodeaux, 2010). After drying and cleaning, ginning was conducted to obtain lint samples for fiber quality analysis. The single fiber quality-related traits for NEPS, length, fineness, and maturity (**Supplementary Table S2**), were assessed three times for each sample to avoid error and then averaged for further statistical analysis.

Preparation for GWAS

The genotype data, which was above an average 6-fold sequencing depth, were derived from a previously published article by our team (Du et al., 2018), and the sequence of the same 215 (GA0001 to GA0215) accessions was considered for phenotyping in this experiment was picked out and used in this study. The genotype data were filtered before GWAS, SNPs with MAF <1% and SNPs that exhibited deviations from Hardy–Weinberg equilibrium (HWE) were removed, leaving 1,425,002 SNPs for GWAS. HWE theorem states that both the allele and genotype frequencies in a population remain constant, so testing for HWE is a standard quality control procedure in population genetic studies. Due to the large number of SNPs, subsets were analyzed separately; for this purpose, we employed PLINK v1.07, an open-source GWAS toolset (Purcell et al., 2007), using an R plugin. In a GWAS, while we hope for some true associations, most of the SNPs are not associated with the trait in question, so almost all *p*-values should come from a uniform distribution.

SNP annotation information was generated *via* utilizing ANNOVAR software (Wang et al., 2010) based on the *G. arboreum* reference genome. The genomic regions were categorized into different groups using genome annotation such as: downstream or upstream regions, the annotated SNP lying within 1 kb region either downside of transcription stop site/upside of transcription start site OR upside and downside of transcription sites simultaneously; intronic (non-coding) region; exonic (coding) region; splicing sites, lying within 2bp of splicing junction; and intergenic regions. Further, the SNPs harbored by the exonic regions were categorized as synonymous (didn't cause any change in amino acids); non-synonymous (caused changes of amino acids); as well as stop-gain and stop-loss type of mutations also grouped in this category.

Statistical Analyses

The data collected were subjected to multivariate analysis using hierarchical clustering analysis and principal component analysis (PCA). After confirming sufficient variation among the traits, GWAS was conducted, making the traits suitable for further genetic analyses. Multivariate correlation analysis was performed to assess the relationships among the 17 traits. The analyses were performed using JMP Pro 14.0 software (SAS Institute Inc.).

Phylogenetic and Population Structure Analyses

To understand the phylogenetic relationships among the accessions, phylogenetic analysis was performed using the SNPhylo pipeline (Li et al., 2014). The SNPs were pruned to reduce SNP redundancy due to linkage disequilibrium (LD; SNPs in a specific LD block provide redundant lineage information). SNPhylo uses only one informative SNP in each LD block, so the phylogenetic analysis was based on high-quality SNPs. A subset of 707,968 high-quality SNPs [>2000, minor allele frequency (MAF) >5%], with missing data rate <20%, was utilized. A neighbor-joining tree was constructed using PHYLIP v3.696 (Felsenstein, 1993), with 100 bootstrap replicates.

To infer the population structure, ADMIXTURE software was used (Falush et al., 2003). This software uses a mode-based clustering method that considers different numbers of clusters (*K*). A set of 431,985 SNPs, excluding missing genotypes, was included in the analysis. Additionally, SMARTPCA in EIGENSOFT software (Patterson et al., 2006) was used to conduct PCA on a set of 4,329,838 SNPs.

GWAS

A set of 1,425,002 high-quality SNPs (MAF >5%, <20% missing rate) from 215 diverse *G. arboreum* accessions were used for the GWAS. The GWAS was conducted using EMMAX (Efficient Mixed-Model Association eXpedited) software (Kang et al.,

2010), which can handle large datasets for GWAS. The 17 AFIS-related fiber quality traits in three ecological locations/environments (Anyang, Henan; Sanya, Hainan; and Akesu, Xinjiang) were considered in this analysis. Population stratification and hidden relatedness were modeled with a kinship (*K*) matrix using the EMMAX-kin-intel EMMAX package. Bonferroni correction (dividing the desired *p*-value by the number of comparisons (Yang et al., 2005; Pearson and Manolio, 2008) was used to avoid a high false-positive rate. Hence, the adjusted *p*-value threshold for the *G. arboreum* accessions was $p < 4.9 \times 10^{-5}$. Manhattan plots and quantile-quantile (Q–Q) plots were constructed using the CMplot R package to visualize the results. Lead SNPs were identified from the Manhattan plots by selecting the SNPs from each peak having the higher $-\log P = 6.15$ using formula P < P = 1/N (where N is the total number of SNPs used for GWAS.

LD Analysis

Linkage Disequilibrium (LD) decay was visualized by plotting r^2 against the physical distance (kb) between the SNPs using PopLDdecay software (Zhang et al., 2019). The GWAS results related to four chromosomes (Chr02, Chr03, Chr05, and Chr06), and the four trait categories are depicted in detail (**Figures 5–8**). The LD (D′) decay distance between the paired genes and SNPs ranges 115.5 kb (57 kb on either side of each key SNP). The GWAS and LD combined plot of each selected area were plotted using LDBlockShow software (Dong et al., 2021). All genomic positions provided were based on the *G. arboreum* reference genome (Li et al., 2014) v1.1. Differential expression data regarding the genes near lead SNPs were obtained from our Institute database (http://grand.cricaas.com.cn/) and the Cotton Functional Genomics Database (www.cottonfgd.org). The upregulated genes (>1) specially in ovule and fiber were considered candidate genes and plotted in a heatmap.

RESULTS

Phenotypic Characteristics

The following 17 phenotypic single fiber quality traits of 215 diploid *G. arboreum* accessions were assessed (**Supplementary Table S2**): NEPs (TNN, TNS, FNN, FNS, SCN, SCS), maturity (MR and IMM), fineness (MTex), and length (LW, LWC, Ln, Ln5, LCV, SFCW, SFCn, and UQLW). A considerable range of variation and normal distribution with insignificant skewness and kurtosis has been illustrated regarding investigated fiber quality traits under three different environments (**Supplementary Figure S1**, **Supplementary Table S3**). Normal distribution values in the diagonals and range of variation *via* boxplots on the extreme right of the scatterplot matrix has been shown in **Supplementary Figure S1** provided sufficient grounds for further processing of data for GWAS (**Supplementary Table S3**). The highest standard deviation (Std Dev) has been displayed by NEPs related trait followed by length related traits and lowest Std Dev was exhibited by maturity related trait. Similar trend was followed by traits across all locations and further detailed description has been provided in **Supplementary Table S3**.

The strength and direction of correlations among the various traits are shown in the scatterplot matrix in **Supplementary Figure S1**. The upper values in black represent the overall correlations. The lower values in different colors represent the correlations in each location. Almost all the correlations were statistically significant. The significant or highly significant positive correlations were determined among SFCn, SFCW, TNN, FNN, SCN, TNS, Ln, LN5, LW, UQLW. However, highly significant negative correlations were among MAT, IMM, SFCW, SFCn, SCN, Ln, Ln5, LW, LCV, LWC and UQLW (**Supplementary Figure S1**). PCA was also used to investigate the relationships among the phenotypic traits and the factors underlying trait variation. The first two principal components (PCs) explained 56.8% of the total variation of the traits. PC1 explained 38.6% of the total variation, and loading on this PC was highest for FNN, LWC, LCV, and Mtex. Both FNS and SCS exhibited maximum loadings on PC4 (**Figure 1**).

Population Structure, PCA, Phylogenetic, and LD Analyses

A set of 215 *G. arboreum* lines were evaluated across two major conventional cotton-growing regions (Anyang and Akesu) and one non-conventional (Sanya) cotton-growing region in China for association studies. Population structure, PCA, and phylogenetic analyses (**Figures 2A–C**). The phylogenic tree indicated the division of accessions into YZR, SC, and YER groups, revealing substantial geographical distributions (**Figure 2C**). PCA confirmed this clustering. Also, high nucleotide diversity was observed, as compared to and within clusters. LD analysis is a helpful tool to locate causal loci in GWAS, and we found that the LD decay distance in the *G. arboreum* accessions was ~115.5 kb (indicating the physical distance between SNPs) (**Figure 2D**). These findings indicate a significant population structure in these accessions.

GWAS of AFIS Fiber Quality Traits

Halting the stagnation of fiber quality has been the primary objective of cotton breeding programs in the last few decades. GWAS was performed on the 215 *G. arboreum* accessions evaluated across three locations to identify genetic factors linked to single fiber quality traits while considering population structure and phylogenetic relatedness (Yu et al., 2006).

A set of 1,425,002 high-quality SNPs with MAF >5% were utilized for GWAS (**Supplementary Table S4**). The EMMAX software detected 10,434 association signals with a threshold probability value $p < 4.9 \times 10^{-5}$ (**Supplementary Table S5**). For the first time, such associations were identified for single fiber quality traits evaluated across multiple environments (**Supplementary Table S6**). The distribution of 7,429 identified and annotated significantly associated SNPs is presented in **Table 1** which were categorized into different regions or groups. Among these SNPs, 4,328 were grouped

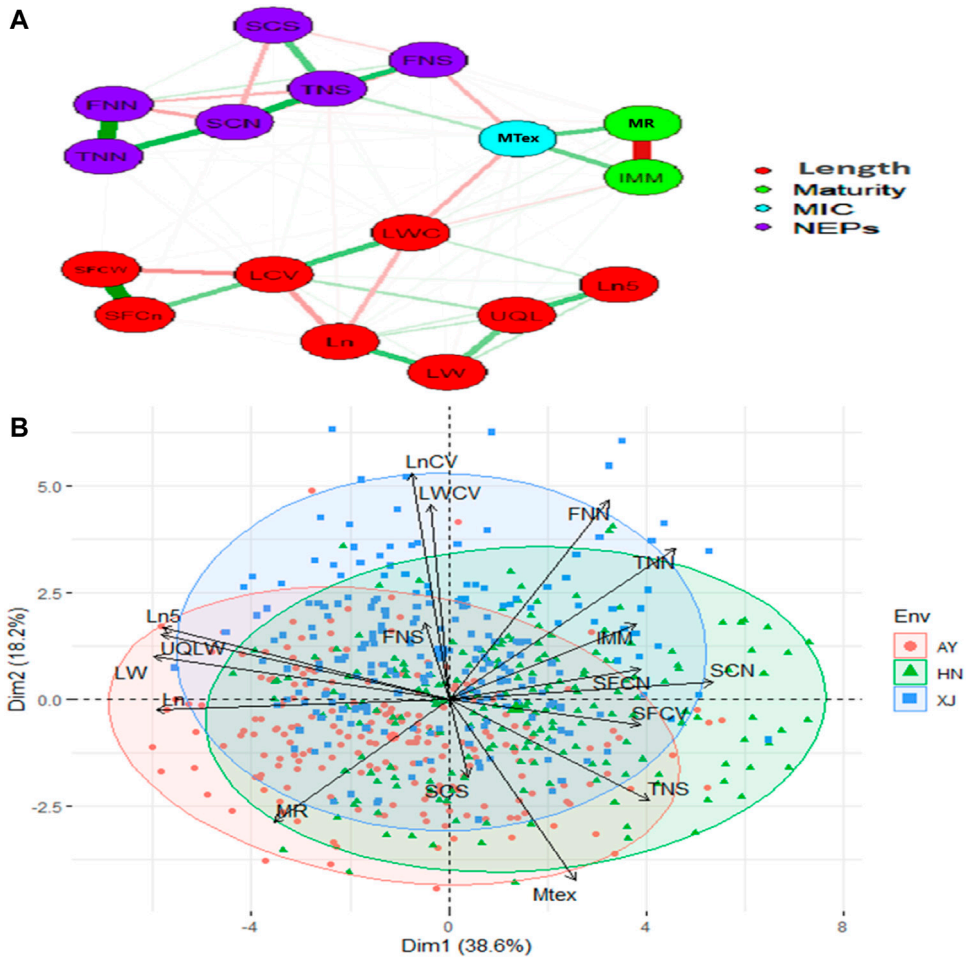


FIGURE 1 | Relationship network and PCA biplot of AFIS Relate Fiber Quality Traits **(A)** Relationship network between different single fiber quality traits of *G. arboreum* accessions. Red and green color connection lines between each trait illustrate negative and positive relationships between them, respectively. All the investigated traits are categorized into four major groups of Length (Red), Maturity (Green), MIC (Blue), and NEPs (Purple) **(B)** PCA-based biplot on the phenotypic variation of single fiber quality traits projected in the PCA1 and PCA2 planes. Legends on the top right: Three different colors represent different environments in which G. arboreum accessions were evaluated; Black to red color gradient represents the contribution of each trait in variation.

into the intergenic (non-coding) regions and 3,101 annotated SNPs were grouped in the genic (coding) regions with a division detail as: 766 significantly associated SNPs of upstream region; 760 downstream region SNPs; 509 upstream/downstream regions SNPs and 339 exonic SNPs with 164 non-synonymous, 161 synonymous, 1 splicing, 7 stop codon loss, 6 stop codon gain SNPs which triggered amino acid changes and premature stopping or elongated transcripts production (**Table 1**).

There were 3271, 3816, and 3345 SNPs significantly associated with single fiber quality traits AY, HN and XJ, respectively (**Figure 3A**, **Supplementary Table S7**). There were 35 SNPs common to all three locations, 1112 common in AY and HN, 605 common to AY and XJ, and 20 commons to XJ and HN (**Figure 3B**). 4984 SNPs were associated with length traits, 4121 with nep traits, 432 with fineness traits, 897 with maturity traits (**Figure 3A**). Two SNPs were associated with length, fineness, and nep traits, 29 with length and nep traits, 15 with length and maturity traits, 7 with length and fineness traits, 32 with nep and maturity traits, and 25 with fineness and maturity traits (**Figure 3C**).

A total of 7,429 of the significant trait-associated SNPs distributed on the 13 chromosomes of diploid *G. arboreum* accessions were located in quantitative trait nucleotide (QTN)-rich regions (**Supplementary Figure S3**). The detailed chromosomal distribution is presented in **Figure 3A**. The maximum number of associations were on chromosome 5, while the minimum on chromosome 12 (**Figure 3A**). A similar trend of peak associations hits has also been observed on different chromosomes and represented in **Figure 4**.

A trend of pleiotropy was discovered for significantly associated SNPs. Out of 7,429 significant SNPs, 1,852 SNPs displayed pleiotropy for single fiber quality traits (**Figure 3D**). A sum of 228 association signals were observed based on multiple corrections (1/1430002 = 6.99E-07 or ~0.000001 value) for 177 Lead SNPs ($-\log P = 6.15$) for 14 single fiber quality traits across three environments with 155 association signals detected from

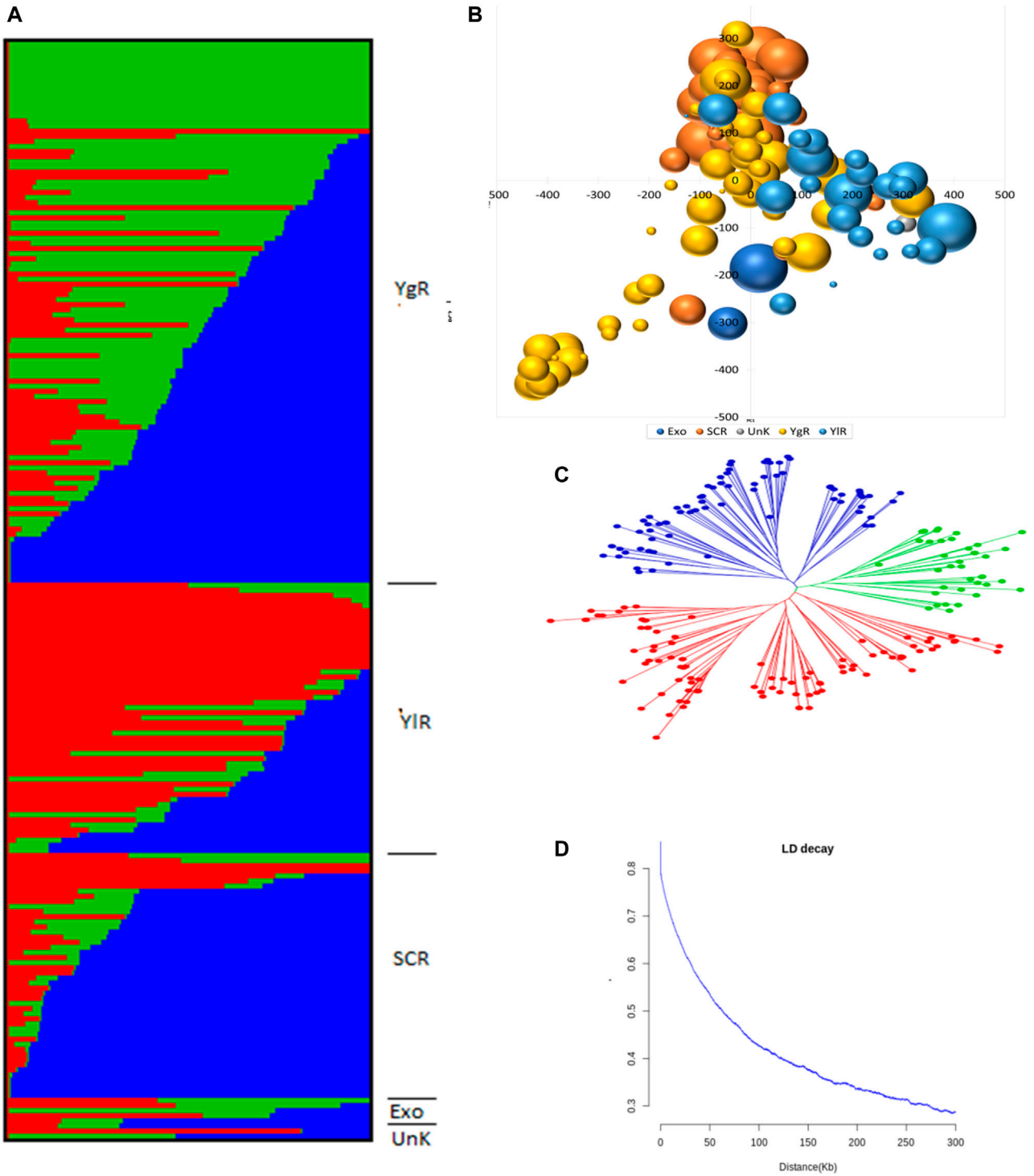


FIGURE 2 | Population Genetic Study **(A)** Population structure of 215 *G. arboreum* cotton accessions in the association panel based on modeled ADMIXTURE analysis at K = 3. The component structural groups of Asiatic cotton are divided into cotton growing zones i.e., YgR (Yangtze River Region), YIR (Yellow River Region), SCR (South China Region), Exo (Exotic) and UnK (Unknown) **(B)** Principal component analysis (PCA) plot of PC1 and PC2 based on 1430005 genomewide SNPs scored on 215 *G. arboreum* accessions **(C)** Phylogenetic tree constructed using whole-genome SNP data, distributing genotypes into three clades as per original classification **(D)** Decay of linkage disequilibrium (LD) expressed as a function of physical distance (kb) and r^2.

TABLE 1 | Counts of significantly associated SNP for 215 *G. arboreum* genotypes across three geographical locations.

Genomic region	SNP count	Percentage
Intergenic	4328	58.26
Genic	3101	41.74
Intronic	727	9.79
Upstream	766	10.31
Downstream	760	10.23
Upstream; downstream	509	6.85
Exonic (non-synonymous)	164	2.21
Exonic (synonymous)	161	2.17
Exonic (stop codon loss)	7	0.09
Exonic (stop codon gain)	6	0.08
Splicing	1	0.01
Grand total	7,429	100.00

AY, 47 from HN and 26 from XJ. Thorough examination revealed that chromosomal distribution of lead SNPs is uneven across the entire genome. Chromosome 5 had a maximum number of 65 lead SNPs while Chromosome 12 had a minimum of 1 lead SNP (**Supplementary Tables S8, S9**).

Candidate Gene Prediction and Annotation

A total of 113 genes were discovered in the 115.5 kb flanking window of lead SNPs. Of these genes, 25 and 1 were identified in maximum and minimum quantities on Chromosome 7 and Chromosome 12, respectively (**Supplementary Table S10**). These genes were further scrutinized based on their differential gene expression data from our Institute database (http://grand.cricaas.com.cn/) and the Cotton Functional Genomics Database (http://www.cottonfgd.org/) (**Supplementary Figure S3**; **Supplementary Tables S10A,B**). As a result, 56 candidate genes were identified *via* validation from gene expression data of various cotton tissues, organs, or growth and developmental stages (**Supplementary Table S12**). Of these 56 genes, 28 were associated with length, 25 with NEPs, 1 with maturity and 2 with fineness (**Supplementary Table S11**). Further, Gene Ontology (GO) annotation (cellular components, biological processes as well as molecular functions) of putative genes was performed (**Supplementary Table S12**). The GWAS summaries of lead SNPs related to NEPs, maturity, fineness, and length comprised of Manhattan plots with a significance threshold horizontal line drawn at [−log (p) > 6], quantile-quantile plots, GWAS-LD blocks for the depiction of haplotype region surrounding the peak associated significantly with respective trait coupled with regional plot revealing particular key SNP and its nearby gene, a boxplot for the demonstration of differential among trait related favorable haplotypes as well as expression plots of prominently selected annotated genes concerning traits at different developmental stages of cotton genotypes.

The NEPs related two traits; AY_TNN and AY_SCN showed highly significant associations with pleiotropic lead SNP Chr05_91033229 at peak with log(p) > 6. The haplotypes of this SNP exhibited significant differences depicted *via* boxplots. The gene expression results gave following three genes namely, *Ga05G3958* and *Ga05G3959* in vicinity of this lead SNP. These candidate genes presented their higher expression levels regarding FPKM value in ovule, fiber, and seed tissues (**Figure 5**). The maturity related two traits; AY_MR and AY_IMM presented highly significant relationships with pleiotropic lead SNP Chr11_117679776 with log(p) > 6 found at peak of Manhattan plots. This lead SNP revealed distinct haplotypes mentioned in boxplots. The gene expression outcomes provided a gene viz, *Ga11G3320* in which this SNP was found. This candidate gene exhibited high FPKM values in almost all tissues of ovule, fiber, seed, stem, and root with highest expression in ovule at 20DPA (**Figure 6**). The fineness related trait Mtex gave maximum significant [log(p) > 6] association with the lead SNP i.e., Chr05_86618304. Its haplotypes showed significant differences from each other represented *via* boxplots. The gene expression findings provided a gene *Ga05G3785* in close vicinity of this lead SNP. This candidate gene presented high expression FPKM values which validated the selection of this gene as candidate one (**Figure 7**). The length related four traits namely, XJ_Ln5, XJ, XJ_LW, and XJ_UQLW displayed highly significant [log(p) > 6] associations with pleiotropic lead SNP Chr06_8554709 found at peak of Manhattan plots. The haplotypes boxplots of this lead SNP were significantly different. The relevant gene expression profiles showed a gene *Ga06G0559* in its close vicinity. This candidate gene presented its high expression FPKM values in ovule and seed tissues (**Figure 8**).

DISCUSSION

Cotton is a prominent natural source of fiber. The cultivated cotton species include both diploid and tetraploid genomes (Sarfraz et al., 2018). A general perception about present-day allotetraploid American Cotton (*Gossypium hirsutum)* confers to the diploid species, i.e., *Gossypium arboreum* and *Gossypium raimondii* (Cronn et al., 2002; Wendel and Cronn, 2003). Asian diploid cotton, a potential A-genome donor of upland cotton, is renowned for harboring many genetic factors coding high fiber quality-related features and resistances against several biotic and abiotic stresses (Shaheen et al., 2013). Due to the scarcity of available genetic divergence in the founder parents of global cotton cultivars, global climate change is posing continual threats to the development and survival of *G. hirsutum* cultivars. To restore the broad genetic base, it is a dire need to explore potential genetic diversity that might have eroded from the cultivated cotton collection during the breeding period to restore their broad genetic base.

The significance of fiber quality for premium textiles has prompted breeders to generate, and farmers to harvest new cotton varieties with superior fiber qualities. Several single fiber quality traits that HVI does not measure, i.e., fiber length distribution, short fiber content, maturity and fineness, have a considerable impact on processing performance. A critical concern in cotton research is the need for precise and accurate methods for measuring fiber quality traits. The dependency of maturity on fiber length, modeled using network analysis, confirmed their strong relationship. This relationship of

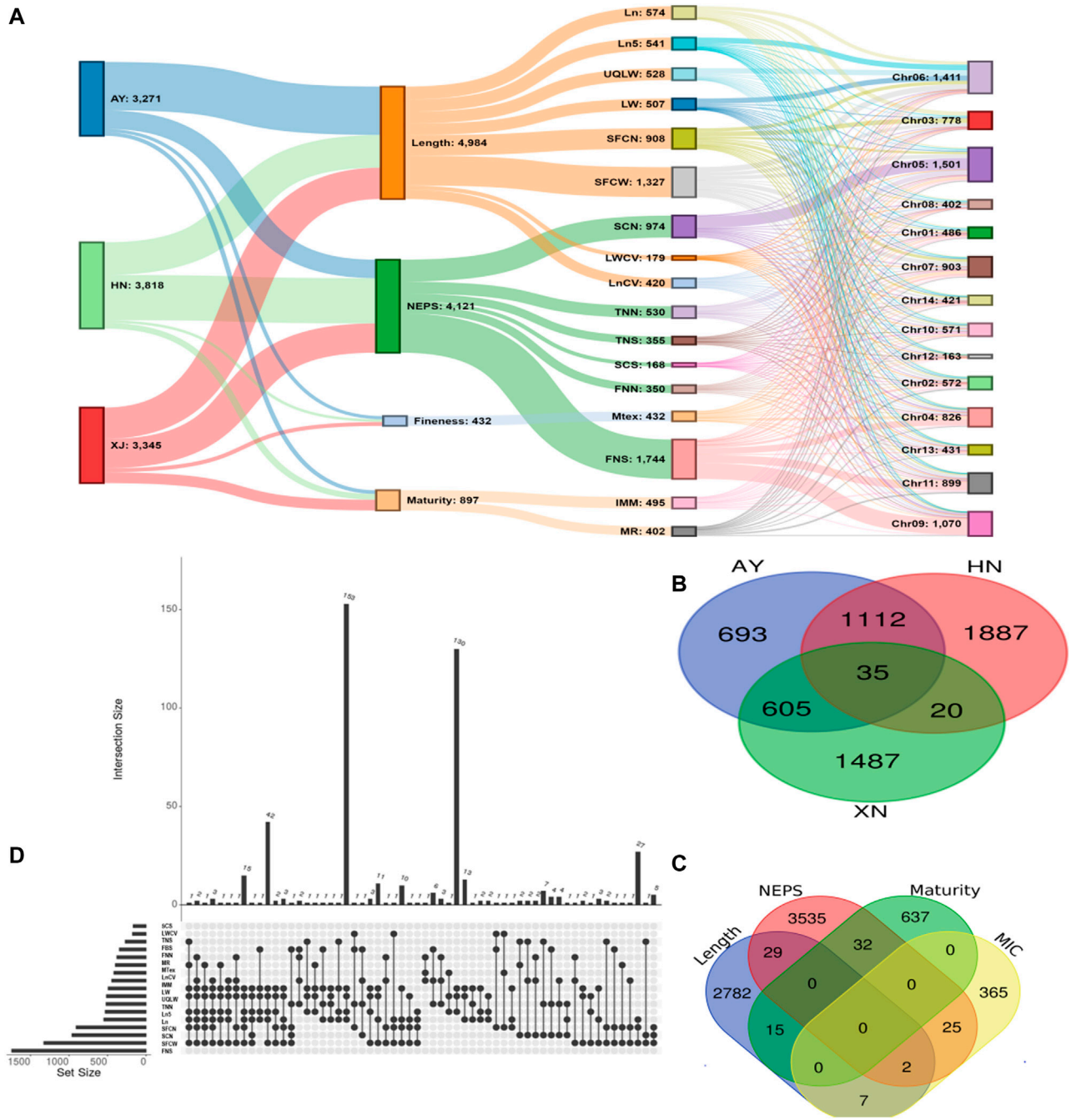


FIGURE 3 | GWAS Association Summary **(A)** Sankey diagram illustrating the number of significantly ($p < 0.00001$) associated SNPs among evaluated environments (AY: Anyang, HN: Sanya, Hainan, XJ: Xinjiang) and single fiber quality traits along with their distributions 13 chromosomes of Asiatic cotton **(B)** Venn diagram displaying the number of significantly ($p < 0.00001$) associated SNPs shared among three evaluated environments **(C)** Venn diagram depicting number of significantly ($p < 0.00001$) associated SNPs shared among four major categories of studied single fiber quality traits of *G. arboreum* accessions **(D)** Details of significantly associated SNPs showing pleiotropism among various investigated single fiber quality traits. All SNPs with one or more associated phenotypes are shown here.

length with fineness and maturity showed that the information for maturity and fineness of fiber is probably embedded in its length distribution (Paudel, 2012). It illustrates that the shorter fibers are immature than the longer fibers. Mature fibers have a secondary cell wall that is thicker in width, so less inclined to breakage during processing. Consequently, the breaking of immature fiber could increase short fibers for immature cotton.

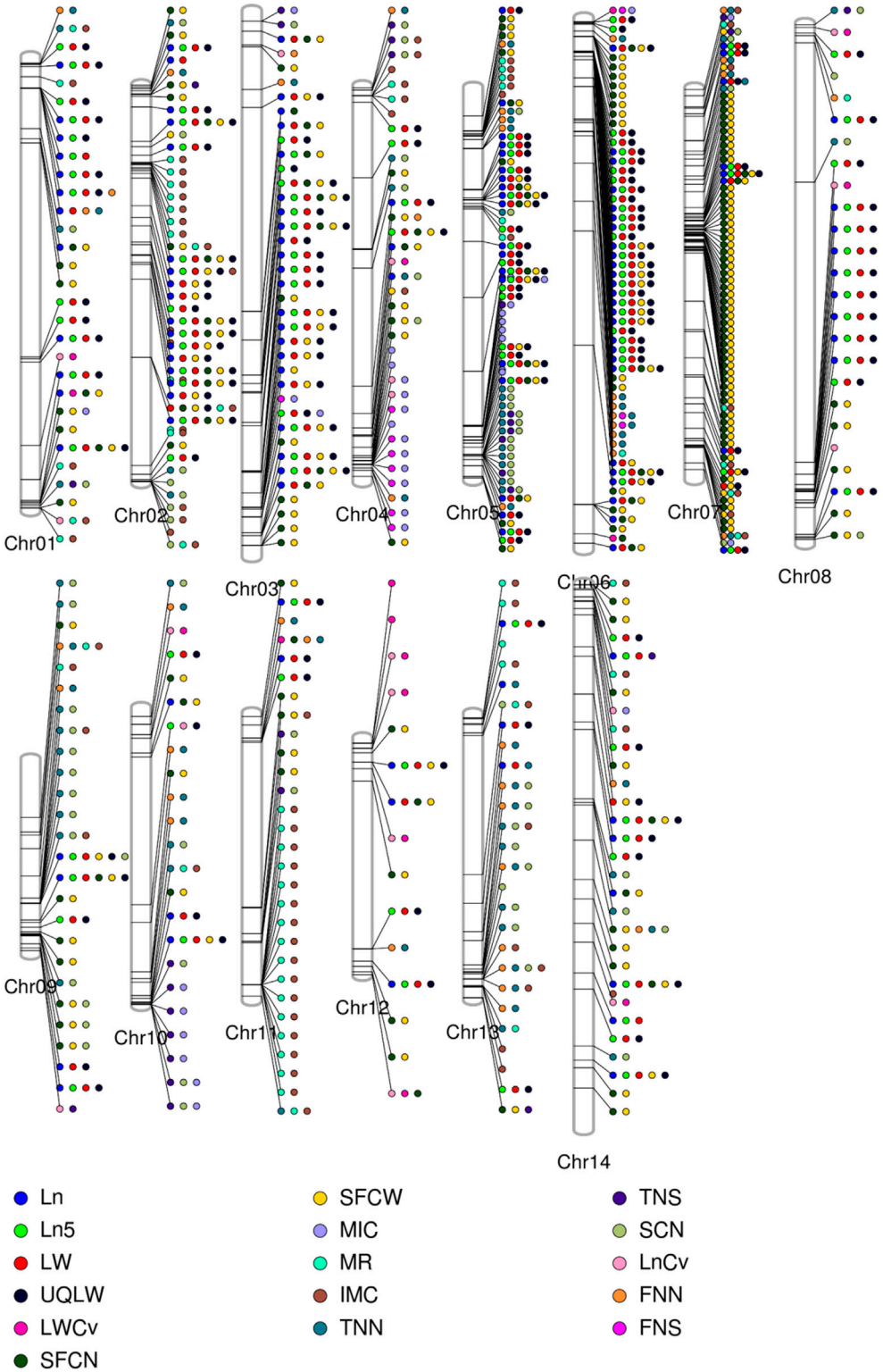


FIGURE 4 | Phenogram is displaying peak associations signals from 7429 significantly associated SNPs for single fiber quality traits across different chromosomes of Asiatic Cotton. Legends at the bottom are for distinguishing different studied traits.

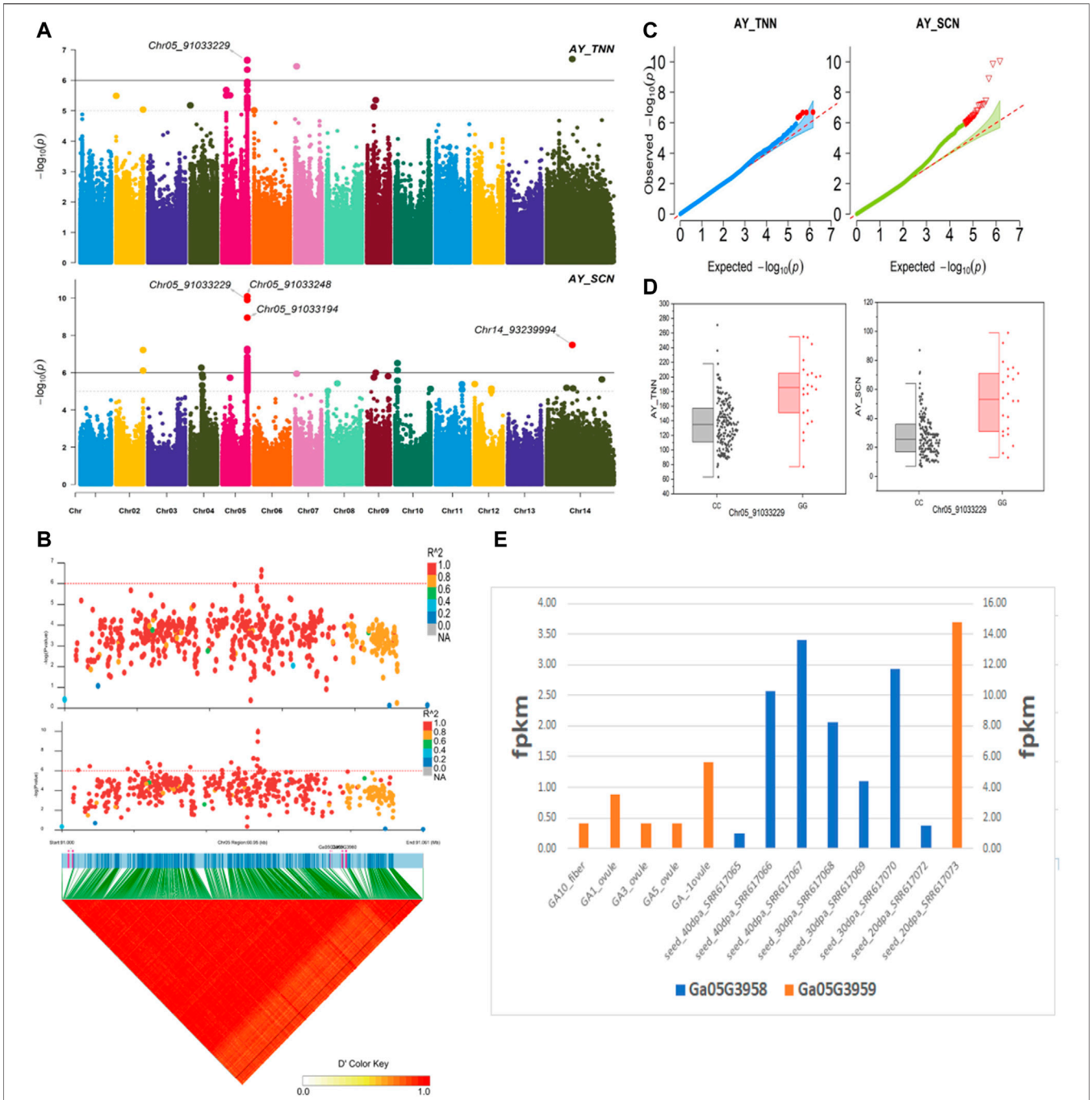


FIGURE 5 | Genome-wide association studies for NEPS in *G. arboreum* **(A)** Manhattan plots for NEPs related traits: AY_TNN and AY_SCN; dotted-horizontal line in each plot depicts significance threshold value kept at [–log (*p*) > 6] **(B)** GWAS-LD block demonstrating haplotype region (60.95 kb) surrounding the displaying peak on chromosome Chr05 associated with NEPs related traits along with GWAS regional plot exposed key SNPs and nearby gene(s) in the region **(C)** Quantile-Quantile plots for NEPs related traits **(D)** Box plots revealing differences for NEPs related traits among two favorable haplotypes of SNP Chr05_91033229:AY_TNN (left) and Chr05_91033229:AY_SCN (right) **(E)** Expression levels of representative annotated genes *Ga05G3958* and *Ga05G3959* at different developmental stages of *G. arboreum*.

The associations among the *G. arboreum* accessions were evaluated through hierarchical clustering based on genetic distance. There were three distinct clusters. The distribution of accessions in the distant Clusters 2 and 3 reflect their sizeable geographic distance, i.e., Southwest China and Northern China. Additionally, the closeness between Cluster 1 and Cluster 3 confirms that *G. arboreum* originated and expanded into the Yangtze River region from South China. A general inference

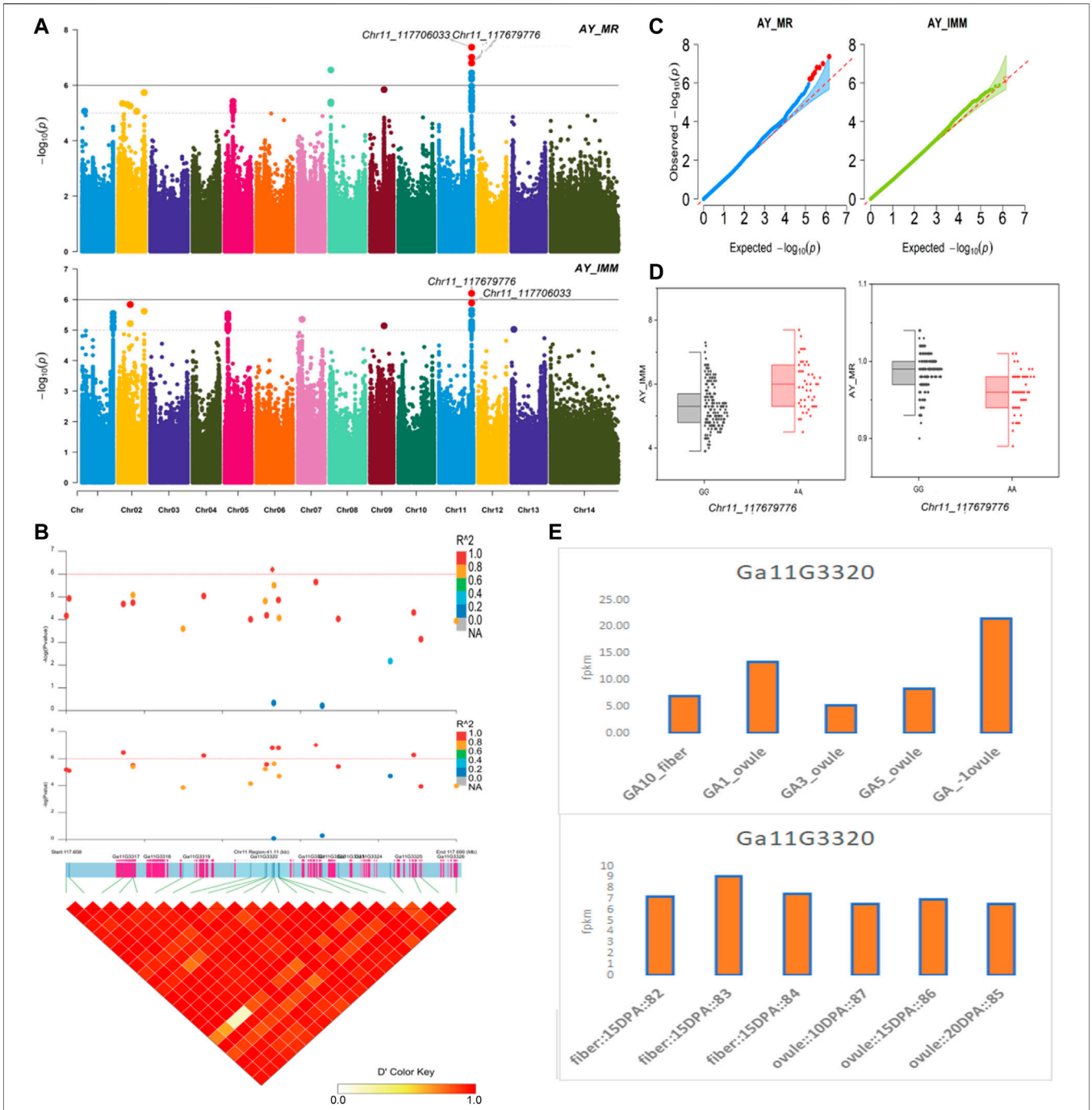


FIGURE 6 | Genome-wide association studies for Maturity in *G. arboreum* **(A)** Manhattan plots for Maturity related traits: AY_MR and AY_IMM; dotted-horizontal line in each plot illustrates significance threshold value kept at [–log (*p*) > 6] **(B)** GWAS-LD block revealing haplotype region (41.11 kb) surrounding the exposed peak on chromosome Chr11 associated with Maturity related traits along with GWAS regional plot presenting key SNPs and nearby gene(s) in the region **(C)** Quantile-Quantile plots for Maturity related traits **(D)** Box plots demonstrating differences for Maturity related traits among two favorable haplotypes of SNP Chr11_117679776: AY_IMM (left) and Chr11_117679776:AY_MR (right) **(E)** Expression levels of representative annotated gene *Ga11G3320* at different developmental stages of *G. arboreum* from two different sources.

drawn from this finding was that *G. arboreum* species first originated in South China and then extended to the Yellow and Yangtze River regions (Guo et al., 2006). The three clusters, with Clusters 1 and 2 being close together and Cluster 3 being farther away, were also revealed by PCA. This clustering corroborates earlier findings (Yinhua et al., 2018).

The efficacy of GWAS decreases when the variations in the population under study increase (Tyagi et al., 2014). Also, the

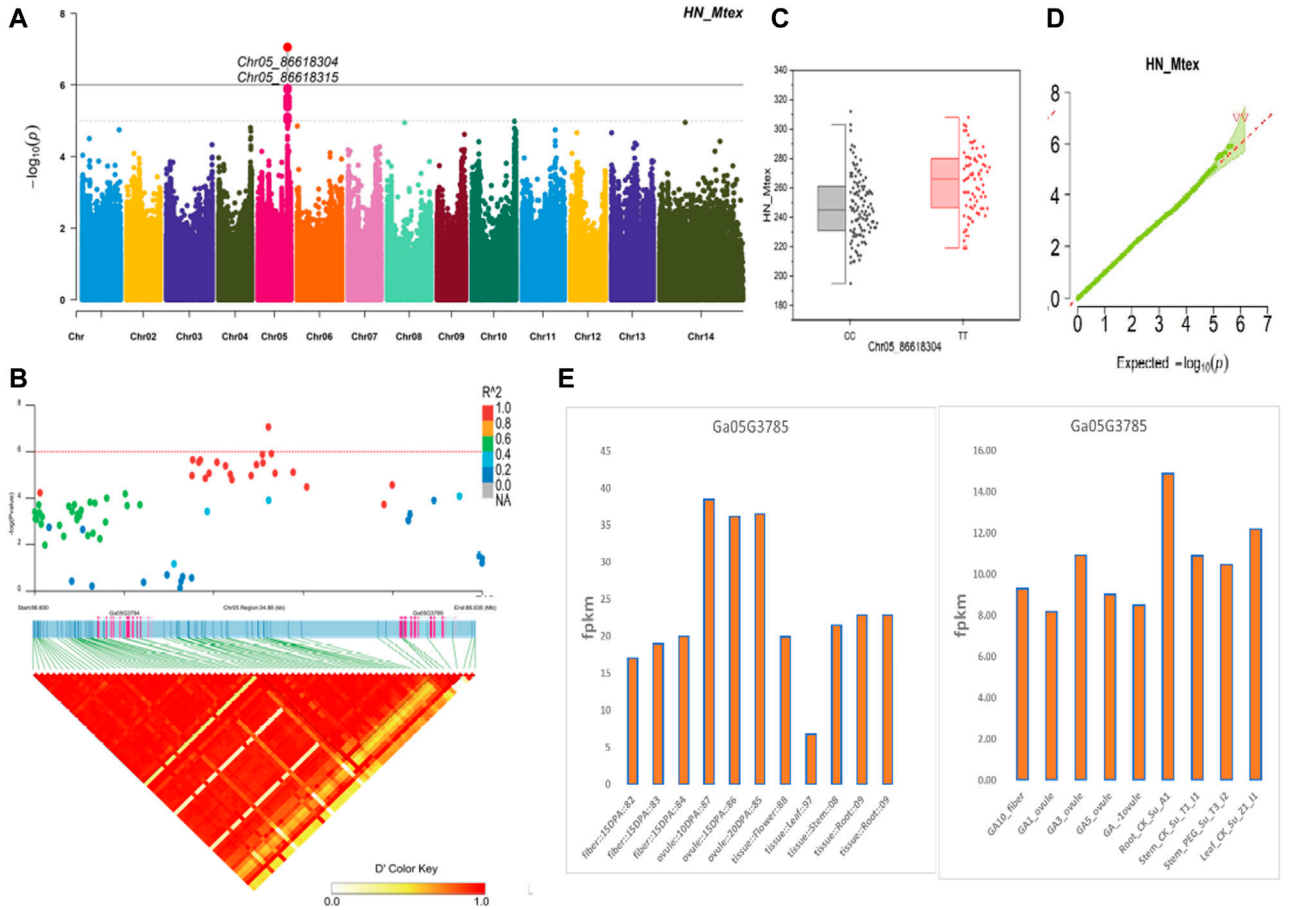


FIGURE 7 | Genome-wide association studies for Fineness in *G. arboreum* **(A)** Manhattan plots for Fineness related trait HN_Mtex; dotted-horizontal line describes significance threshold value kept at [–log (*p*) > 6] **(B)** GWAS-LD block displaying haplotype region (34.88 kb) surrounding the displaying peak on chromosome Chr05 associated with Fineness related traits along with GWAS regional plot revealing key SNPs and nearby gene(s) in the region **(C)** Quantile-Quantile plots for Fineness related traits **(D)** Box plots presenting differences for Fineness related traits among two favorable haplotypes of SNP Chr05_86618304:HN_Mtex **(E)** Expression levels of representative annotated gene *Ga05G3785* at different developmental stages of *G. arboreum* from two different sources.

power of structure-based studies to detect single gene effects decreases with increases in population differences (Flint-Garcia et al., 2005). *G. arboreum* is better than tetraploid cotton for identifying trait-related genes *via* GWAS due to its low population differentiation and smaller genome. Hence, *G. arboreum* is useful for exploring population differences and genetic diversity to facilitate identifying genes significantly associated with critical traits (Yinhua et al., 2018).

LD analysis plays a unique role in GWAS in evaluating the density and number of associated markers between loci where LD persists. Factors affecting LD include population size, genetic diversity, admixture level, marker system, mating design, and selection techniques (Flint-Garcia et al., 2003; Stich et al., 2006). The LD decay distance in cross-pollinated crops such as maize (1–100 kb) is lower than in self-pollinated species such as cotton (Flint-Garcia et al., 2003), and also *Arabidopsis thaliana* (~250 kb) (Nordborg et al., 2002), rice (75–500 kb) (Mather et al., 2007), and soybean (100–600 kb) (Hyten et al., 2007). However, LD decay was 25 cM in tetraploid cotton based on microsatellites (Abdurakhmonov et al., 2009). We determined an LD decay distance of ~115.5 kb with r^2 = 0.42, similar to other reports on self-pollinated species. In the last two decades, GWAS has been extensively used to map various quantitative traits in plants, being considered an important milestone (Zhao et al., 2014). The power of GWAS depends on rich genetic diversity, accurate phenotypic data, marker density, and adequate statistical methods.

The gene *Ga05G3958* found in close vicinity of lead SNP Chr05_91033229 associated with NEPs related traits, encodes a protein S-norcoclaurine synthase 2 (Q4QTJ1) plays role in defense mechanism again pathogens reported earlier in *Opium poppy* (Lee and Facchini, 2010). The expression of this protein was found in its roots, leaves, stem, flower buds and germinating seeds (Samanani and Facchini, 2001) like our findings having this gene with higher expression in seed tissues at different stages for seed coat nep (SCN) and total nep count (TNN). The other NEPs related gene *Ga05G3959* encodes Nucleoporin nup107 protein reported in Fission yeast. As the name indicates, the protein is

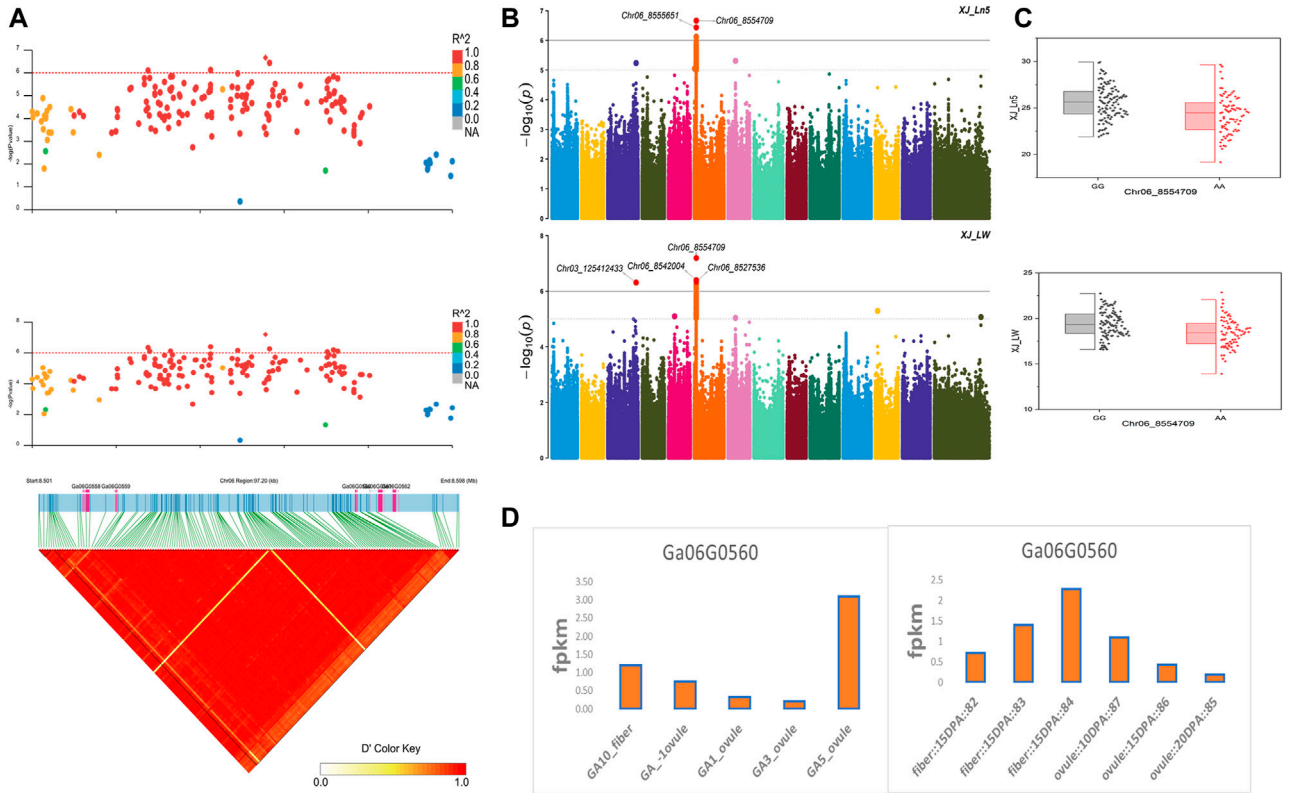


FIGURE 8 | Genome-wide association studies for Length in *G. arboreum* **(A)** GWAS-LD block demonstrating haplotype region (97.20 kb) surrounding the exposed peak on chromosome Chr06 associated with Length related traits along with GWAS regional plot revealing key SNPs and nearby gene(s) in the region **(B)** Manhattan plots for Length related traits: XJ_Ln5 and XJ_LW dotted-horizontal line in each plot describes significance threshold value kept at [–log (*p*) > 6] **(C)** Box plots exhibiting differences for Length related traits among two favorable haplotypes of SNP Chr06_8554709: XJ_Ln5 and Chr06_8554709: XJ_LW **(D)** Expression levels of representative annotated gene *Ga06G0559* at different developmental stages of *G. arboreum* from two different sources.

located in nucleus and related functions involved transport of mRNA, rRNA, and various proteins across nuclear envelope (Gaudet et al., 2011). In our findings, this gene gave its higher expression in seed tissues at various stages. The gene *Ga05G3785* was found in vicinity of lead SNPs related to fineness gave a higher expression in all tissues of fiber, ovule, and seed at different stages. It was previously described in *Arabidopsis* encoding for DJ-1 homolog B (DJ1B) protein which is involved in the oxidative stress response (Kwon et al., 2013). Two genes *Ga03G2389* and *Ga03G2390* were observed in close vicinity of lead SNPs related to length and NEPs, respectively. They were previously described in *Arabidopsis*. *Ga03G2389* encodes LIM domain-containing protein (WLIM1) which binds to actin filaments to promote cross-links to form thick bundles (Papuga et al., 2010), strengthening the fibers and ultimately may play roles in its lengthening. *Ga03G2390* encodes ADP-ribosylation factor-like A1C protein from GTPase family (Nishikiori et al., 2011; Du et al., 2018), involves in GTP binding is located in nucleus, plasma membrane, male gametophyte (pollen tube) during maturation, and fiber during development stages. A length related gene *Ga06G0559* was found in vicinity of Chr06_8554709 lead SNP encodes Sporulation-specific glucan 1,3-beta-glucosidase (SPR1) protein found in Baker's yeast. This enzyme is expressed in later stages of sporulation for the modification of glucan linkages in order to strengthen the ascospore wall or providing it plasticity (Gaudet et al., 2011). The reported function of cell wall organization in yeast spores is validating this gene association with our length related traits for provision of shape, strength, and plasticity to fiber.

Ga02G1729, *Ga02G1738*, *Ga02G1741*, *Ga08G0324*, *Ga11G3317*, *Ga11G3319*, *Ga11G3320*, and *Ga11G3321* were found nearby the lead SNPs related to NEPs and maturity. *Ga08G0324*, *Ga02G1738*, *Ga11G3317*, *Ga11G3319*, and *Ga11G3320* were previously described in *Arabidopsis*. *Ga02G1729* encodes RETICULATA protein and is located in chloroplastic DNA. It may play a prominent role in leaf development as it is key for mesophyll cell division during initial leaf organogenesis. It is highly expressed during embryo development and in leaf primordia, margins of fully expanded leaves, stipules, lamina, root tips, and stamens (Barth and Conklin, 2003; González-Bayon et al., 2006; Pérez-Pérez et al., 2013). *Ga08G0324* encodes DNA damage and repair/toleration protein (DRT100). *Ga02G1738* encodes mediator of RNA polymerase II transcription subunit protein. It regulates flowering time and plant defense and is involved in pollen tube growth (Lalanne et al., 2004; Kidd et al., 2009).

Ga11G3317 (non-synonymous mutation) encodes FAR1-related sequence 5 (FRS5) protein. It is involved in zinc ion binding and transcription regulation and is expressed in hypocotyl tissues, leaves, stems, and flowers. It is upregulated in hypocotyl tissues (Lin and Wang, 2004). *Ga11G3320* encodes membrane-anchored ubiquitin-fold protein 1 (MUB1). It is involved in stability at temperatures >90°C and is located at the plasma membrane (Downes et al., 2006). *Ga11G3319* (non-synonymous mutation) encodes galacturonosyltransferase 13 (GAUT13). It is involved in pectin biosynthesis in cell walls, pollen tube growth, and pollen development, and it is expressed in flowers, roots, stems, and leaves (Caffall et al., 2009).

GO annotation revealed that the abovementioned genes were associated with several biological processes, cellular components, and molecular functions. These genes encode for proteins involved in the transport/metabolism of amino acids, coenzymes, inorganic ions, lipids and carbohydrates, cell wall/membrane/envelope/ribosomal structure and biogenesis (deposition), energy production or conversion, intercellular trafficking, secretion, vesicular transport, transcription, translation, post-translational modification, protein turnover, chaperons, replication, recombination, repair, signal transduction mechanisms, and general functions prediction.

CONCLUSION

Continuous improvements of cotton fiber quality are required to maintain the superiority of cotton fiber over manmade yarn. Significant research is needed to improve the measurement accuracy of key cotton fiber traits. Basically, 4 categories of 17 AFIS-related fiber quality traits, including NEPS, Fiber Length, Fiber Fineness and Fiber maturity, were evaluated in the current study. Rapid and precise measurement of quality traits will help cotton breeders quickly select key traits to develop varieties with superior fiber quality suitable for industrial use. A substantial amount of highly significant SNP markers (QTNs) for these traits were identified and further validated *via* gene expression analysis. Highly significant genes present in the vicinity of these key SNPs were considered as candidate genes. These compendia connecting traits, genes and cell types may allow further prioritization of genes in GWAS loci to enable mechanistic studies. These identified QTNs can possibly be helpful to cotton breeders regarding fiber quality improvement as well as revival of eroded genetic factors of *G. hirsutum via* introgression and marker-assisted breeding approaches.

DATA AVAILABILITY STATEMENT

All raw sequencing available at the NCBI BioProject database under accession number PRJNA349094.

AUTHOR CONTRIBUTIONS

XD obtained the funding; XD and SW conceived the study; SW and XD supervised the study; MI, ST, SH, YJ, and ZP executed the study; ZP, LW, MSI, ZS, BC, and ST extracted the DNA; SH, GS, and MI constructed the Illumina sequencing libraries; ST, MSI, HL, GS, ZP, LW, BP, MN, HL, BC, and GX collected and compiled the data; MI, SH, GS, ZS, AM, and XD analyzed and interpreted the data; ZS, MSI, and MN constructed the figures and tables; MI and ZS drafted the manuscript; SW, XD, SA, SH, ST, HL, AM, and ZS provided helpful edits regarding the manuscript; all authors critically reviewed and approved the final manuscript.

FUNDING

This research was supported by grants from The National Key Research and Development Program of China (grant no. 2016YFD0102105) and Central Public-Interest Scientific Institution Basal Research Fund (grant no. 1610162019010100). The funders had no role in the design or conduct of the study, the collection, management, analysis, or interpretation of the data, the preparation, review, or approval of the manuscript, or the decision to submit the manuscript for publication.

ACKNOWLEDGMENTS

We thank the National Mid-term Genebank for Cotton at the Institute of Cotton Research of the Chinese Academy of Agricultural Sciences for providing the seeds.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.758665/full#supplementary-material.

Supplementary Figure S1 | Summary of correlations among different fiber quality traits of *G. arboreum* represented in upper triangle. The density distribution of all accessions for different traits evaluated in different environments is displayed at diagonal with distinct colors (Pink: Anyang, Green: Henan, Blue: XJ: Xinjiang). In the lower triangle, bivariate scatterplots of all investigated traits are presented. Boxplots are illustrating the presence of variability among all experimented accessions. Central box displays middle half data extending from upper to lower quartile whereas, a horizontal line is drawn to show the median. The endpoints of vertical projections designate minimum and maximum data points until the presence of outliers. The rows at the bottom depict frequency distributions of each environment for all the traits.

Supplementary Figure S2 | Quantitative Trait Nucleotides (QTNs) rich regions with significant ($p < 0.00001$) associations among fiber quality traits across all the chromosomes of Asiatic Cotton.

Supplementary Figure S3 | Heat map of gene expression (fpkm) levels of 113 genes in the vicinity of lead SNPs based on transcriptome data from two sources.

REFERENCES

Abdurakhmonov, I. Y., Saha, S., Jenkins, J. N., Buriev, Z. T., Shermatov, S. E., Scheffler, B. E., et al. (2009). Linkage Disequilibrium Based Association Mapping of Fiber Quality Traits in G. Hirsutum L. Variety Germplasm. *Genetica* 136, 401–417. doi:10.1007/s10709-008-9337-8

Barth, C., and Conklin, P. L. (2003). The Lower Cell Density of Leaf Parenchyma in theArabidopsis Thalianamutantlcd1-1is Associated with Increased Sensitivity to Ozone and virulentPseudomonas Syringae. *Plant J.* 35, 206–218. doi:10.1046/j.1365-313x.2003.01795.x

Batool, S., Khan, N. U., Makhdoom, K., Bibi, Z., Hassan, G., Marwat, K. B., et al. (2010). Heritability and Genetic Potential of upland Cotton Genotypes for Morpho-Yield Traits. *Pak. J. Bot.* 42, 1057–1064.

Berkley, E. E. (1948). Cotton-A Versatile Textile Fiber. *Text. Res. J.* 18, 71–88. doi:10.1177/004051754801800201

Bradow, J. M., Hinojosa, O., Wartelle, L. H., Davidonis, G., Sassenrath-Cole, G. F., and Bauer, P. J. (1996). Applications of AFIS Fineness and Maturity Module and X-ray Fluorescence Spectroscopy in Fiber Maturity Evaluation. *Text. Res. J.* 66, 545–554. doi:10.1177/004051759606600902

Caffall, K. H., Pattathil, S., Phillips, S. E., Hahn, M. G., and Mohnen, D. (2009). *Arabidopsis thaliana* T-DNA Mutants Implicate GAUT Genes in the Biosynthesis of Pectin and Xylan in Cell walls and Seed Testa. *Mol. Plant* 2, 1000–1014. doi:10.1093/mp/ssp062

Cronn, R., Cedroni, M., Haselkorn, T., Grover, C., and Wendel, J. F. (2002). PCR-mediated Recombination in Amplification Products Derived from Polyploid Cotton. *Theor. Appl. Genet.* 104, 482–489. doi:10.1007/s001220100741

Dong, S.-S., He, W.-M., Ji, J.-J., Zhang, C., Guo, Y., and Yang, T.-L. (2021). LDBlockShow: A fast and Convenient Tool for Visualizing Linkage Disequilibrium and Haplotype Blocks Based on Variant Call Format Files. *Brief. Bioinform.* 22 (4), bbaa227. doi:10.1093/bib/bbaa227

Downes, B. P., Saracco, S. A., Lee, S. S., Crowell, D. N., and Vierstra, R. D. (2006). MUBs, a Family of Ubiquitin-fold Proteins that Are Plasma Membrane-Anchored by Prenylation. *J. Biol. Chem.* 281, 27145–27157. doi:10.1074/jbc.m602283200

Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., et al. (2018). Resequencing of 243 Diploid Cotton Accessions Based on an Updated A Genome Identifies the Genetic Basis of Key Agronomic Traits. *Nat. Genet.* 50, 796–802. doi:10.1038/s41588-018-0116-x

Erpelding, J. E., and Stetina, S. R. (2013). Genetics of Reniform Nematode Resistance in Gossypium Arboreum Germplasm Line PI 529728. *World J. Agric. Res.* 1, 48–53.

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* 164, 1567–1587. doi:10.1093/genetics/164.4.1567

Felsenstein, J. (1993). *PHYLIP (Phylogeny Inference Package) version 3.5 c. Distributed by the author*. Seattle: Department of Genetics, University of Washington.

Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et al. (2005). Maize Association Population: a High-Resolution Platform for Quantitative Trait Locus Dissection. *Plant J.* 44, 1054–1064. doi:10.1111/j.1365-313X.2005.02591.x

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. (2003). Structure of Linkage Disequilibrium in Plants. *Annu. Rev. Plant Biol.* 54, 357–374. doi:10.1146/annurev.arplant.54.031902.134907

Frydrych, I., and Thibodeaux, D. (2010). *Fiber Quality Evaluation-Current and Future Trends/intrinsic Value of Fiber Quality in cottonCotton: Technology for the 21st century*. Washington DC: International Cotton Advisory Committee, 251–296.

Gaudet, P., Livstone, M. S., Lewis, S. E., and Thomas, P. D. (2011). Phylogenetic-based Propagation of Functional Annotations within the Gene Ontology Consortium. *Brief. Bioinf.* 12, 449–462. doi:10.1093/bib/bbr042

González-Bayón, R., Kinsman, E. A., Quesada, V., Vera, A., Robles, P., Ponce, M. R., et al. (2006). Mutations in the RETICULATA Gene Dramatically Alter Internal Architecture but Have Little Effect on Overall Organ Shape in Arabidopsis Leaves. *J. Exp. Bot.* 57, 3019–3031. doi:10.1093/jxb/erl063

Guo, W.-Z., Zhou, B.-L., Yang, L.-M., Wang, W., and Zhang, T.-Z. (2006). Genetic Diversity of Landraces in Gossypium Arboreum L. Race Sinense Assessed with Simple Sequence Repeat Markers. *J. Integr. Plant Biol.* 48, 1008–1017. doi:10.1111/j.1744-7909.2006.00316.x

Hedin, P. A., Jenkins, J. N., and Parrott, W. L. (1992). Evaluation of Flavonoids inGossypium Arboreum (L.) Cottons as Potential Source of Resistance to Tobacco Budworm. *J. Chem. Ecol.* 18, 105–114. doi:10.1007/bf00993746

Hequet, E. F., Wyatt, B., Abidi, N., and Thibodeaux, D. P. (2006). Creation of a Set of Reference Material for Cotton Fiber Maturity Measurements. *Text. Res. J.* 76, 576–586. doi:10.1177/0040517506064710

Hyten, D. L., Choi, I.-Y., Song, Q., Shoemaker, R. C., Nelson, R. L., Costa, J. M., et al. (2007). Highly Variable Patterns of Linkage Disequilibrium in Multiple Soybean Populations. *Genetics* 175, 1937–1944. doi:10.1534/genetics.106.069740

Iqbal, M. A., Ammad, A., and Zafar, Y. (2015). Characterization of Indigenous Gossypium Arboreum L. Genotypes for Various Fiber Quality Traits. *Pakistan J. Bot.* 47, 2347–2354.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., et al. (2010). Variance Component Model to Account for Sample Structure in Genome-wide Association Studies. *Nat. Genet.* 42, 348–354. doi:10.1038/ng.548

Khan, N. U., Basal, H., and Hassan, G. (2010). Cottonseed Oil and Yield Assessment via Economic Heterosis and Heritability in Intraspecific Cotton Populations. *Afr. J. Biotechnol.* 9, 7418–7428. doi:10.5897/ajb10.915

Kidd, B. N., Edgar, C. I., Kumar, K. K., Aitken, E. A., Schenk, P. M., Manners, J. M., et al. (2009). The Mediator Complex Subunit PFT1 Is a Key Regulator of Jasmonate-dependent Defense inArabidopsis. *Plant Cel.* 21, 2237–2252. doi:10.1105/tpc.109.066910

Kwon, K., Choi, D., Hyun, J. K., Jung, H. S., Baek, K., and Park, C. (2013). Novel Glyoxalases fromArabidopsis Thaliana. *Febs J.* 280, 3328–3339. doi:10.1111/febs.12321

Lalanne, E., Michaelidis, C., Moore, J. M., Gagliano, W., Johnson, A., Patel, R., et al. (2004). Analysis of Transposon Insertion Mutants Highlights the Diversity of Mechanisms Underlying Male Progamic Development in Arabidopsis. *Genetics* 167, 1975–1986. doi:10.1534/genetics.104.030270

Lee, E.-J., and Facchini, P. (2010). Norcoclaurine Synthase Is a Member of the Pathogenesis-Related 10/Bet V1 Protein Family. *Plant Cel.* 22, 3489–3503. doi:10.1105/tpc.110.077958

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., et al. (2014). Genome Sequence of the Cultivated Cotton Gossypium Arboreum. *Nat. Genet.* 46, 567–572. doi:10.1038/ng.2987

Lin, R., and Wang, H. (2004). Arabidopsis FHY3/FAR1 Gene Family and Distinct Roles of its Members in Light Control of Arabidopsis Development. *Plant Physiol.* 136, 4010–4022. doi:10.1104/pp.104.052191

Maqbool, A., Abbas, W., Rao, A. Q., Irfan, M., Zahur, M., Bakhsh, A., et al. (2010). Gossypium Arboreum GHSP26 Enhances Drought Tolerance in Gossypium Hirsutum. *Biotechnol. Prog.* 26, 21–25. doi:10.1002/btpr.306

Mather, K. A., Caicedo, A. L., Polato, N. R., Olsen, K. M., Mccouch, S., and Purugganan, M. D. (2007). The Extent of Linkage Disequilibrium in rice (Oryza Sativa L.). *Genetics* 177, 2223–2232. doi:10.1534/genetics.107.079616

Mehetre, S., Aher, A., Gawande, V., Patil, V., and Mokate, A. (2003). Induced Polyploidy in Gossypium: a Tool to Overcome Interspecific Incompatibility of Cultivated Tetraploid and Diploid Cottons. *Curr. Sci.* 84, 1510–1512.

Meredith, W. R., Jr (2005). Minimum Number of Genes Controlling Cotton Fiber Strength in a Backcross Population. *Crop Sci.* 45, 1114–1119. doi:10.2135/cropsci2003.0425

Nazeer, W., Tipu, A. L., Ahmad, S., Mahmood, K., Mahmood, A., and Zhou, B. (2014). Evaluation of Cotton Leaf Curl Virus Resistance in BC1, BC2, and BC3 Progenies from an Interspecific Cross between Gossypium Arboreum and Gossypium Hirsutum. *PloS one* 9, e111861. doi:10.1371/journal.pone.0111861

Nishikiori, M., Mori, M., Dohi, K., Okamura, H., Katoh, E., Naito, S., et al. (2011). A Host Small GTP-Binding Protein ARL8 Plays Crucial Roles in Tobamovirus RNA Replication. *Plos Pathog.* 7, e1002409. doi:10.1371/journal.ppat.1002409

Nordborg, M., Borevitz, J. O., Bergelson, J., Berry, C. C., Chory, J., Hagenblad, J., et al. (2002). The Extent of Linkage Disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 30, 190–193. doi:10.1038/ng813

Papuga, J., Hoffmann, C., Dieterle, M., Moes, D., Moreau, F., Tholl, S., et al. (2010). ArabidopsisLIM Proteins: A Family of Actin Bundlers with Distinct Expression Patterns and Modes of Regulation. *Plant cel.* 22, 3034–3052. doi:10.1105/tpc.110.075960

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, 423–427. doi:10.1038/nature11798

Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *Plos Genet.* 2, e190. doi:10.1371/journal.pgen.0020190

Paudel, D. (2012). *Evaluating the Potential of New Testing Methods for Cotton (Gossypium Hirsutum L.) Breeding*. Texas Tech University. Available at: http://hdl.handle.net/2346/45622

Pérez-Pérez, J. M., Esteve-Bruna, D., González-Bayón, R., Kangasjärvi, S., Caldana, C., Hannah, M. A., et al. (2013). Functional Redundancy and Divergence within the Arabidopsis RETICULATA-RELATED Gene Family. *Plant Physiol.* 162, 589–603. doi:10.1104/pp.113.217323

Pearson, T. A., and Manolio, T. J. J. (2008). How to Interpret a Genome-wide Association Study. *Jama* 299, 1335–1344. doi:10.1001/jama.299.11.1335

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., 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

Sacks, E. J., and Robinson, A. F. (2009). Introgression of Resistance to Reniform Nematode (Rotylenchulus Reniformis) into upland Cotton (Gossypium Hirsutum) from Gossypium Arboreum and a G. hirsutum/Gossypium Aridum Bridging Line. *Field Crops Res.* 112, 1–6. doi:10.1016/j.fcr.2009.01.006

Samanani, N., and Facchini, P. J. (2001). Isolation and Partial Characterization of Norcoclaurine Synthase, the First Committed Step in Benzylisoquinoline Alkaloid Biosynthesis, from Opium Poppy. *Planta* 213, 898–906. doi:10.1007/s004250100581

Sarfraz, Z., Iqbal, M. S., Pan, Z., Jia, Y., He, S., Wang, Q., et al. (2018). Integration of Conventional and Advanced Molecular Tools to Track Footprints of Heterosis in Cotton. *BMC Genomics* 19, 776. doi:10.1186/s12864-018-5129-4

Shaheen, T., Zafar, Y., and Rahman, M.-U. (2013). QTL Mapping of Some Productivity and Fibre Traits in Gossypium Arboreum. *Turk J. Bot.* 37, 802–810. doi:10.3906/bot-1209-47

Shofner, F. M., Chu, Y.-T., and Thibodeaux, D. P. (1990). "An Overview of the Advanced Fiber Information System," in Proc. Int. Cotton Conf., Faserinstitut, Bremen, Germany, March 15, 1990, 173–181.

Stanton, M., Stewart, J. M., and Tugwell, N. (1992). Evaluation of Gossypium Arboreum L. Germplasm for Resistance to Thrips. *Genet. Resour. Crop Evol.* 39, 89–95.

Stich, B., Melchinger, A. E., Piepho, H.-P., Heckenberger, M., Maurer, H. P., and Reif, J. C. (2006). A New Test for Family-Based Association Mapping with Inbred Lines from Plant Breeding Programs. *Theor. Appl. Genet.* 113, 1121–1130. doi:10.1007/s00122-006-0372-5

Tahir, M. S., and Khan, N.-U. -I. (2011). Development of an Interspecific Hybrid (Triploid) by Crossing Gossypium Hirsutum and G. Arboreum. *Cytologia* 76, 193–199. doi:10.1508/cytologia.76.193

Thibodeaux, D., Senter, H., Knowlton, J., Mcalister, D., and Cui, X. (2007). "Measuring the Short Fiber Content of Cotton," in Cotton: Nature's High-Tech Fiber. Proc. World Cotton Res. Conf.-4, Lubbock, TX, September 10, 2007, 10–14.

Tyagi, P., Gore, M. A., Bowman, D. T., Campbell, B. T., Udall, J. A., and Kuraparthy, V. (2014). Genetic Diversity and Population Structure in the US Upland Cotton (Gossypium Hirsutum L.). *Theor. Appl. Genet.* 127, 283–295. doi:10.1007/s00122-013-2217-3

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38, e164. doi:10.1093/nar/gkq603

Wendel, J. F., and Cronn, R. C. (2003). Polyploidy and the Evolutionary History of Cotton. *Adv. Agron.* 78, 139–186. doi:10.1016/s0065-2113(02)78004-8

Yang, Q., Cui, J., Chazaro, I., Cupples, L. A., and Demissie, S. (2005). Power and Type I Error Rate of False Discovery Rate Approaches in Genome-wide Association Studies. *BMC Genet.* 6 Suppl 1 (Suppl. 1), S134. doi:10.1186/1471-2156-6-S1-S134

Yinhua, J., Zhaoe, P., Shoupu, H., Wenfang, G., Xiaoli, G., Baoyin, P., et al. (2018). Genetic Diversity and Population Structure of Gossypium Arboreum L. Collected in China. *J. Cotton Res.* 1, 1–8. doi:10.1186/s42397-018-0011-0

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J, F., et al. (2006). A Unified Mixed-Model Method for Association Mapping that Accounts for Multiple Levels of Relatedness. *Nat. Genet.* 38, 203–208.

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2019). PopLDdecay: a Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files. *Bioinformatics* 35, 1786–1788. doi:10.1093/bioinformatics/bty875

Zhao, Y., Wang, H., Chen, W., and Li, Y. (2014). Genetic Structure, Linkage Disequilibrium and Association Mapping of Verticillium Wilt Resistance in Elite Cotton (Gossypium Hirsutum L.) Germplasm Population. *Plos one* 9, e86308. doi:10.1371/journal.pone.0086308

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 December 2021
doi: 10.3389/fgene.2021.807210

Meta-QTL Analysis in Rice and Cross-Genome Talk of the Genomic Regions Controlling Nitrogen Use Efficiency in Cereal Crops Revealing Phylogenetic Relationship

Nitika Sandhu[1]†, Gomsie Pruthi[1†], Om Prakash Raigar[1†], Mohini Prabha Singh[1†], Kanika Phagna[2], Aman Kumar[1], Mehak Sethi[1], Jasneet Singh[1], Pooja Ankush Ade[1] and Dinesh Kumar Saini[1]*

[1]Punjab Agricultural University, Ludhiana, India, [2]Indian Institute of Science Education and Research, Berhampur, India

OPEN ACCESS

Edited by:
Ahmed Sallam,
Assiut University, Egypt

Reviewed by:
Jose Miguel Soriano,
Institute of Agrifood Research and Technology (IRTA), Spain
Agata Gadaleta,
University of Bari Aldo Moro, Italy
Karansher Singh Sandhu,
Washington State University, United States

***Correspondence:**
Nitika Sandhu
nitikasandhu@pau.edu

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 01 November 2021
Accepted: 30 November 2021
Published: 21 December 2021

Citation:
Sandhu N, Pruthi G, Prakash Raigar O, Singh MP, Phagna K, Kumar A, Sethi M, Singh J, Ade PA and Saini DK (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. 12:807210. doi: 10.3389/fgene.2021.807210

The phenomenal increase in the use of nitrogenous fertilizers coupled with poor nitrogen use efficiency is among the most important threats to the environment, economic, and social health. During the last 2 decades, a number of genomic regions associated with nitrogen use efficiency (NUE) and related traits have been reported by different research groups, but none of the stable and major effect QTL have been utilized in the marker-assisted introgression/pyramiding program. Compiling the data available in the literature could be very useful in identifying stable and major effect genomic regions associated with the root and NUE-related trait improving the rice grain yield. In the present study, we performed meta-QTL analysis on 1,330 QTL from 29 studies published in the past 2 decades. A total of 76 MQTL with a stable effect over different genetic backgrounds and environments were identified. The significant reduction in the confidence interval of the MQTL compared to the initial QTL resulted in the identification of annotated and putative candidate genes related to the traits considered in the present study. A hot spot region associated with correlated traits on chr 1, 4, and 8 and candidate genes associated with nitrate transporters, nitrogen content, and ammonium uptake on chromosomes 2, 4, 6, and 8 have been identified. The identified MQTL, putative candidate genes, and their orthologues were validated on our previous studies conducted on rice and wheat. The research-based interventions such as improving nitrogen use efficiency *via* identification of major genomic regions and candidate genes can be a plausible, simple, and low-cost solution to address the challenges of the crop improvement program.

Keywords: candidate gene, grain yield, meta-QTL, NUE, orthologue, rice, root

INTRODUCTION

Asian rice (*Oryza sativa*) is a major cereal crop grown worldwide and an essential food source for over half of the world's population (Nayar, 2014). The area under rice has continued to grow despite severe unsustainability issues due to national food security concerns and assured procurement on a minimum support price. Nitrogen is the essential macronutrient, and it is considered as the limiting factor for crop productivity. The nitrogen requirement of crops is highly dependent on the

exogenous supply of nitrogen (Kraiser et al., 2011). The world's agriculture is facing new challenges, and the global grain and food security problem persists (Kalugina, 2014). Globally, 119.41 million tons of nitrogen (N) fertilizers are being applied to the cereal crops to achieve the desirable crop yield (FAO, 2018). Rice needs nitrogen during whole stages of growth and development (Cassman et al., 1998; Ye et al., 2019). The rice crop has the lowest NUE among the major cereal crops (Norton et al., 2015). The increasing use of nitrogenous fertilizers for crop production coupled with poor nitrogen use efficiency (NUE) has led to the degradation of soil, water, and environment. The uptake, transport, assimilation/utilization, and remobilization of N are controlled by a complex and interconnected network of genes involved in various biological processes (Kant et al., 2011).

To date, a number of QTLs associated with nitrogen uptake, utilization, and nitrogen use efficiency have been detected in different cereal crop species (Agrama et al., 1999; Gallais and Hirel, 2004; Ribaut et al., 2007; Fontaine et al., 2009; Wei et al., 2012; Hu et al., 2015; Sandhu et al., 2015; Li et al., 2017; Zhou et al., 2017; Pozzo et al., 2018; Van Deynze et al., 2018; Zhang et al., 2019; Brasier et al., 2020). To the best of our knowledge, none of the identified genomic regions have been deployed in the marker-assisted introgression and pyramiding program. The QTL mapping approach generally influenced by the parents is used to develop the mapping population, choice of marker sets, population size and types, and the testing environments (Li et al., 2013; Lei et al., 2018; Zhao et al., 2018). Integrating QTL results from different independent experiments performed on the related cereal crop species provide useful information to study the genetic diversity of loci/alleles underlying the quantitative traits and highlight the potential targets to be further used in molecular breeding or the QTL cloning program.

The efficiency of any genomics-assisted breeding (GAB) program depends upon the consistency of the QTL effect across different genetic backgrounds and multiple environments (Collard and Mackill, 2008). The major effect QTLs which have been consistently reported in various studies in a common genomic region of the chromosome pinpoint its major role in regulating the particular trait to be further used efficiently in the GAB program.

Meta-QTL (MQTL) analysis is the approach to compile the information regarding consensus QTLs for a particular trait and to validate their effect across environments and backgrounds (Goffinet, 2000). Several studies have already been conducted to successfully locate the regions (meta-QTLs) in the genome of multiple crops for multiple traits (Li et al., 2013; Semagn et al., 2013; Acuña-Galindo et al., 2015; Van and McHale, 2017; Zhang et al., 2017; Khahani et al., 2021). The additional advantage of MQTL is the reduction of the confidence interval (CI) of the MQTLs controlling a particular trait of interest and identification of targeted candidate genes, thus improving the genetic resolution of the marker-assisted breeding program (Khahani et al., 2019). So, keeping in line the usefulness of meta-QTL analysis for narrowing down the major target genes, the present study is planned to aggregate all the reported QTLs for nitrogen use efficiency (NUE) and related traits and to fetch out relevant meta-QTL which can be further used in the breeding program with more precise knowledge about the genomic regions that can be targeted for NUE-associated traits in rice.

MATERIALS AND METHODS

The MQTL analysis for the nitrogen use efficiency (NUE) and related traits in rice was performed in the present study. It involves three main steps in identifying consensus QTLs associated with NUE and their related traits in rice:

1. An extensive bibliographic search of the QTL mapping studies and compilation of the reliable data on QTLs associated with NUE and related traits.
2. Creation of a consensus map on which QTLs of the individual studies were projected.
3. Meta-analysis on the QTL clusters to identify the consensus MQTL.

Bibliographic Search and Data Mining

An extensive bibliographic search on the QTLs associated with the traits controlling NUE and related traits in rice was conducted and retrieved from Google Scholar, PubMed (http://www.ncbi.nlm. nih. gov/pubmed), and Mendeley using appropriate keywords such as the agro-morphological trait, developmental stage, grain yield, nitrogen application, nitrogen use efficiency, nitrogen deficiency, nitrogen uptake, nitrogen utilization, nitrogen remobilization, QTL, root trait, and rice. The literature reports published in journal articles and dissertations from 2000 to 2021 were considered, and all information on QTLs pertaining to nitrogen use efficiency (NUE) and its related traits was compiled for carrying forward the genome-wide meta-QTL analysis. QTL data including details on the parents used to develop the mapping population, type of mapping population (F_2; recombinant inbred lines (RILs); backcross population (BC); near isogenic lines (NILs); chromosome segment substitution lines (CSSLs); double haploids (DHs)), size of population assayed, logarithm of the odds (LOD) score, phenotypic variance (R^2 or PVE), molecular markers flanking the QTL along with its genetic position (QTL position), and the genetic position of the QTL interval (confidence interval; CI) were compiled. The mid-point position between the two flanking markers was considered as the peak position wherever the information about the peak position was missing. The actual LOD score reported in the study was used, and in case, when the actual LOD score for an individual QTL was not available, a threshold LOD score of 3.0 was chosen for the study. The different QTLs located on the same chromosome were distinguished using the numerical identifiers following the chromosome. After careful examination, 28 relevant studies including 11-BC, 13-RILs, 3-DH, and 1-CSSL with all complete information from the 200 studies searched were compiled (**Table 1**). The studies that lacked the required data, such as the genetic position, LOD, and phenotypic variance, were excluded from the analysis. The size of the mapping populations ranged from 75 to 611 for BC lines,

TABLE 1 | Summary of the QTL studies used in the QTL meta-analysis for the nitrogen use efficiency and related traits in rice.

Cross	Cross type	Pop size	No of markers	Types of markers	Total no of QTL	Traits (QTL) identified per traits	Year of study	Country	References
Habataki × Koshihikari	BC_5F_3/F_4	611	23	SSR	2	Photosynthetic rate (PS) (2)	2011	Japan	Adachi et al. (2011)
XieqingzaoB × Zhonghui9308	BC_4F_6-CSSL	75	55	SSR	9	plant height (PH) (3), shoot dry weight (SDW)(2), root dry weight (RDW)(1), total dry biomass (TDB)(1), root length (RL)(1), root number (RN)(1)	2016–2017	China	Anis et al. (2018)
CSSL45× Zhonghui9308	BC_5F_2/F_3	75	10	SSR	1	root dry weight (RDW) (1)	2017	China	Anis et al. (2019)
XieqingzaoB × Zhonghui_9308	RIL	281	118	SSR	13	nitrogen content (NC)(4), harvest index (HI)(2), nitrogen use efficiency (NUE)(2), agricultural nitrogen-absorption efficiency (ANAE)(5)	2008–2009	China	Dai et al. (2015)
ZS97 × MH63	RIL	127	141	SSR, RFLP	30	grain yield (GY)(9), biomass yield (BY)(6), nitrogen content (NC)(8), nitrogen use efficiency (NUE)(7)	2006–2007	China	Wei et al. (2012)
ZS97 × MH63	RIL	127	115	SSR, RFLP	68	number of grains per panicle (NGP)(10), grain yield (GY)(14), panicle number per plant (PNP)(12), spikelet fertility (SF)(10), grain weight (GW)(22)	2006–2007	China	Wei et al. (2012)
IR64 × Azucena	DH	123	113	RFLP, RAPD	18	nitrogen content (NC)(6), tillering number (TN)(4), days to 50% flowering (DTF)(18)	2000	China	Fang and Wu, (2001)
R9308 × XieqingzaoB	RIL	238	82	SSR	7	shoot dry weight (SDW)(3), total dry biomass (TDB)(2), root length (RL)(1), plant height (PH)(1)	2008	China	Feng et al. (2010)
XQZB × R9308	RIL	138	127	SSR	21	plant height (PH)(11), days to 50% flowering (DTF)(10)	2009	China	Feng et al. (2011)
Lemont × Teqing	CSSL	247	118	SSR, RAPD	31	plant height (PH)(8), number of panicles per plant (PNP)(7), chlorophyll content (SPAD)(8), shoot dry weight (SDW)(5), grain yield (GY)(3)	2003	China	Han-Hua et al., 2006
WTR-1 × HAN	BC_1F_5	230	98	SNP	261	grain yield (GY)(50), nitrogen use efficiency (NUE)(4), biomass yield (BY)(40), partial factor productivity (PFP)(22), grain weight (GW)(61), spikelet fertility (SF)(84)	2014	Philippines	Jewel et al. (2019)
WTR1 × HAN_CH448_Z413	BC_1F_5	243	38	SNP	19	chlorophyll content (SPAD)(5), plant height (PH)(5), tiller number (TN)(9)	2017	Philippines	Mahender et al. (2019)
US-2 × Malay-2	BC_1F_4-RIL	168	83	SSR	8	root length (RL)(2), root number (RN)(3), root dry weight (RDW)(2), plant height (PH)(1)	2007	Philippines	Manangkil et al. (2019)
IR64 × Azucena	DH	84	135	RFLP, RAPD	16	days to 50% flowering (DTF)(1), number of grains/panicle (NGP)(2), nitrogen content (NC)(2), nitrogen use efficiency (NUE)(1), plant height (PH)(3), panicle length (PL)(1), tiller number (TN)(4), grain yield (GY)(1), spikelet fertility (SF)(1)	2008	Philippines	Senthilvel et al. (2008)
BPT5204 × PTB1	RIL	291	25	SSR	37	days to 50% flowering (DTF)(8), number of grains per panicle (NGP)(1), grain yield (GY)(2), leaf length (LL)(4), plant height (PH)(5), biomass yield (BY)(1), nitrogen content (NC)(8), chlorophyll content (SPAD)(2), total dry biomass (TDB)(1), grain weight (GW)(5)	2014–2015	India	Vishnukiran et al. (2020)

(Continued on following page)

TABLE 1 | (*Continued*) Summary of the QTL studies used in the QTL meta-analysis for the nitrogen use efficiency and related traits in rice.

Cross	Cross type	Pop size	No of markers	Types of markers	Total no of QTL	Traits (QTL) identified per traits	Year of study	Country	References
Lijiangxintu-anheigu × Towada	BC_4F_{10}/ F_{11}	105	94	SSR	47	soluble protein (SP)(17), free amino acid (FAA)(5), proline (PRO)(6), catalase (CAT)(7), peroxidase (POD)(5), nitrate reductase (NR)(7)	2017	China	Yang et al. (2019)
XieqingzaoB × Zhonghui9308	RIL	138	165	SSR	52	panicle length (PL)(10), spikelet fertility (SF)(3), number of panicles per plant (PNP)(20), grain yield (GY)(3), number of grains per panicle (NGP)(13), grain weight (GW)(3)	2009	China	Yue et al. (2015)
*indica*cultivar9311× *japonica* Nipponbare	BC_4F_2	119	190	SSR	44	plant height (PH)(9), root length (RL)(5), root dry weight (RDW)(2), shoot dry weight (SDW)(17), total dry biomass (TDB)(11)	2012	China	Zhao et al. (2014)
IR64 × INRC10192	RIL	140	60	ISSR	46	total dry biomass (TDB)(1), number of grains per panicle (NGP)(5), spikelet fertility (SF)(6), grain weight (GW)(1), grain yield (GY)(3), harvest index (HI)(9), number of tillers (NT)(5), plant height (PH)(15), panicle length (PL)(1)	2005	India	Akkareddy et al. (2010)
IR64 × Azucena	RIL	174	228	SSR	446	nitrogen content (NC)(63), agricultural nitrogen-absorption efficiency (ANAE)(50), nitrogen use efficiency (NUE)(56), biomass yield (BY)(1), grain weight (GW)(3), grain yield (GY)(5), leaf area (LA)(1), number of grains per panicle (NGP)(30), number of leaves (NL)(9), plant height (PH)(23), panicle number per plant (PNP)(26), photosynthetic rate (PS)(17), root dry weight (RDW)(19), shoot dry weight (SDW)(30), chlorophyll content (SPAD)(23), total dry biomass (TDB)(65), total fresh weight (TFW)(5), tiller number (TN)(20)	2011	Belgium	Nguyen et al. (2014)
Dasanbyeo × TR22183	RIL	166	6	SSR, STS	5	nitrogen content (NC)(1), harvest index (HI)(1), grain yield (GY)(1), biomass yield (BY)(2)	2002	Korea	Cho et al. (2007)
ZYQ8XJ × 17	DH	127	233	SSR	28	plant height (PH)(15), tiller number (TN)(13)	2003	China	Jiang et al. (2008)
Nipponbare × kasalath	BC_1F_6	98	45	RFLP	12	nitrogen content (NC)(6), NADH-glutamate synthetase content (NGOC)(6)	2001	Japan	Obara et al. (2014)
YTH183 × IR64	BC_3F_8	334	17	SSR	5	root length (RL)(5)	2010–2011	Japan	Obara et al. (2014)
NPT × IR72	RIL	101	170	SSR, RFLP	61	days to 50% flowering (DTF)(7), grain weight (GW)(3), spikelet fertility (SF)(3), nitrogen content (NC)(20), nitrogen use efficiency (NUE)(3), plant height (PH)(3), panicle number per plant (PNP)(4), specific leaf area (SLA)(6), chlorophyll content (SPAD)(3), total dry biomass (TDB)(4)	2001–2004	Japan	Laza et al. (2006)
HHZ × Teqing, CDR22, OM1723	BC_1F_4	206	4	KASP SNP	2	grain yield (GY)(1), spikelet fertility (SF)(1)	2014	China	Feng et al. (2018)
Azucena × Bala	RIL	205	65	SSR	17	root length (RL)(2), maximum root thickness (MRT)(2), root dry weight	2009	United Kingdom	MacMillan et al. (2006)

(Continued on following page)

TABLE 1 | (*Continued*) Summary of the QTL studies used in the QTL meta-analysis for the nitrogen use efficiency and related traits in rice.

Cross	Cross type	Pop size	No of markers	Types of markers	Total no of QTL	Traits (QTL) identified per traits	Year of study	Country	References
Zhenshan 97 × Minghui 63	RIL	127	108	SSR	24	(RDW)(6), total dry biomass (TDB)(4), plant height (PH)(3) agricultural nitrogen-absorption efficiency (ANAE)(10), grain yield (GY)(14)	2006–2007	China	Wei et al. (2011)

BC: backcross, RIL: recombinant inbred lines, DH: double haploids, CSSL: chromosome segment substitution lines, RAPD: random amplified polymorphic DNA, RFLP: restriction fragment length polymorphism, SSR: simple sequence repeats, ISSR: inter-simple sequence repeats, SNP: single nucleotide polymorphism, KASP: kompetitive allele specific PCR, STS: sequence-tagged sites. The numeric number in the bracket () represents the number of QTL, for that particular trait.

TABLE 2 | The trait category and number of QTL compiled per category in the QTL meta-analysis for the nitrogen use efficiency and related traits in rice.

Trait category	Number of QTL
Agricultural nitrogen-absorption efficiency	65
Nitrogen content	118
Nitrogen use efficiency	73
Nitrogen-related enzymes and amino acids (proline, peroxidases, nitrate reductase, free amino acid, soluble proteins, catalase, NADH-glutamate synthetase)	53
Photosynthetic rate and chlorophyll content	60
Plant height	105
Shoot dry weight	57
Total dry biomass	89
Leaf number and area	20
Root traits (root length, root number, root thickness, root dry weight)	53
Grain yield and related traits (harvest index, biomass yield, number of grains per panicle, panicle length, panicle number per plant, grain weight, days to 50% flowering, number of productive tillers, spikelet fertility, partial factor productivity)	637
Total	1,330

101 to 291 for RILs, 113 to 233 for DH, and 247 for CSSL lines. The information on the 33 traits associated with NUE and its related traits was compiled for carrying forward the genome-wide meta-QTL analysis. The 33 traits were accommodated into 11 major trait categories (**Table 2**). The detailed information on 1,330 QTLs associated with 11 major trait categories including the agricultural nitrogen-absorption efficiency (ANAE)((65 QTLs)); nitrogen content (NC)(118 QTLs); nitrogen use efficiency (NUE)(73 QTLs); photosynthetic rate and chlorophyll content (SPAD) (60 QTLs); nitrogen-related enzymes and amino acids (NEAA) [proline (PRO), free amino acid (FAA), soluble proteins (SP), peroxidases (POD), nitrate reductase (NR), catalase (CAT), glutamine synthetase (GSI), NADH-glutamate synthetase (NGOC)] (53 QTLs); root traits (RT) [root length (RL), root number (RN), root thickness (RT), root dry weight (RDW)] (53 QTLs); plant growth and morphological traits such as plant height (PH) (105 QTLs); shoot dry weight (SDW) (57 QTLs); total dry biomass (TDB) (89 QTLs); leaf number and area (LN and A) (20 QTLs); and grain yield (GY) and yield-related traits [biomass yield (BY), grain weight (GW), harvest index (HI), days to 50% flowering (DTF), tiller number (TN), panicle length (PL), panicle number per plant (PNP), number of grains per panicle (NGP), spikelet fertility (SF), partial factor productivity (PFP)](630 QTLs) were compiled. The compiled information on the 1,330 QTLs associated with the 11 trait categories was subjected to MQTL analysis. The QTLs have been renumbered based on their location on each chromosome.

QTL Projection and Development of the Consensus Map

To construct a consensus map, the LPmerge tool in the R package was employed (Endelman and Plomion, 2014) where the maps with the markers and QTLs were iteratively projected on a composite reference map (integrated rice genetic linkage maps). The reference maps used were those from Orjuela et al. (2010), Supplementary table 18 from GRAMENE (https://archive.gramene.org (International Rice Genome Sequencing Project, 2005), Cornell SSR 2001, and IRMI-2003 (https://archive.gramene.org). It is based on the common marker between the original map and reference map by means of a homothetic function described by Chardon et al. (2004). Any marker that did not comply (inverted) in terms of linkage was automatically discarded. After the integration of all the maps, the consensus map contained 21,280 markers, including SSR, RFLP, AFLP, SNP markers, and genes. The total number of markers discarded was 560 markers. The consensus map covered a total length of 1821 cM, with an average distance of 3.5 cM between markers. Only 1,330 QTLs from the 2,763 available QTLs

associated with the 11 trait categories with complete information (estimated CIs, original LOD scores, peak positions, and phenotypic variance) required for the QTL projection were used for projection on the consensus map using BioMercator V4.2 (Sosnowski et al., 2012) (**Table 2**).

Meta-QTL Analysis

Following projection, the meta-QTL analysis was performed, for each chromosome individually following the Veyrieras et al. (2007) two-step algorithm available in the software BioMercator V4.2. The Akaike (AIC) statistics values were used to find the best QTL model for determining the actual number of MQTLs on each of the rice chromosome. The detailed statistical procedures and the algorithms used in this software have been well-described in the study by Sosnowski et al. (2012). The first step involved the use of five different models based on the presence of 1, 2, 3, 4, or N real QTL. The best model was selected using Akaike (AIC) statistics. The second step involved setting up the suitable parameters for further analysis. The parameters used were the actual number of MQTLs or the real QTL to be mapped on the concerned chromosome. The phenotypic variance and LOD values of the MQTLs were calculated as the averages of the phenotypic variance and LOD values of the QTLs involved.

Functional Analysis and Identification of Nitrogen Use Efficiency–Associated Candidate Genes

The selected 42 meta-QTL regions possessing QTLs with average phenotypic variance >8%, average LOD >4, and the involvement of ≥10 initial QTLs within the MQTL were subjected to the functional analysis to identify candidate genes related to the traits associated with nitrogen use efficiency in rice. First, genetic markers flanking the confidence intervals of each rice MQTL (rMQTL) were selected, and their physical positions on respective chromosomes were obtained from the Gramene database (www.gramene.org), or gene models present within the original or estimated physical regions were retrieved using the "BioMart" of the Ensembl Plants database. All the genes physically located within or near each rice MQTL region were considered as candidate genes and retrieved from the Rice Annotation Project Database (RAPDB) (http://rapdb.dna.affrc.go.jp) as batch download, or the primer sequence of the marker is subjected to nucleotide blast in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn) to identify the range of the sequence in the reference genome of Nipponbare using RAP and Build 5 (www.rapdatabase.org). It is assumed that the genes identified in Nipponbare regions are homologous and collinear to those underlying the nitrogen use efficiency and associated QTLs mapped in different studies involving different donor wild species and recipients. The functional annotations of the identified gene models were explored for the best candidate genes within each MQTL. All genes with the gene ontology (GO) term or description related to NUE, nitrogen uptake, its assimilation, and all related traits were filtered and considered as candidate genes (CGs). The QTLs associated with nitrogen use efficiency and related traits identified previously by Sandhu et al. (2015), Sandhu et al. (2016), Sandhu et al. (2019), Subedi et al. (2019), and Sandhu et al. (2021) were projected on the MQTLs identified in the present study to find consistency in the genomic regions.

Rice MQTL Regions Homologous to the Other Cereals

Additionally, to evaluate transferability of information to other cereals, ortho-MQTLs were investigated based on the genomic collinearity between rice–maize, rice–wheat, and rice–barley. The information on maize, wheat, and barley genes associated with NUE and related traits was collected from the available literature and used for the retrieval of corresponding protein sequences for the identification of homologous MQTL genomic regions. The amino acid sequences for the relevant genes were then retrieved from NCBI https://www.ncbi.nlm.nih.gov/) and used for BLASTP search to identify the rice protein (available in Ensembl Plants) at an E-value of $<10^{-10}$, with 60% coverage, and >60% identity. The physical positions of the corresponding genes and rice MQTLs were then compared to detect the MQTL regions homologous to known genes in other cereals.

To detect the ortho-MQTLs between rice and maize, rice and wheat, and rice and barley syntenic regions between the two respective species were identified by using the Ensembl Plants database.

The distribution of the aforementioned factors, number of MQTL related to nitrogen use efficiency and all related traits, and candidate genes underlying MQTL over the rice genome were shown by using Circos (Krzywinski et al., 2009).

Development of the Validation Panel to Check the Efficacy of the Identified MQTL and Candidate Genes.

The information on the marker trait associated with root traits improving nitrogen use efficiency and grain yield/yield-related traits in rice (Sandhu et al., 2015; Sandhu et al., 2019; Subedi et al., 2019) and wheat (Kumar et al., 2021; Sandhu et al., 2021) was collected from our previous studies. The information on the nitrate transporter gene in wheat and rice was collected from the Ensembl Plants database. To check the nitrogen responsiveness of candidate genes, the expression data of all the candidate genes were downloaded from the RiceXPro database. Specifically, root gene expression profiles in response to nitrogen were considered. The dataset RXP_5002, which contained microarray-based expression data of 7-day-old seedlings exposed to nitrogen deficiency treatments and control conditions, was analyzed.

RESULTS

Salient Features of the QTL Studies

In order to identify the consensus genomic regions associated with the 33 nitrogen use efficiency and related traits, we compiled

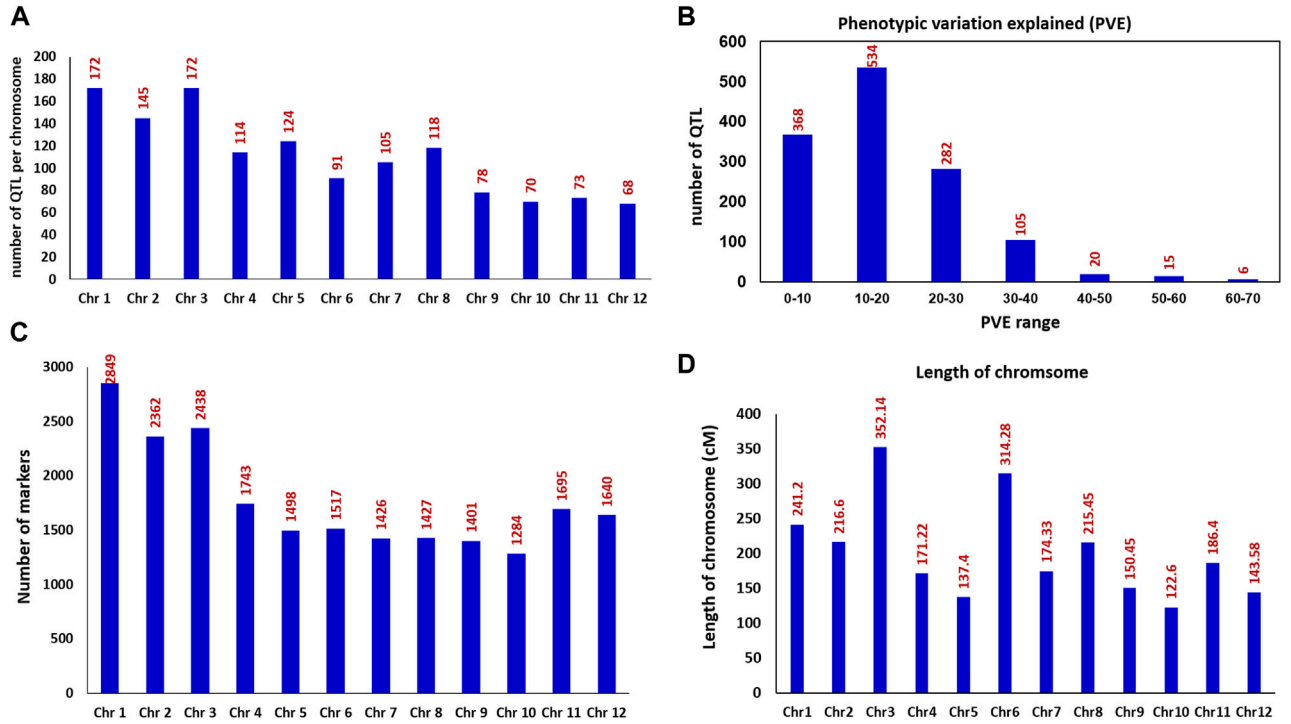


FIGURE 1 | Frequencies of QTL and details on the mapped markers used for the identification of MQTLs in rice **(A)** Number of QTLs present on the 12 rice chromosomes **(B)** Frequencies of the QTL with different levels of phenotypic variation explained (PVE%) **(C)** Frequencies of the mapped markers on each of the 12 rice chromosomes **(D)** Length/size of the rice genetic map in cM (centimorgan) across 12 chromosomes.

the information on a total of 1,330 QTLs derived from 28 relevant studies including 11-BC, 13-RILs, 3-DH, and 1-CSSL with the population size that ranged from 75 to 611 reported between 2001 and 2021. All the 33 traits were grouped into 11 main trait categories (**Table 2**). The 1,330 QTLs were distributed across all 12 chromosomes (chr) of rice (**Figure 1A**). The chr 1 and 3 harbored the maximum number of QTL (172 QTL) for all the studied traits followed by chr 2 (145 QTL) and chr 5 (126 QTL). The chr 12 harbored the lowest number of QTL (68 QTL). Among the studied trait category, GY and related trait categories had the highest number of QTL with 630 QTLs followed by 118 QTLs for nitrogen content and 105 QTLs for plant height. The phenotypic variance of the individual QTLs ranged from 1.9 to 66.9%, with an average of 17.43%. The frequency distribution of the QTL with different levels of phenotypic variation explained is presented in **Figure 1B**. The LOD value of the individual QTL ranged from 2.02 to 28.7 with an average of 4.77.

Rice Consensus Map

The rice consensus genetic map included a total of 22,280 markers (**Figure 1C**) covering a total length of 2425.65 cM, thus giving a density of 8.77 markers/cM for the whole rice genome. The size of the rice genetic map for an individual chr ranged from 122.60 cM (chr 10) to 352.14 cM (chr 3) (**Figure 1D**), and the number of markers on an individual chr ranged from 1,284 on chr 10 to 2849 on chr 1. The marker density on individual chr varied from 4.83 markers/cM for chr 6 to 11.81 markers/cM for chr 1.

MQTLs Detected for Grain Yield/Related Traits and Nitrogen Content/Nitrogen Use Efficiency

Only 915 QTLs from the 1,330 available QTLs could be used for successful projection onto the consensus map; the remaining 415 QTLs could not be projected as either the associated markers were absent in the consensus map, or they had comparatively low PVE values and/or a large confidence interval. The MQTL analysis significantly summarized the total number of projected QTLs from 915 to 76 MQTLs (8.3%) (**Table 3**). The number of MQTLs per rice chromosome ranged from three (chr 1) to nine (chr 3) with an average of 6.33 MQTLs per chromosome. The 76 MQTLs involving 636 QTLs were identified using the MQTL analysis. The remaining 279 QTLs could not be assigned to any of the identified MQTL since they either lacked the common markers between the consensus map and initial maps or the QTL had a relatively low phenotypic variance or LOD value and/or a large confidence interval.

The phenotypic variance of the detected MQTL ranged from 8.14 to 21.16, and the LOD ranged from 3.29 to 9.47. The confidence interval of the MQTLs, on an average, was reduced 5.24-fold relative to the confidence intervals of the QTLs used in the present study. The average confidence interval of the QTL

TABLE 3 | Summary of the detected MQTLs for the nitrogen use efficiency and related traits in rice.

MQTL ID	Chr	Position (cM)	CI (95%)	From	To	Flanking marker from	Flanking marker to	Start (Mb)	End (Mb)	No. of QTLs involved in MQTL	QTLs involved
MQTL1.1	1	54.16	1.16	53.16	54.32	RM10009	RM10111	188501	2104274	47	SF, BY, GW, PNP, GD, GY, POD, PNP, RDW, TN, CAT, DTF, NC, PL, NUE, SP, SPAD, NGP, PFP, PH, HI
MQTL1.2	1	79.92	1.54	78.92	80.46	RM10523, SNP_1_20238919	RM10527, SNP_1_20706894	20238919	20706894	10	PH, GW, PNP, NGOC, DTF, NC, BY, SF, GY
MQTL1.3	1	110.58	0.12	109.58	109.7	RM11226	RM11235	22330477	22473962	26	BY, SPD, GY, PH, SDW, DTF, TDB, ANUE, NC, RT, PS, RL, PNP, SF, TN
MQTL2.1	2	16.71	2.96	15.23	18.19	SNP_2_4342883	SNP_2_4481943	4342883	4481943	16	GY, HI, BY, GW, PFP, SF
MQTL2.2	2	28.81	1.47	28.08	29.55	RM12349	RM12353	977799	1078287	2	NUE, PNP
MQTL2.3	2	39.76	2.14	38.69	40.83	RM12446	RM12461	2331733	2675901	13	PH, SF, NC, PNP, GW, ANAE, PL, GY, ANUE
MQTL2.4	2	51.72	1.58	50.93	52.51	RM12658	RM12693	5366392	6121755	11	DTF, SF, ANUE, TDB, GY, PNP, TN, PH
MQTL2.5	2	68.63	0.28	68.49	68.77	RM12913	RM12914	9243319	9409837	2	NC, DTF
MQTL2.6	2	69.67	1.09	69.13	70.22	RM12914	RM12918	9409837	9459499	5	DTF, NC, NGP
MQTL2.7	2	87.85	4.47	85.62	90.09	S2-136	G1314A	1291679	25433673	6	GY, SF, NC, GW
MQTL2.8	2	119.49	0.01	119.49	119.50	C747	G57	27078652	28307630	5	PNP, NC, DTF, PH
MQTL3.1	3	3.67	2.46	2.44	4.90	SNP_3_1270943	SNP_3_1670761	1270943	1670761	19	GW, SF, GY, BY, PFP, PH
MQTL3.2	3	14.49	4.68	12.15	16.83	SNP_3_3542519	RM14441	3542519	3597356	5	GY, SF, PFP, HI
MQTL3.3	3	49.05	1.44	48.33	49.77	RM4992	RM14526	4706112	5024749	12	RDW, SDW, PS, NUE, ANAE, SF, NT, TDB
MQTL3.4	3	65.42	9.43	60.71	70.14	SNP_3_16294363	RM15181	16294363	16310268	3	GY, DTF, GW
MQTL3.5	3	77.1	3.01	75.60	78.61	C6	RM545	2985072	4916484	11	PNP, NUE, GY, FAA, TN, GW, NC, RL
MQTL3.6	3	85.08	2.49	83.84	86.33	RM14782	RM14820	10140286	10790118	2	SDW, PH
MQTL3.7	3	95.69	0.49	95.45	95.94	RM14936	RM14945	12917488	12998073	10	DTF, SF, GY, PRO, PNP, POD
MQTL3.8	3	117.9	1.61	117.10	118.71	RM15288	RM15313	18945114	20112723	9	TN, PNP, NC, DTF, GY, SF
MQTL3.9	3	138.2	0.45	137.98	138.43	RM15626	RM15639	25803162	26022473	5	PHT, PL, SF, NUE, NC
MQTL4.1	4	4.37	1.18	3.78	4.96	RM16393	RM16399	3393833	3498871	6	ANAE, NC, DTF, SF, PS
MQTL4.2	4	32.74	3.17	31.16	34.33	RM16767	RM16788	17456500	18062531	4	NUE, TN, SF, NC
MQTL4.3	4	46.99	3.16	45.41	48.57	SNP_4_14609247	RM16926	14609247	18062531	10	SF, GW, PHT, TN, RDW, PNP, NC
MQTL4.4	4	59.1	2.39	57.91	60.30	RM16996	RM17007	21359732	21484908	21	NC, BY, GW, SF, PFP, GY, TDB, RDW, SDW, NUE, PNP, TN
MQTL4.5	4	72.97	0.98	72.48	73.46	SNP_4_21815986	SNP_4_21833014, RM17090	21815986	21833014	6	NGOC, PH, NGP, RL, TDB, GY
MQTL4.6	4	87.7	2.75	86.33	89.08	RM17231	RM17272	25746072	26405863	3	NC, GY, PNP
MQTL4.7	4	104.9	1.13	104.34	105.47	RM17475	RM17489	31082124	31418566	3	DTF, PH, SF
MQTL5.1	5	6.73	0.78	6.34	7.12	RM17846	RM17852	1777550	1942400	2	SF, DTF

(Continued on following page)

TABLE 3 | (*Continued*) Summary of the detected MQTLs for the nitrogen use efficiency and related traits in rice.

MQTL ID	Chr	Position (cM)	CI (95%)	From	To	Flanking marker from	Flanking marker to	Start (Mb)	End (Mb)	No. of QTLs involved in MQTL	QTLs involved
MQTL5.2	5	18.02	2.52	16.76	19.28	GA478	R2232	2011188	4107103	20	PHT, PNP, SF, GY, BY, GW, NUE, SDW, PFP, NC
MQTL5.3	5	35.88	6.38	32.69	39.07	RM18408	SNP_5_15469279	15116392	15469279	2	TN, GY
MQTL5.4	5	44.56	3.03	43.05	46.08	RM18033	RM18071	5273339	6155982	9	PS, RL, PFP, GY, BY
MQTL5.5	5	51.56	3.09	50.02	53.11	RM18115	RM18176	7121001	8674028	5	RN, SF, SPD, NC, TDB
MQTL5.6	5	60.33	2.79	58.94	61.73	RM18302	RM18343	12708022	13572971	7	TDB, SF, RDW, ANAE, SDW
MQTL5.7	5	70.03	0.6	69.73	70.33	RM18624	RM18632	19183516	19346291	10	PH, RT, RL, TN, PNP, RDW, TN
MQTL5.8	5	110.22	0	110.22	110.22	RM178	RM6972	25101829	25208346	4	SF, RL
MQTL6.1	6	7.27	7.24	3.65	10.89	RZ242	G342	28963386	30822714	2	SF
MQTL6.2	6	44.18	1.6	43.38	44.98	SNP_6_12183428	SNP_6_13250266	12183428	13250266	21	TN, PFP, NUE, GW, GY, SF, BY, NC, RDW
MQTL6.3	6	62.07	1.71	61.22	62.93	R2549	RZ516	24916395	2560318	9	PL, DTF, TDB, BY, RDW, NUE, PNP, SDW, ANAE
MQTL6.4	6	93	2.52	91.74	94.26	RM19715	RM19746	7868951	8518493	12	PH, CAT, PRO, NC, SP, FAA, PNP, SF
MQTL6.5	6	107.6	1.9	106.65	108.55	SNP_6_29056693	SNP_6_29416997	29056693	29416997	14	GY, SF, TDB, TN, RN, BY, PNP, RL
MQTL6.6	6	159.12	1.43	158.41	159.84	Pho2	G329	27384548	27612443	6	NC, RL, DTF, SDW, GY
MQTL7.1	7	13.77	5.57	10.99	16.56	SNP_7_4569035	SNP_7_5704192	4569035	5704192	4	SF, BY
MQTL7.2	7	38.56	0.94	38.09	39.03	RM21034	RM21050	3577411	3771121	17	PNP, SF, PH, PL, BY, DTF, RL, POD, HI, SP, GW
MQTL7.3	7	48.82	2.24	47.70	49.94	RM21660	RM22027	18996018	26301017	13	NC, SP, GY, PRO, TDB, PS, TN, SDW, BY, DTF
MQTL7.4	7	80.47	2.6	79.17	81.77	RM22132	RM22157	28665611	29170514	12	GY, PNP, NUE
MQTL7.5	7	107.07	0.76	106.69	107.45	SNP_7_28234334	SNP_7_28303039	28234334	28303039	11	GW, SF, PFP, BY, GY
MQTL8.1	8	1.48	4.28	0.66	3.62	SNP_8_389278	G278	389278	1193267	5	PFP, GY, BY
MQTL8.2	8	18.73	2.35	17.56	19.91	RM6863	RM22416	2005990	3285143	9	BY, GW, NGP, HI, SF
MQTL8.3	8	31.49	3	29.99	32.99	SNP_8_8437588	SNP_8_8580913, RM122191	8437588	8580913	10	NC, BY, GW, SDW, PFP, GY
MQTL8.4	8	48.35	1.21	47.75	48.96	RM22335	RM22351	2144719	2470019	8	GY, DTF, PNP, SF, PL, PS
MQTL8.5	8	58.94	5.75	56.07	61.82	RM22981	S8_8206216, RM339	2470019	8206216, 17945059	2	SDW, PH
MQTL8.6	8	88.55	0.24	88.43	88.67	RM22979	RM22982	17250943	17401871	26	NC, TDB, GW, GY, SF, PH, NUE, RDW, ANAE, SPAD, TN, PNP
MQTL9.1	9	24.5	4.38	22.31	26.69	SNP_9_12154616	C397B	12154616	12289001	5	PH, GY, SF, BY
MQTL9.2	9	37.47	4.04	35.45	39.49	R1164	RZ698	6015994	7222547	10	GY, SF, BY, PFP, BY, GW
MQTL9.3	9	51.3	0.86	50.87	51.73	RM23654	RM23655	99249	159416	3	TDB, PNP, PS
MQTL9.4	9	56.73	3.49	54.99	58.48	RM23820	RM23888	5036985	6546861	12	ANAE, NC, TDB, GY, NUE, RDW, SF, GW
MQTL9.5	9	68.16	3.32	66.50	69.82	RM23967	RM23999	8127499	8970550	12	

(Continued on following page)

TABLE 3 | (*Continued*) Summary of the detected MQTLs for the nitrogen use efficiency and related traits in rice.

MQTL ID	Chr	Position (cM)	CI (95%)	From	To	Flanking marker from	Flanking marker to	Start (Mb)	End (Mb)	No. of QTLs involved in MQTL	QTLs involved
MQTL9.6	9	86.82	0.49	86.58	87.07	RM24130	RM6839	11610977	14512398	6	SDW, TDW, GW, NC, GY, RDW, PNP, SPAD, TN GW, TN, PN, RDW, NC
MQTL10.1	10	10.33	4.71	7.98	12.69	SNP_10_2056123	RM24990	2056123	2768779	4	PH, ANAE, NC, SF
MQTL10.2	10	27.76	4.42	25.55	29.97	RM25084	RM25178	4818641	8255065	8	FAA, NC, NUE, TN, SDW, TDB
MQTL10.3	10	41.89	6.53	38.63	45.16	RM25271	S10_14563405	10748123	14563405	2	NC, PS
MQTL10.4	10	50.86	3.79	48.97	52.76	RM25308	RM25331	11746315	12461164	2	PNP, GY
MQTL10.5	10	56.87	3.37	55.19	58.56	RM467	RM25401	13044511	13797426	6	FAA, SP, PNP, TN
MQTL10.6	10	74.14	4.94	71.67	76.61	RM25601	SNP10_18820606	17570667	18820606	5	NC, GY, BY, TDB
MQTL10.7	10	88.16	4.24	86.04	90.28	RM25852	RM25934	21605343	22626576	2	PNP, NC
MQTL11.1	11	23.05	5.99	20.06	26.05	RM26044	RM26108	1885593	2828718	4	HI, NC, SDW, NC
MQTL11.2	11	44.9	2.33	43.74	46.07	RM26306	RM26341	7073686	7650323	9	TDB, TN, NC, PNP, TDB, ANAE, SF
MQTL11.3	11	63.45	3.05	61.93	64.98	RM26687	RM26727	15903868	16610716	4	GY, NUE, PH
MQTL11.4	11	99.77	3.13	98.21	101.34	RM27045	RM27097	22663165	23487875	5	NR, RDW, ANUE, GY
MQTL12.1	12	14.75	5.57	11.97	17.54	RM27494	S12_5905028	1709959	5905028	3	TN, PNP, PH
MQTL12.2	12	49.07	6	46.07	52.07	RM28004	RM28064	1,3161862	14701301	2	PNP, HI
MQTL12.3	12	57.06	2.36	55.88	58.24	RM28095	RM28117	15679490	16360229	3	NUE, PNP
MQTL12.4	12	66.78	2.39	65.59	67.98	RM27712	RM27800	5104402	7237077	8	NC, PS, TDB, ANAE, SPAD, SDW, TDW, RDW
MQTL12.5	12	74.49	4.54	72.22	76.76	RM27855	RM28004	8458858	1,3161862	2	NC, DTF
MQTL12.6	12	86.42	5.01	83.92	88.93	RM28093	RM28455	15616573	22776128	2	PH
MQTL12.7	12	116.74	0.77	116.36	117.13	RM28511	RM28523	23594375	23753217	5	PNP, GY, TN, NGP, NC

ANAE: agricultural nitrogen-absorption efficiency, BY: biomass yield, CAT: catalase, DTF: days to 50% flowering, FAA: free amino acid, GY: grain yield, GW: grain weight, GD: grain density, HI: harvest index, NGP: number of grains per panicle, NGOC: NADH-glutamate synthetase content, NC: nitrogen content, NUE: nitrogen use efficiency, NR: nitrate reductase, PL: panicle length, PH: plant height, RDW, root dry weight; RL, root length, NR: number of roots, RT, root thickness, SPAD: chlorophyll content, POD: peroxidase, PNP: panicle number per plant, PFP: partial factor productivity, PS: photosynthetic rate, SF: spikelet fertility, SDW: shoot dry weight, SP: soluble protein, TN: tiller number, TDB: total dry biomass, Chr: chromosome, Mb: megabase, cM: centimorgan, CI: confidence interval.

used in the present study ranged from 8.72 to 20.20 cM with an average confidence interval of 12.26 cM. The frequencies of QTL with different sizes of confidence interval (in cM) are presented in **Figure 2A**. The average confidence interval of the MQTLs detected in the present study varied from 0.94 to 4.57 cM with an average confidence interval of 2.73. The average reduction in the size of the confidence interval for individual MQTLs was as high as 13.3-fold for chromosome 1 and 8.95-fold for chromosome 2, and as low as 2.9-fold for chromosome 10 and 2.55-fold for chromosome 12 (**Figure 2B**).

Each of the individual MQTL differed for the number of traits controlling the MQTL (**Table 3**). The association of MQTL with traits varied from a single trait for MQTL6.1 and MQTL12.6 to 21 traits for MQTL1.1. The 20 MQTLs out of 76 detected MQTLs were reported to be associated with more than 10 traits (**Table 3**). The 55 MQTLs were reported to be associated with nitrogen use efficiency–related traits, root traits, and grain yield/yield-related traits (**Table 3**). The major colocation of genomic regions associated with root traits improving the nitrogen use efficiency and grain yield/related traits was observed on chr 2, 3, 4, 5, 6, and 10 (**Table 3**). The MQTL analysis for the 630 QTLs associated with the grain yield and yield-related traits indicates the presence of MQTL for yield and related traits in the same genomic region (**Supplementary Table S1**). A total of 11 MQTLs (MQTLYRT1.3 on chr 1, MQTLYRT2.1 and MQTLYRT2.4 on chr 2, MQTLYRT3.4 and MQTLYRT3.5 on chr 3, MQTLYRT4.3 on chr 4, MQTLYRT5.2 on chr 5, MQTLYRT6.5 on chr 6, MQTLYRT7.2 on chr 7, and MQTLYRT8.2 and MQTLYRT8.4 on chr 8) were observed to be associated with more than 6 yield and yield-related traits (**Supplementary Table S1**). The 13 MQTLs (MQTLN&R1.1 on chr 1; MQTLN&R2.4 on chr 2; MQTLN&R3.1 and MQTLN&R3.2 on chr 3; MQTLN&R4.3, MQTLN&R4.4, and MQTLN&R4.5 on chr 4; MQTLN&R5.2 and MQTLN&R5.3 on chr 5; MQTLN&R6.1 and MQTLN&R6.9 on chr 6; MQTLN&R7.4 on chr 7; and MQTLN&R11.4 on chr 11) showed association with the

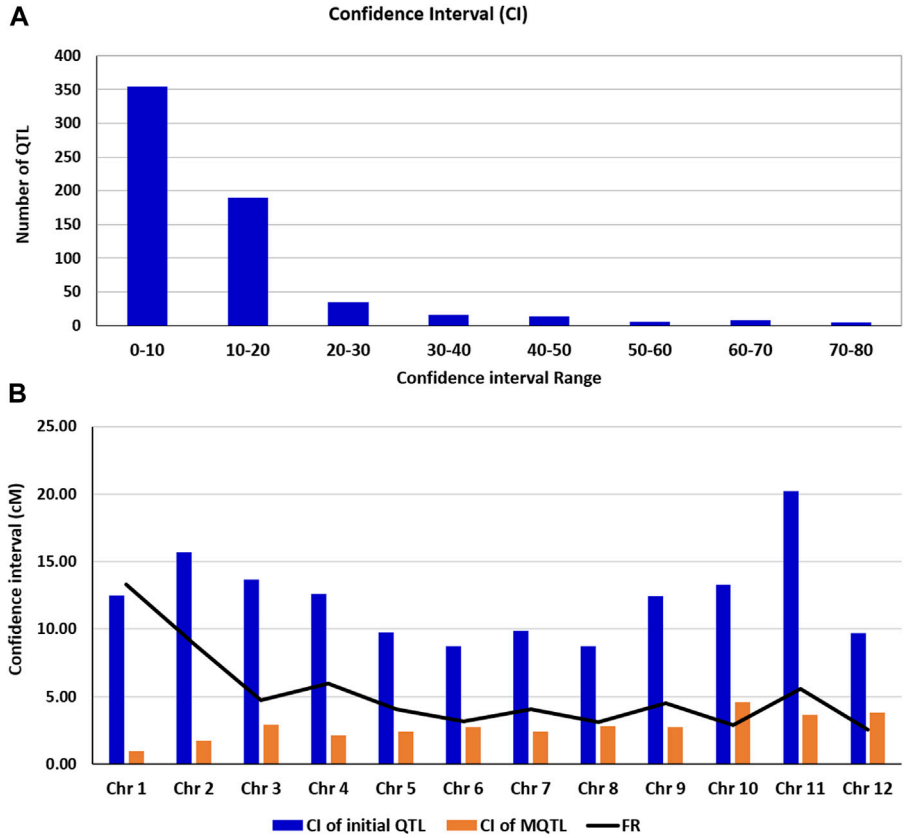


FIGURE 2 | (A) Frequencies of QTL with different sizes of confidence interval (in cM) **(B)** (b) Comparison of confidence intervals (CIs) of the QTL and those of meta-QTL (MQTL), showing the fold level of reduction (FR) in the size of CI.

root traits improving nitrogen use efficiency in rice (**Supplementary Table S2**). The 15 MQTLs were solely associated with nitrogen use efficiency in rice (**Supplementary Table S2**).

Identification of Candidate Genes and Orthologues

The 76 MQTLs reported in the present study were used for further selection of some of the promising MQTLs using the criteria; average phenotypic variance >8%, average LOD >4, and the involvement of ≥10 initial QTLs within the MQTL. This screening of MQTL resulted in the selection of 42 promising MQTLs (**Supplementary Table S3**), which was further used for the identification of candidate genes and orthologues in other crops such as wheat, barley, and maize.

A total of 2665 genes were present in the genomic region constituting the 42 promising MQTLs (**Supplementary Table S4**). A total of 158 candidate genes associated with plant growth and development, amino acid biosynthesis, nitrogen assimilation and transport, and stress resistance/tolerance were chosen (**Supplementary Table S5**) to identify the orthologues in barley, maize, and wheat. Out of the 158 candidate genes, 39 candidate genes showed no syntenic relationship with any of the three genomes i.e., barley, maize, and wheat. The 109, 104, and 94 of these 158 rice candidate genes could be utilized to identify the 376 wheat orthologues, 149 maize orthologues, and 109 barley orthologues, respectively, in the MQTL region (**Supplementary Table S6**).

Collinearity Within the Rice Genome and Synteny With Other Genomes

The investigation of collinear genomic regions within the rice genome resulted in the identification of duplicated regions containing the MQTLs associated with the same traits. The candidate genes underlying MQTL7.4 and MQTL9.6 were reported to be associated with the cellular response to nitrate, and MQTL1.2 and MQTL5.2, with nutrient reservoir activity (**Supplementary Table S6**). The candidate genes in the genomic region of MQTLs on chr 3, 5, 6, 7, 8, 10, 11, and 12 were observed to be associated with the transmembrane transport activity (**Supplementary Table S6**). The collinearity in the genomic region on chr 1,5; chr 2,3; chr 3,6; chr 3, 7; chr 4, 6; chr 5, 6; chr 5, 7; chr 6,9; and chr 7,9 was observed (**Supplementary Table S6**). The MQTLs associated with the nitrogen content, root traits, and grain yield such as MQTL1.1 and MQTL1.3 on chr 1; MQTL3.3 and MQTL3.5 on chr 3; MQTL4.3 and MQTL4.5

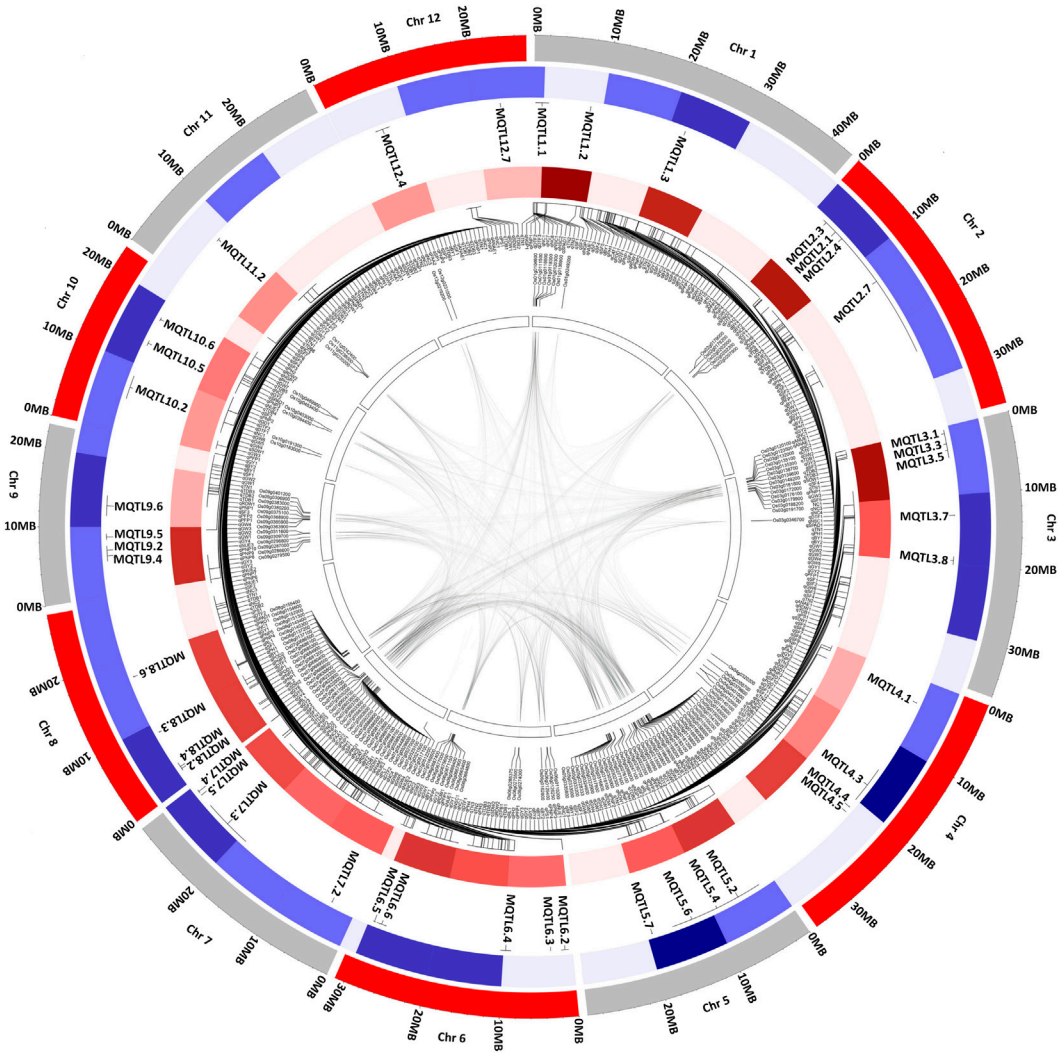


FIGURE 3 | Schematic representation of the distribution pattern of identified MQTL, QTL, and candidate genes on rice chromosomes. From the center of the plot moving to the outer circle (1) The innermost circle representing the collinearity within the rice genome (2) Candidate genes identified in the major MQTL region (3)(query) QTL density in the MQTL region (4) MQTLs associated with nitrogen use efficiency and related traits (5) Outermost circle represents the rice genome in MB. The color density indicates the number of MQTL detected and number of QTLs present in the MQTL region. The denser color indicates a greater number of MQTL or QTL within the MQTL region.

on chr 4; MQTL5.5 on chr 5; MQTL6.2 and MQTL6.3 on chr 6; MQTL7.2 on chr 7; MQTL8.6 on chr 8; MQTL9.4, MQTL9.5, and MQTL9.6 on chr 9; and MQTL11.4 on chr 11 were co-located in the rice genome duplicated regions (**Figure 3**). The already reported genes have been investigated in the MQTL region using the Q-TARO database. The MQTL1.1 comprises the *OsDET1* gene associated with photosynthetic capacity, *sui1* gene with plant height, and *AIP1* gene with root hair development; MQTL2.7 comprises the *GW2* gene associated with grain weight and size and *OsNAR2.1* gene associated with nitrogen uptake; MQTL3.5 comprises the *OsMDP1* gene associated with root elongation; MQTL4.6 comprises the *OsAMT1;1* gene associated with ammonium uptake; and MQTL12.6 comprises the *kch1* gene associated with coleoptile elongation (**Supplementary Table S7**).

The syntenic relationship was observed for rice candidate genes present in the MQTL region with the wheat, maize, and barley genome (**Supplementary Table S8**). The comparison of rice genome with wheat (**Figure 4A**), maize (**Figure 4B**), and barley (**Figure 4C**) genomes suggested that most of the orthologues were retained during evolution.

Checking the Efficacy of MQTL

To check the efficacy of identified MQTL and candidate genes underlying MQTL, a validation panel of the marker-trait association identified in our previous study (Sandhu et al., 2015; Sandhu et al., 2019; Subedi et al., 2019) in rice and wheat (Sandhu et al., 2021) and nitrate transporters reported in rice was made. The identification of previously identified marker-trait associations associated with the nitrogen use

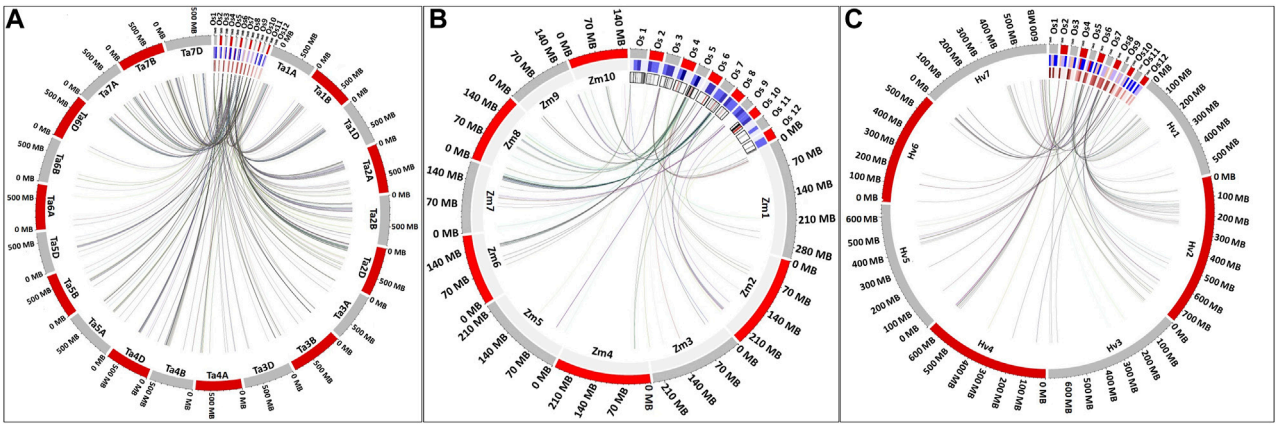


FIGURE 4 | Syntenic relationship of rice candidate genes present in the MQTL region with the **(A)** wheat, **(B)** maize, and **(C)** barley genome.

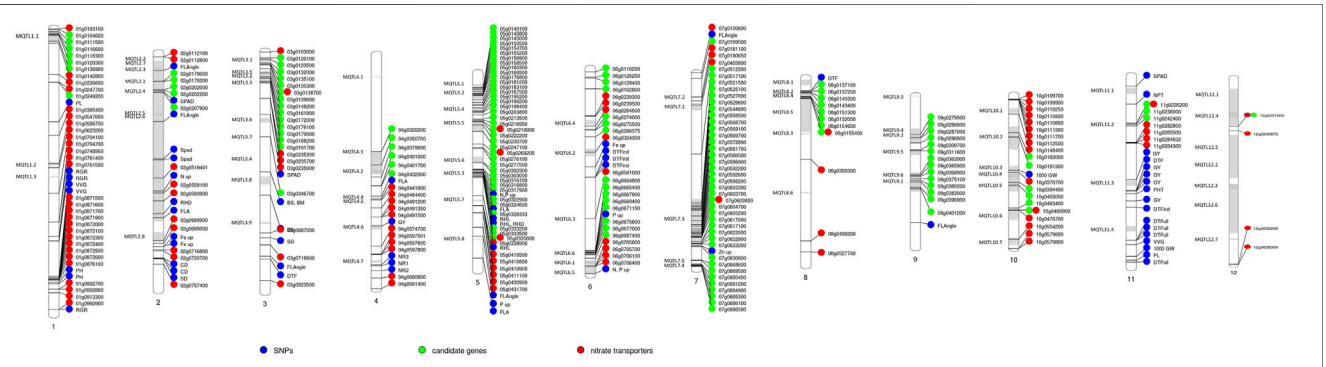


FIGURE 5 | Schematic representation of the distribution of previously reported single nucleotide polymorphism (SNP) associated with traits of interest, candidate genes reported in the present study, and the nitrate transporter along the 7 chromosomes of rice. The chromosome map showing MQTL identified in the present study and their co-location with the previously reported single nucleotide polymorphism (SNP) associated with different nitrogen use efficiency (NUE)–related trait, root traits, yield, and yield-related traits (Sandhu et al., 2015; Sandhu et al., 2019; Subedi et al., 2019), candidate genes reported in the present study, and the nitrate transporter reported in rice. The numbers below each rice chromosome indicate chromosome numbers.

efficiency–related traits in rice and nitrate transporter genes in close proximity to the MQTL and candidate genes reported in the present study indicated the robustness of the reported MQTL. The marker-trait association associated with the root traits (root hair length and root hair density) improving nutrient uptake (nitrogen and phosphorus) was reported to be collocated or in close proximity to the nitrogen transporter genes and candidate genes identified in the present study on chr 2, 5, and 6 (**Figure 5**). The marker-trait associations associated with grain yield/yield-related traits and plant morphological traits reported in our previous studies were present in the MQTL region reported in the present study. Similarly, the wheat orthologues were present in close proximity to the marker-trait association and nitrate transporter genes identified in our previous study (Kumar et al., 2021; Sandhu et al., 2021) (**Figure 6**). The identified 20 candidate genes in wheat are nitrogen transporters.

Expression Profiles of Candidate Genes

To identify the target genes for NUE, we analyzed the expression data of all 158 candidate genes identified in this study. We studied the response of candidate genes against nitrogen deficiency in the microarray-based expression dataset from the RiceXPro database. Of all the 158 candidate genes, 15 genes were selected which showed significant changes in expression values in response to nitrogen deficiencies in roots (**Table 4**). Twelve out of 15 genes showed upregulation in response to nitrogen deficiency, while three genes were downregulated (**Figure 7**). Eight genes out of twelve were upregulated in both 6 and 24 h after treatment in comparison to control. These genes consisted of glutamine amidotransferase, *OsVIT2*, *PLA1*, *GL7*, ERF transcription factor (*Sub1C*), *CEF1*, HD zip TF (*Oshox14*), and TGF beta receptor (**Table 4**). Three genes (*OsWRKY67*, acid phosphatase, and *OsABCG12*) showed a similar expression to control at 6 h after treatment, but their expression increased at

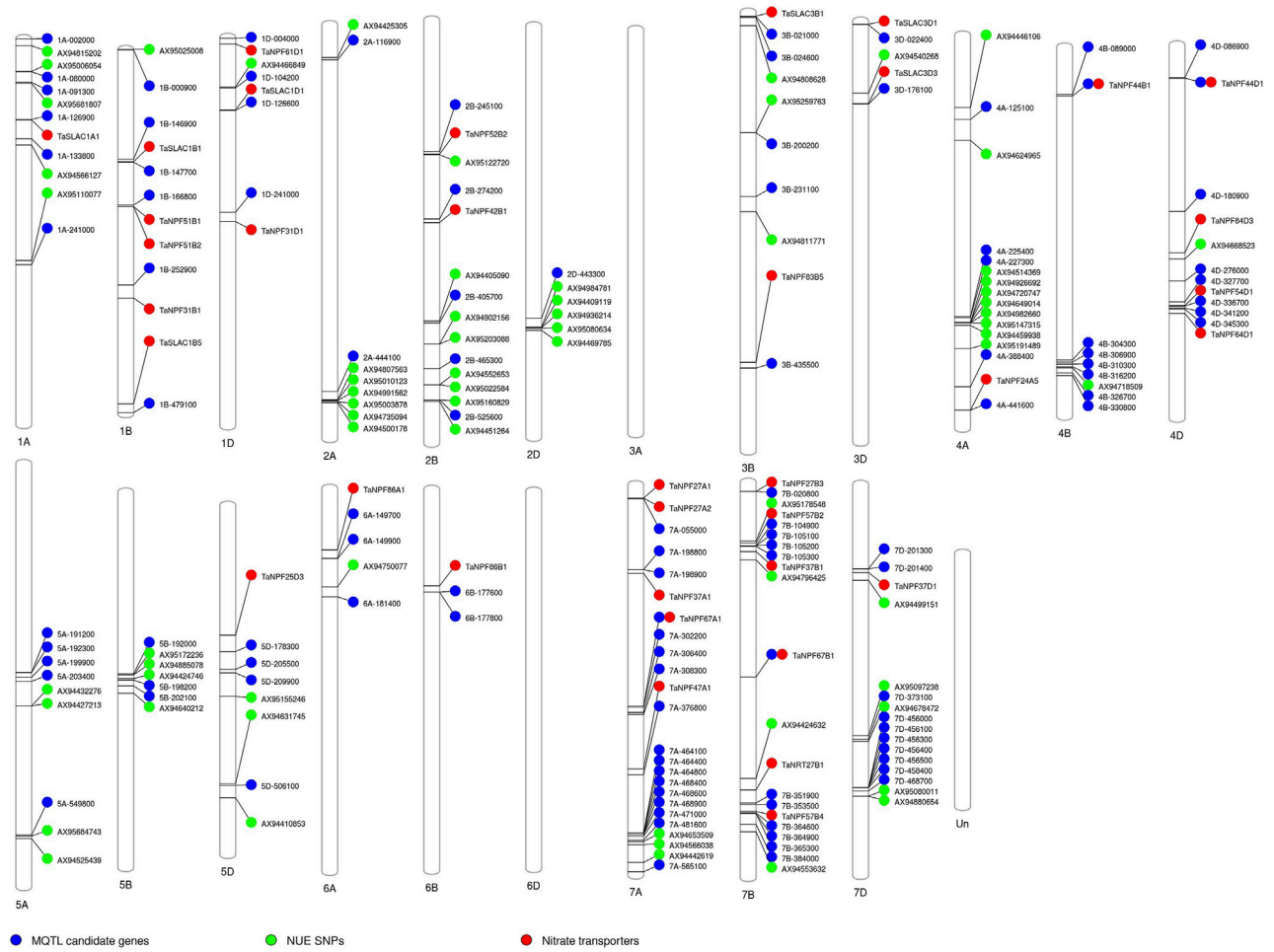


FIGURE 6 | Schematic representation of the distribution of previously reported single nucleotide polymorphism (SNP) associated with traits of interest, candidate genes reported in the present study, and the nitrate transporter along the 21 chromosomes of wheat. The chromosome map showing the candidate genes reported in the present study in the MQTL gene based on the synteny and their co-location with the previously reported single nucleotide polymorphism (SNP) associated with different nitrogen use efficiency (NUE)–related trait, root traits, yield, and yield-related traits (Sandhu et al., 2021), and the nitrate transporter reported in wheat (Kumar et al., 2021). The numbers below each rice chromosome indicate chromosome numbers.

TABLE 4 | Selected 15 candidate genes showing response to nitrogen treatment.

MQTL	Gene stable ID	Gene description	Gene name
MQTL2.1	*Os02g0179200*	Glutamine amidotransferase class-I domain containing protein	-
MQTL4.3	*Os04g0401000*	Proline-rich protein, Blast resistance	*Pi21*
MQTL5.2	*Os05g0183100*	Similar to WRKY transcription factor 16 (Fragment)	*OsWRKY67*
MQTL5.2	*Os05g0222200*	ABC transporter-like domain containing protein	*OsABCG12*
MQTL5.2	*Os05g0247100*	Similar to Glycosyl hydrolases family 18	*DIP3*
MQTL7.3	*Os07g0581700*	Homeodomain-leucine zipper (HD-Zip) transcription factor	*Oshox14*
MQTL7.3	*Os07g0603300*	TON1 RECRUIT MOTIF (TRM)-containing protein	*GL7*
MQTL7.5	*Os07g0668600*	Zinc finger, CCCH-type domain containing protein	*OsGZF1*
MQTL7.4	*Os07g0681200*	Vegetative storage protein/acid phosphatase domain containing protein	-
MQTL8.2	*Os08g0151300*	R2R3-MYB transcription factor	*CEF1*
MQTL9.2	*Os09g0286600*	Pathogenesis-related transcriptional factor and ERF domain containing protein	*Sub1C*
MQTL9.6	*Os09g0396900*	Protein of unknown function DUF125, transmembrane family protein	*OsVIT2*
MQTL10.5	*Os10g0403000*	Cytochrome P450 protein, CYP78A11	*PLA1*
MQTL10.6	*Os10g0463400*	B-type response regulator	*EF1*
MQTL10.6	*Os10g0469900*	TGF-beta receptor, type I/II extracellular region family protein	-

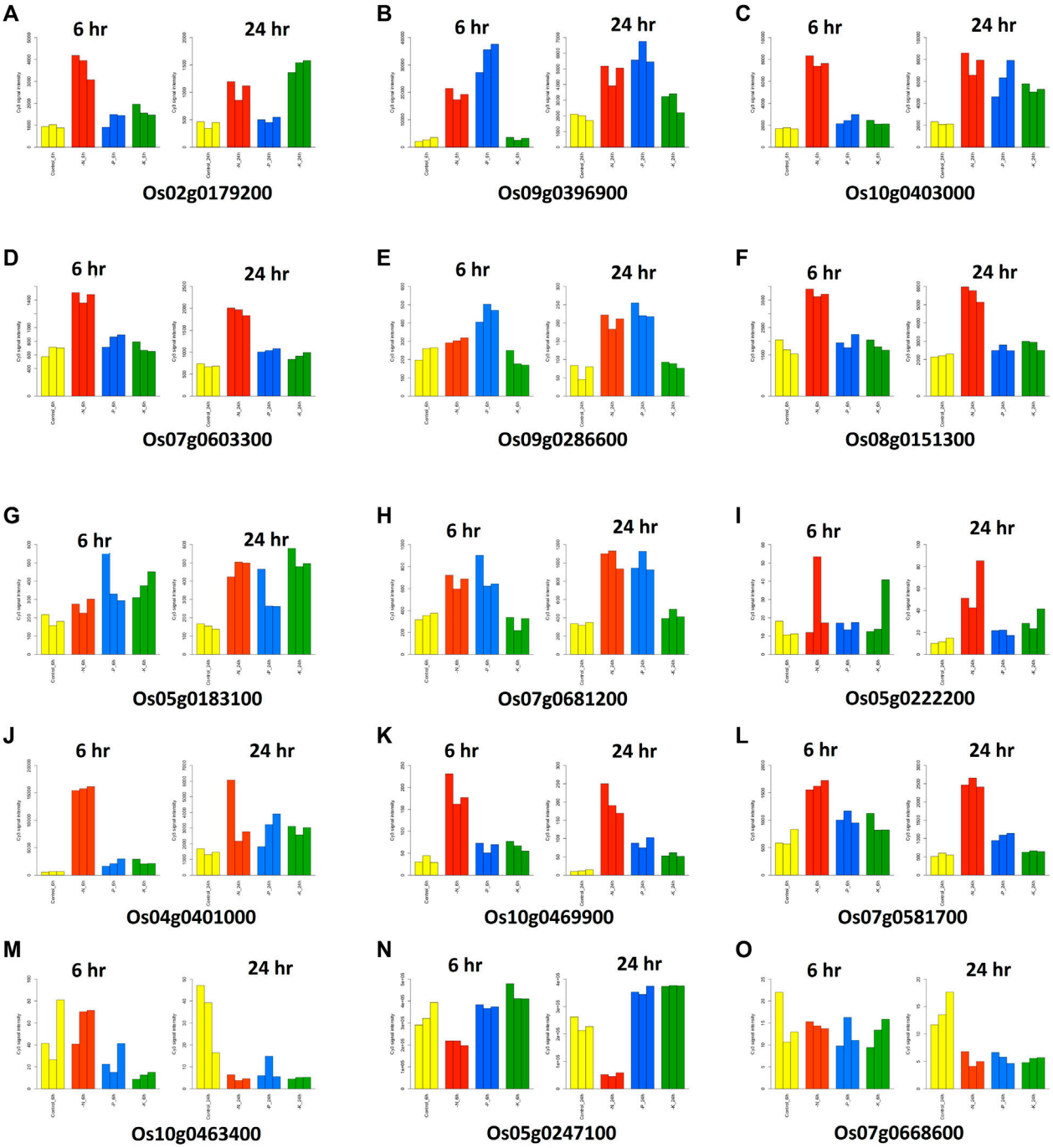


FIGURE 7 | Graphical representation of the significant changes in expression values of the candidate genes in response to nitrogen deficiencies in roots.

24 h. Meanwhile, one gene showed more expression at 6 h after treatment, but its expression was similar to control at 24 h after treatment (**Figure 7**). Three genes showed similar expressions at 6 h after treatment, but their expression decreased significantly after 24 h in comparison to control (**Figure 7**). These genes consisted of *EF1*(B type response regulator), *DIP3* (glycosyl hydrolase family 18), and *OsGZF1* (Zn finger CCCH domain containing protein) (**Table 4**). There were 5 genes that had multiple transcripts or splice variants, but no variation was detected between transcripts of the same gene.

DISCUSSION

Among all the plant nutrients basic for crop development, N is the nutrient limiting the crop productivity. Overabundance

use of nitrogenous fertilizers creates imbalance in ecosystem function and services (Fowler et al., 2013). There is a strong requirement to enhance nitrogen use efficiency for the sustainable agriculture production (Paul et al., 2014). The identification of genomic regions associated with traits of interest using molecular markers is an accurate and useful approach in the marker-assisted breeding program (Ashikari and Matsuoka, 2006; Price, 2006). The complex nature of the QTL and their interaction with other QTLs, genetic background, and environment are some of the important constraints in identifying their precise location. The identification of a major effect and consistent QTL across different genetic backgrounds and environments is an important requirement for the precise use of the identified genomic regions in marker-assisted selection.

The meta-analysis of the genomic regions reported in different studies helps in identification of most accurate and confined genomic regions to be further used in the marker-assisted introgression program. The trait NUE is not a biological measure by itself; it is a complex derivative of biological measures such as the grain yield/related traits and nitrogen response. A number of studies reported QTL associated with NUE in rice (Anis et al., 2019; Jewel et al., 2019; Mahender et al., 2019; Zhang et al., 2019). To the best of our knowledge, many QTLs involved in the NUE were reported separately, but a comprehensive listing and analysis of the genomic region associated with the grain yield and nitrogen use efficiency were not available. In the present study, the meta-QTL analysis was performed to identify the consistent and major effect QTL associated with nitrogen use efficiency in rice.

The present study explored 1,330 QTLs associated with NUE and related traits in rice and identified a total of 76 MQTLs, suggesting the power of MQTL analysis in narrowing down the genomic regions controlling the different traits of interest (Khahani et al., 2020). The number of available QTLs that we listed here was 1,330 QTLs of which 915 major QTLs were used for the identification of as many as 76 MQTLs, suggesting that to the best of our knowledge, the present study is so far the most comprehensive study for the identification of NUE-associated MQTLs in rice. The first attempt to map the genomic region associated with NUE and related traits was initiated at IRRI. The distribution of the QTL on different rice chromosomes with the highest number of QTL on chr 1 and 3 was similar to that of the previous reports (Swamy et al., 2011; Swamy and Sarla, 2011; Khahani et al., 2020). Some of the MQTLs identified in the present study constituted as high as 47 initial QTLs per MQTL (MQTL1.1), indicating the robustness of the detected MQTL. The consensus map developed in the present study is much informative relative to those prepared and used in the earlier studies for the identification of MQTL associated with the traits (Courtois et al., 2009; Swamy et al., 2011; Khahani et al., 2021; Kumari et al., 2021). The fold reduction in the size of the confidence interval of the MQTL detected in the present study is in contrast to those reported in the previous studies (Courtois et al., 2009; Danan et al., 2011). The reduction in the confidence interval allows for the exploration of most promising and less number of candidate genes per MQTL.

Here, we ran the MQTL analysis separately for the grain yield/yield-related traits and nitrogen use efficiency–related traits also to identify the major and consistent genomic regions associated solely with the trait. However, it is more effective and robust to pool different correlated traits measured in the same population (Goffinet and Gerber, 2000), assuming that a number of traits studied are pleiotropically related. The analysis of pooled correlated traits may be more powerful than the analysis of an individual trait ensuring better coverage of the genome than the single trait analysis. This enabled us to identify the stable QTL and "hotspots" for the NUE-related traits in rice across different genetic and environmental backgrounds to be further used in the marker-assisted introgression program.

Our approach of using all correlated traits altogether in the MQTL analysis in the present study was more robust, not only because we used all correlated traits but also we targeted only the genomic region (phenotypic variance >8) associated with NUE-related traits to shortlist the target genes. The co-location of root traits improving NUE and yield/yield-related traits in the duplicated region of the rice genome might be useful in identification of promising candidate genes controlling the traits of interest. The MQTLs reported in the present study showed a positive effect on the grain yield/yield-related traits, root traits improving nutrient uptake, NUE, and related traits. The presence of candidate genes *Os02g0207900* (MQTL2.4), *Os04g0379600* (MQTL4.3), *Os06g0286375* (MQTL6.4), and Os08g0155400 (MQTL8.2) encoded nitrate transporters, suggesting the efficacy of the detected MQTL. The MQTL on chr 1, 4, and 8 reported as a hotspot with 21, 12, and 12 yield/yield-related and NUE-related traits, respectively. The MQTL on chr 2 and 6 constituting genes associated with the nitrate/ammonium uptake/content (*OsNAR2.1, OsAMT1;1, OsPTR9*, *OsAAT49*) (Yan et al., 2011; Lu et al., 2012; Fang et al., 2013; Ranathunge et al., 2014). The identification of genes encoding the nitrate transporter may be very useful to target for improving nitrogen use efficiency under direct seeded cultivation conditions where the soil conditions are dry and nitrate is the major source of nitrogen. Various other reported genes and candidate genes associated with the floral organ development, grain size, grain weight, fertility, and root development have been found in the MQTL region detected in the present study (**Supplementary Tables S7, S8, S9**).

The rice genomic region subtending the MQTL4.3 (*Os04g0379600*) and MQTL8.2 (*Os08g0155400*) harbor the candidate genes encoding nitrate transporters; we identified the orthologue of these genes in wheat (*Os04g0379600: TraesCS2D02G279200, TraesCS2A02G280400, TraesCS2B02G297700; Os08g0155400: TraesCS7B02G201900, TraesCS7A02G301700, TraesCS7D02G297000*), maize (*Os04g0379600: Zm00001eb085850; Os08g0155400: Zm00001eb416090*), and barley (*Os08g0155400: HORVU7Hr1G071600*). The other genes associated with nitrogen uptake/transport/assimilation, amino acid synthesis/transport, protein synthesis/transport/phosphorylation in the MQTL region, and their orthologues in wheat, maize, and barley have been reported in the present study (**Supplementary Tables S7, S8**), providing a better

understanding of candidate genes controlling NUE with a similar evolutionary background and conserved the function between different cereal crops. The root-specific expression of the candidate genes was identified in the MQTL region under the nitrogen-deficit condition, suggesting the role of root traits in improving NUE in plants. The identification of nitrate transporter genes and previously identified marker-trait associations in rice and wheat (Sandhu et al., 2015; Sandhu et al., 2019; Subedi et al., 2019; Kumar et al., 2021; Sandhu et al., 2021) in close proximity to the MQTL/candidate genes/their orthologues associated with the NUE-related traits in the present study validates the efficacy of the reported MQTL. The 76 MQTLs reported in the present study were used for a further selection of some of the promising MQTL using the criteria; average phenotypic variance >10%, average LOD >4, and the involvement of ≥10 initial QTLs within the MQTL. Furthermore, the MQTLs were screened on the basis of phenotypic variance, LOD, and number of QTL in the MQTL region and resulted in the selection of 42 promising MQTLs, which were further used for the identification of candidate genes and expression studies. Once validated, the identified 15 target genes for NUE in the MQTL present on chromosomes 2,4,5,7,8,9, and 10 (**Table 4**) may be targeted for identification of donors possessing candidate genes and for marker development for further deployment in the marker-assisted breeding program.

CONCLUSIONS

NUE is a quantitative trait controlled by multiple genes and the co-localization of genomic regions associated with yield/yield-related traits and root traits improving NUE providing key candidate genes for the rice crop improvement. The MQTL analysis approach is used in the present study overcoming the limitation of QTL mapping while felicitating identification of robust markers and fine-mapped genomic regions to be further used in the marker-assisted introgression program. The present study identified 76 MQTLs associated with NUE and related traits in rice. The study also verified the evolutionary relationship of cereal crops through the mining of orthologues using the ortho-MQTL approach. The results reported in the present study will be applicable to improve the selection for the yield/yield-related traits and root traits improving nitrogen use efficiency in rice-breeding programs. The detailed crosstalk between the genome and proteome and the validation of identified putative candidate genes in the MQTL region through gene expression and gene editing studies may lay down the foundation to improve the nitrogen use efficiency of cereal crops.

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

NS designed this study and wrote the manuscript. GP, OP, and MPS collected the literature and performed the MQTL analysis. KR helped in collecting information on candidate genes in the MQTL region. NS, GP, and KP performed ortho-MQTL analysis. NS and AK worked on the identification of nitrate transporters and checked the efficacy of identified MQTL. MS helped in drafting the introduction section of the manuscript. MPS helped in drafting the material and methods section of the manuscript. GP, JS, and PA helped in the circos plot. DKS provided guidance in the MQTL analysis. All authors contributed to the manuscript and approved the submitted version.

FUNDING

The work was compiled under the projects funded by the Department of Biotechnology, Govt. of India (Grant Nos. BT/PR30871/BIC/101/1159/2018 and BT/PR31462/ATGC/127/6/2019) and GCRF Global Research Translation Awards, UKRI, United Kingdom.

ACKNOWLEDGMENTS

We are thankful to the Department of Biotechnology, Govt. of India, and GCRF Global Research Translation Awards, UKRI, United Kingdom, for providing grants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.807210/full#supplementary-material

1000 GW: | 1,000 grain weight, BS: bending strength, BM: bending moment, CD: culm diameter, DTF: days to 50% flowering, DTFirst: days to first emergence, DTFull: days to full emergence, Fe up: iron uptake, FLA: flag leaf area, FLAngle: flag leaf angle, GY: grain yield, NPT: number of productive tillers, NR: nodal root number, N up: nitrogen uptake, P up: phosphorus uptake, RGR: relative growth rate, RHL: root hair length, RHD: root hair density, PL: panicle length, PH: plant height, SPAD: chlorophyll content, SD: stem diameter, VVG: vegetative vigor, Zn up: zinc uptake.

REFERENCES

Acuña-Galindo, M. A., Mason, R. E., Subramanian, N. K., and Hays, D. B. (2015). Meta-analysis of Wheat QTL Regions Associated with Adaptation to Drought and Heat Stress. *Crop Sci.* 55 (2), 477–492.

Adachi, S., Tsuru, Y., Nito, N., Murata, K., Yamamoto, T., Ebitani, T., et al. (2011). Identification and Characterization of Genomic Regions on Chromosomes 4 and 8 that Control the Rate of Photosynthesis in rice Leaves. *J. Exp. Bot.* 62 (6), 1927–1938. doi:10.1093/jxb/erq387

Agrama, H. A. S., Zakaria, A. G., Said, F. B., and Tuinstra, M. (1999). Identification of Quantitative Trait Loci for Nitrogen Use Efficiency in maize. *Mol. Breed.* 5, 187–195. doi:10.1023/a:1009669507144

Akkareddy, S., Vemireddy, L. R., Hariprasad, A. S., Jayaprada, M., Sridhar, S., Ramanarao, P. V., et al. (2010). Identification and Mapping of Landrace Derived QTL Associated with Yield and its Components in rice under Different Nitrogen Levels and Environments. *Int. J. Plant Breed. Genet.* 4 (4), 210–227.

Anis, G. B., Zhang, Y., Islam, A., Zhang, Y., Cao, Y., Wu, W., et al. (2019). RDWN6XB, a Major Quantitative Trait Locus Positively Enhances Root System Architecture under Nitrogen Deficiency in rice. *BMC Plant Biol.* 19 (1), 12–13. doi:10.1186/s12870-018-1620-y

Anis, G., Zhang, Y., Xu, X., Fiaz, S., Wu, W., Rahman, M. H., et al. (2018). QTL Analysis for rice Seedlings under Nitrogen Deficiency Using Chromosomal Segment Substitution Lines. *Pak. J. Bot.* 50 (2), 537–544.

Ashikari, M., and Matsuoka, M. (2006). Identification, Isolation and Pyramiding of Quantitative Trait Loci for rice Breeding. *Trends Plant Sci.* 11, 344–350. doi:10.1016/j.tplants.2006.05.008

Brasier, K., Ward, B., Smith, J., Seago, J., Oakes, J., Balota, M., et al. (2020). Identification of Quantitative Trait Loci Associated with Nitrogen Use Efficiency in winter Wheat. *PLoS One* 15, e0228775. doi:10.1371/journal.pone.0228775

Cassman, K. G., Peng, S., Olk, D. C., Ladha, J. K., Reichardt, W., Dobermann, A., et al. (1998). Opportunities for Increased Nitrogen-Use Efficiency from Improved Resource Management in Irrigated rice Systems. *Field Crops Res.* 56 (1-2), 7–39. doi:10.1016/s0378-4290(97)00140-8

Chardon, F., Virlon, B., Moreau, L., Falque, M., Joets, J., Decousset, L., et al. (2004). Genetic Architecture of Flowering Time in maize as Inferred from Quantitative Trait Loci Meta-Analysis and Synteny Conservation with the rice Genome. *Genetics* 168 (4), 2169–2185. doi:10.1534/genetics.104.032375

Cho, Y. I., Jiang, W., Chin, J. H., Piao, Z., Cho, Y. G., McCouch, S., et al. (2007). Identification of QTLs Associated with Physiological Nitrogen Use Efficiency in riceMolecules & Cells. *Mol. Cell* 23 (1), 72–79.

Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted Selection: an Approach for Precision Plant Breeding in the Twenty-First century. *Phil. Trans. R. Soc. B* 363 (1491), 557–572. doi:10.1098/rstb.2007.2170

Courtois, B., Ahmadi, N., Khowaja, F., Price, A. H., Rami, J. F., Frouin, J., et al. (2009). Rice Root Genetic Architecture: Meta-Analysis from a Drought QTL Database. *Rice* 2 (2), 115–128. doi:10.1007/s12284-009-9028-9

Dai, G. J., Cheng, S. H., Hua, Z. T., Zhang, M. L., Jiang, H. B., Feng, Y., et al. (2015). Mapping Quantitative Trait Loci for Nitrogen Uptake and Utilization Efficiency in rice (*Oryza Sativa* L.) at Different Nitrogen Fertilizer Levels. *Genet. Mol. Res.* 14 (3), 10404–10414. doi:10.4238/2015.september.8.1

Danan, S., Veyrieras, J. B., and Lefebvre, V. (2011). Construction of a Potato Consensus Map and QTL Meta-Analysis Offer New Insights into the Genetic Architecture of Late Blight Resistance and Plant Maturity Traits. *BMC Plant Biol.* 11 (1), 16–17. doi:10.1186/1471-2229-11-16

Endelman, J. B., and Plomion, C. (2014). LPmerge: an R Package for Merging Genetic Maps by Linear Programming. *Bioinformatics* 30 (11), 1623–1624. doi:10.1093/bioinformatics/btu091

Fang, P., and Wu, P. (2001). QTL× N-Level Interaction for Plant Height in rice (*Oryza Sativa* L.). *Plant and Soil* 236 (2), 237–242. doi:10.1023/a:1012787510201

Fang, Z., Xia, K., Yang, X., Grotemeyer, M. S., Meier, S., Rentsch, D., et al. (2013). Altered Expression of thePTR/NRT1homologueOsPTR9affects Nitrogen Utilization Efficiency, Growth and Grain Yield in rice. *Plant Biotechnol. J.* 11 (4), 446–458. doi:10.1111/pbi.12031

FAO (2018). *World Fertilizer Trends and Outlook to 2018*. Rome: Food and Agriculture Organization of the United Nations.

Feng, B., Chen, K., Cui, Y., Wu, Z., Zheng, T., Zhu, Y., et al. (2018). Genetic Dissection and Simultaneous Improvement of Drought and Low Nitrogen Tolerances by Designed QTL Pyramiding in rice. *Front. Plant Sci.* 9, 306. doi:10.3389/fpls.2018.00306

Feng, Y., Cao, L. Y., Wu, W. M., Shen, X. H., Zhan, X. D., Zhai, R. R., et al. (2010). Mapping QTLs for Nitrogen-Deficiency Tolerance at Seedling Stage in rice (Oryza Sativa L.). *Plant Breed* 129 (6), 652–656. doi:10.1111/j.1439-0523.2009.01728.x

Feng, Y., Zhai, R.-R., Cao, L.-Y., Lin, Z.-C., Wei, X.-H., and Cheng, S.-H. (2011). QTL Analysis for Plant Height and Heading Date in Rice under Two Nitrogen Levels. *A a S* 37 (9), 1525–1532. doi:10.3724/sp.j.1006.2011.01525

Fontaine, J.-X., Ravel, C., Pageau, K., Heumez, E., Dubois, F., Hirel, B., et al. (2009). A Quantitative Genetic Study for Elucidating the Contribution of Glutamine Synthetase, Glutamate Dehydrogenase and Other Nitrogen-Related Physiological Traits to the Agronomic Performance of Common Wheat. *Theor. Appl. Genet.* 119, 645–662. doi:10.1007/s00122-009-1076-4

Fowler, D., Pyle, J. A., Raven, J. A., and Sutton, M. A. (2013). The Global Nitrogen Cycle in the Twenty-First century: Introduction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368 (1621), 1–2. doi:10.1098/rstb.2013.0165

Gallais, A., and Hirel, B. (2004). An Approach to the Genetics of Nitrogen Use Efficiency in maize. *J. Exp. Bot.* 55, 295–306. doi:10.1093/jxb/erh006

Goffinet, B., and Gerber, S. (2000). Quantitative Trait Loci: a Meta-Analysis. *Genetics* 155 (1), 463–473. doi:10.1093/genetics/155.1.463

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

International Rice Genome Sequencing Project (20052005). *IRGSP Releases the Assembled rice Genome Sequences*. Available at: http://rgp.dna.affrc.go.jp/IRGSP/Build2/build2.html.

Jewel, Z., Ali, J., Mahender, A., Hernandez, J., Pang, Y., and Li, Z. (2019). Identification of Quantitative Trait Loci Associated with Nutrient Use Efficiency Traits, Using SNP Markers in an Early Backcross Population of rice (Oryza Sativa L.). *Ijms* 20 (4), 900. doi:10.3390/ijms20040900

Jiang, H., Jiang, L., Guo, L., Gao, Z., Zeng, D., Zhu, L., et al. (2008). Conditional and Unconditional Mapping of Quantitative Trait Loci Underlying Plant Height and Tiller Number in rice (*Oryza Sativa* L.) Grown at Two Nitrogen Levels. *Prog. Nat. Sci.* 18 (12), 1539–1547. doi:10.1016/j.pnsc.2008.05.025

Kalugina, Z. I. (2014). Agricultural Policy in Russia: Global Challenges and the Viability of Rural Communities. *Int. J. Sociol. Agric. Food* 21, 115–131.

Kant, S., Bi, Y.-M., and Rothstein, S. J. (2011). Understanding Plant Response to Nitrogen Limitation for the Improvement of Crop Nitrogen Use Efficiency. *J. Exp. Bot.* 62, 1499–1509. doi:10.1093/jxb/erq297

Khahani, B., Tavakol, E., Shariati, V., and Fornara, F. (2020). Genome Wide Screening and Comparative Genome Analysis for Meta-QTLs, Ortho-MQTLs and Candidate Genes Controlling Yield and Yield-Related Traits in rice. *BMC Genomics* 21 (1), 294–317. doi:10.1186/s12864-020-6702-1

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), 6942–6959. doi:10.1038/s41598-021-86259-2

Khahani, B., Tavakol, E., and Shariati, 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

Kraiser, T., Gras, D. E., Gutierrez, A. G., Gonzalez, B., and Gutierrez, R. A. (2011). A Holistic View of Nitrogen Acquisition in Plants. *J. Exp. Bot.* 62, 1455–1466. doi:10.1093/jxb/erq425

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an Information Aesthetic for Comparative Genomics. *Genome Res.* 19 (9), 1639–1645. doi:10.1101/gr.092759.109

Kumar, A., Sandhu, N., Kumar, P., Pruthi, G., Singh, J., Kaur, S., et al. (2021). Genome-wide Analysis of NPF, NRT2, CLC and SLAC1/SLAH Nitrate Transporters in Hexaploid Wheat (*Triticum aestivum*). *Sci. Rep.*. (under review).

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, 627955. doi:10.3389/fpls.2021.627955

Laza, M. R., Kondo, M., Ideta, O., Barlaan, E., and Imbe, T. (2006). Identification of Quantitative Trait Loci for δ13C and Productivity in Irrigated Lowland Rice. *Crop Sci.* 46 (2), 763–773. doi:10.2135/cropsci2005.05.0084

Lei, L., Zheng, H. L., Wang, J. G., Liu, H. L., Sun, J., Zhao, H. W., et al. (2018). Genetic Dissection of rice (Oryza Sativa L.) Tiller, Plant Height, and Grain Yield Based on QTL Mapping and Metaanalysis. *Euphytica* 214, 1–7. doi:10.1007/s10681-018-2187-2

Li, H., Hu, B., and Chu, C. (2017). Nitrogen Use Efficiency in Crops: Lessons from Arabidopsis and rice. *J. Exp. Bot.* 68, 2477–2488. doi:10.1093/jxb/erx1010.1093/jxb/erx101

Li, W.-T., Liu, C., Liu, Y.-X., Pu, Z.-E., Dai, S.-F., Wang, J.-R., et al. (2013). Meta-analysis of QTL Associated with Tolerance to Abiotic Stresses in Barley. *Euphytica* 189, 31–49. doi:10.1007/s10681-012-0683-3

Lu, Y., Song, Z., Lü, K., Lian, X., and Cai, H. (2012). Molecular Characterization, Expression and Functional Analysis of the Amino Acid Transporter Gene Family (*OsAATs*) in rice. *Acta Physiol. Plant* 34 (5), 1943–1962. doi:10.1007/s11738-012-0995-x

MacMillan, K., Emrich, K., Piepho, H.-P., Mullins, C. E., and Price, A. H. (2006). Assessing the Importance of Genotype × Environment Interaction for Root Traits in rice Using a Mapping Population II: Conventional QTL Analysis. *Theor. Appl. Genet.* 113, 953–964. doi:10.1007/s00122-006-0357-4

Mahender, A., Ali, J., Prahalada, G. D., Sevilla, M. A. L., Balachiranjeevi, C. H., Md, J., et al. (2019). Genetic Dissection of Developmental Responses of Agro-Morphological Traits under Different Doses of Nutrient Fertilizers Using High-Density SNP Markers. *PLoS One* 14 (7), e0220066. doi:10.1371/journal.pone.0220066

Manangkil, J. M., Niones, J. M., Undan, J. R., Obara, M., Mananghaya, T. E., Mallari, R. P., et al. (2019). Quantitative Trait Loci Associated with Root Elongation Ability of rice under Nitrogen-Deficient Condition. *Philip. J. Sci.* 148 (2), 401–409.

Nayar, N. M. (2014). *The Origin of Asian rice. Origin and Phylogeny of Rices*. Chapter 6. Academic Press, Elsevier, 169–253. doi:10.1016/b978-0-12-417177-0.00006-1

Nguyen, H. T. T., Van Pham, C., and Bertin, P. (2014). The Effect of Nitrogen Concentration on Nitrogen Use Efficiency and Related Parameters in Cultivated Rices (*Oryza Sativa* L. Subsp. *Indica* and *Japonica* and *O. Glaberrima* Steud.) in Hydroponics. *Euphytica* 198 (1), 137–151. doi:10.1007/s10681-014-1101-9

Norton, R., Davidson, E., and Roberts, T. (2015). *Position Paper - Nitrogen Use Efficiency and Nutrient Performance Indicators*. Nairobi: Global Partnership on Nutrient Management.

Obara, M., Ishimaru, T., Abiko, T., Fujita, D., Kobayashi, N., Yanagihara, S., et al. (2014). Identification and Characterization of Quantitative Trait Loci for Root Elongation by Using Introgression Lines with Genetic Background of *Indica*-type rice Variety IR64. *Plant Biotechnol. Rep.* 8 (3), 267–277. doi:10.1007/s11816-014-0320-9

Orjuela, J., Garavito, A., Bouniol, M., Arbelaez, J. D., Moreno, L., Kimball, J., et al. (2010). A Universal Core Genetic Map for rice. *Theor. Appl. Genet.* 120 (3), 563–572. doi:10.1007/s00122-009-1176-1

Paul, K., Chopra, N. K., Soni, P. G., Kumar, R., and Mondal, G. (2014). Influence of Different Nitrogen Levels and weed Control on Yield and Chemical Composition of Mustard (*Brassica Rapa* L. Sub. Chinensis) Fodder. *Ind. J. Ani. Nut.* 31 (4), 400–403.

Pozzo, T., Higdon, S. M., Pattathil, S., Hahn, M. G., and Bennett, A. B. (2018). Characterization of Novel Glycosyl Hydrolases Discovered by Cell wall Glycan Directed Monoclonal Antibody Screening and Metagenome Analysis of maize Aerial Root Mucilage. *PLoS One* 13, e0204525. doi:10.1371/journal.pone.0204525

Price, A. H. (2006). Believe it or Not, QTLs Are Accurate!. *Trends Plant Sci.* 11 (5), 213–216. doi:10.1016/j.tplants.2006.03.006

Ranathunge, K., El-Kereamy, A., Gidda, S., Bi, Y.-M., and Rothstein, S. J. (2014). AMT1;1 Transgenic rice Plants with Enhanced NH4+ Permeability Show superior Growth and Higher Yield under Optimal and Suboptimal NH4+ Conditions. *J. Exp. Bot.* 65 (4), 965–979. doi:10.1093/jxb/ert458

Ribaut, J.-M., Fracheboud, Y., Monneveux, P., Banziger, M., Vargas, M., and Jiang, C. (2007). Quantitative Trait Loci for Yield and Correlated Traits under High and Low Soil Nitrogen Conditions in Tropical maize. *Mol. Breed.* 20, 15–29. doi:10.1007/s11032-006-9041-2

Sandhu, N., Subedi, S. R., Singh, V. K., Sinha, P., Kumar, S., Singh, S. P., et al. (2019). Deciphering the Genetic Basis of Root Morphology, Nutrient Uptake, Yield, and Yield-Related Traits in rice under Dry Direct-Seeded Cultivation Systems. *Sci. Rep.* 9 (1), 9334–9350. doi:10.1038/s41598-019-45770-3

Sandhu, N., Kaur, A., Sethi, M., Kaur, S., Varinderpal-Singh, V., Sharma, A., et al. (2021). Genetic Dissection Uncovers Genome-wide Marker-Trait Associations for Plant Growth, Yield, and Yield-Related Traits under Varying Nitrogen Levels in Nested Synthetic Wheat Introgression Libraries. *Front. Plant Sci.* 12, 738710. doi:10.3389/fpls.2021.738710

Sandhu, N., Raman, K. A., Torres, R. O., Audebert, A., Dardou, A., Kumar, A., et al. (2016). Rice Root Architectural Plasticity Traits and Genetic Regions for Adaptability to Variable Cultivation and Stress Conditions. *Plant Physio* 171 (4), 2562–2576. doi:10.1104/pp.16.00705

Sandhu, N., Torres, R. O., Sta Cruz, M. T., Maturan, P. C., Jain, R., Kumar, A., et al. (2015). Traits and QTLs for Development of Dry Direct-Seeded Rainfed rice Varieties. *J. Exp. Bot.* 66 (1), 225–244. doi:10.1093/jxb/eru413

Semagn, K., Beyene, Y., Warburton, M. L., Tarekegne, A., Mugo, S., Meisel, B., et al. (2013). Meta-analyses of QTL for Grain Yield and Anthesis Silking Interval in 18 maize Populations Evaluated under Water-Stressed and Well-Watered Environments. *BMC Genomics* 14 (1), 313–316. doi:10.1186/1471-2164-14-313

Senthilvel, S., Vinod, K. K., Malarvizhi, P., and Maheswaran, M. (2008). QTL and QTL × Environment Effects on Agronomic and Nitrogen Acquisition Traits in Rice. *J. Integ. Plant Bio.* 50 (9), 1108–1117. doi:10.1111/j.1744-7909.2008.00713.x

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

Subedi, S. R., Sandhu, N., Singh, V. K., Sinha, P., Kumar, S., Singh, S. P., et al. (2019). Genome-wide Association Study Reveals Significant Genomic Regions for Improving Yield, Adaptability of rice under Dry Direct Seeded Cultivation Condition. *BMC Genomics* 20 (1), 471–491. doi:10.1186/s12864-019-5840-9

Swamy, B. P., Vikram, P., Dixit, S., Ahmed, H. U., and Kumar, A. (2011). Meta-analysis of Grain Yield QTL Identified during Agricultural Drought in Grasses Showed Consensus. *BMC Genomics* 12 (1), 319–336. doi:10.1186/1471-2164-12-319

Swamy, B. P. M., and Sarla, N. (2011). Meta-analysis of Yield QTLs Derived from Inter-specific Crosses of rice Reveals Consensus Regions and Candidate Genes. *Plant Mol. Biol. Rep.* 29 (3), 663–680. doi:10.1007/s11105-010-0274-1

Van Deynze, A., Zamora, P., Delaux, P.-M., Heitmann, C., Jayaraman, D., Rajasekar, S., et al. (2018). Nitrogen Fixation in a Landrace of maize Is Supported by a Mucilage-Associated Diazotrophic Microbiota. *Plos Biol.* 16, e2006352. doi:10.1371/journal.pbio.2006352

Van, K., and McHale, L. (2017). Meta-analyses of QTLs Associated with Protein and Oil Contents and Compositions in Soybean (Glycine max (L.) Merr.) Seed. *Ijms* 18 (6), 1180. doi:10.3390/ijms18061180

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 Bioinformatics* 8 (1), 49–16. doi:10.1186/1471-2105-8-49

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

Wei, D., Cui, K., Pan, J., Ye, G., Xiang, J., Nie, L., et al. (2011). Genetic Dissection of Grain Nitrogen Use Efficiency and Grain Yield and Their Relationship in rice. *Field Crops Res.* 124 (3), 340–346. doi:10.1016/j.fcr.2011.07.003

Wei, D., Cui, K., Ye, G., Pan, J., Xiang, J., Huang, J., et al. (2012). QTL Mapping for Nitrogen-Use Efficiency and Nitrogen-Deficiency Tolerance Traits in rice. *Plant and Soil* 359 (1), 281–295. doi:10.1007/s11104-012-1142-6

Yan, M., Fan, X., Feng, H., Miller, A. J., Shen, Q., and Xu, G. (2011). Rice *OsNAR2.1* Interacts with *OsNRT2.1, OsNRT2.2* and *OsNRT2.3a* Nitrate Transporters to Provide Uptake over High and Low Concentration Ranges. *Plant Cel Environ.* 34 (8), 1360–1372. doi:10.1111/j.1365-3040.2011.02335.x

Yang, S. M., Zhang, F. F., Zhang, S. H., Li, G. Y., Zeng, L. Q., Liu, G. S., et al. (2019). QTL Mapping of Physiological Traits at the Booting Stage in rice under Low

Temperature Combined with Nitrogen Fertilization. *Czech J. Genet. Plant Breed.* 55 (4), 146–155. doi:10.17221/67/2018-cjgpb

Ye, T., Li, Y., Zhang, J., Hou, W., Zhou, W., Lu, J., et al. (2019). Nitrogen, Phosphorus, and Potassium Fertilization Affects the Flowering Time of rice (Oryza Sativa L.). *Glob. Ecol. Conservation* 20, e00753. doi:10.1016/j.gecco.2019.e00753

Yue, F., Rong-rong, Z., Ze-chuan, L., Li-yong, C., Xing-hua, W., and Shi-hua, C. (2015). Quantitative Trait Locus Analysis for rice Yield Traits under Two Nitrogen Levels. *Rice Sci.* 22 (3), 108–115. doi:10.1016/j.rsci.2015.05.014

Zhang, M., Gao, M., Zheng, H., Yuan, Y., Zhou, X., Guo, Y., et al. (2019). QTL Mapping for Nitrogen Use Efficiency and Agronomic Traits at the Seedling and Maturity Stages in Wheat. *Mol. Breed.* 39, 71. doi:10.1007/s11032-019-0965-8

Zhang, X., Shabala, S., Koutoulis, A., Shabala, L., and Zhou, M. (2017). Meta-analysis of Major QTL for Abiotic Stress Tolerance in Barley and Implications for Barley Breeding. *Planta* 245 (2), 283–295. doi:10.1007/s00425-016-2605-4

Zhao, C.-F., Zhou, L.-H., Zhang, Y.-D., Zhu, Z., Chen, T., Zhao, Q.-Y., et al. (2014). QTL Mapping for Seedling Traits Associated with Low-Nitrogen Tolerance Using a Set of Advanced Backcross Introgression Lines of rice. *Plant Breed* 133 (2), 189–195. doi:10.1111/pbr.12123

Zhao, X., Peng, Y., Zhang, J., Fang, P., and Wu, B. (2018). Identification of QTLs and Meta-QTLs for Seven Agronomic Traits in Multiple Maize Populations under Well-Watered and Water-Stressed Conditions. *Crop Sci.* 58, 507–520. doi:10.2135/cropsci2016.12.0991

Zhou, Y., Tao, Y., Tang, D., Wang, J., Zhong, J., Wang, Y., et al. (2017). Identification of QTL Associated with Nitrogen Uptake and Nitrogen Use Efficiency Using High Throughput Genotyped CSSLs in rice (Oryza Sativa L.). *Front. Plant Sci.* 8, 1166. doi:10.3389/fpls.2017.01166

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 March 2022
doi: 10.3389/fgene.2022.819849

SNP-Based Genome-Wide Association Mapping of Pollen Viability Under Heat Stress in Tropical *Zea mays* L. Inbred Lines

Zubair Ahmed[1,2], Maria Khalid[3], Abdul Ghafoor[4], Muhammad Kausar Nawaz Shah[1], Ghazala Kaukab Raja[3], Rashid Mehmood Rana[1], Tahir Mahmood[1] and Addie M. Thompson[5]**

[1]Department of Plant Breeding and Genetics, Pir Mehar Ali Shah Arid Agriculture University, Rawalpindi, Pakistan, [2]Crop Disease Research Institute, National Agricultural Research Center (Pakistan), Islamabad, Pakistan, [3]Institute of Biochemistry and Biotechnology, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan, [4]Pakistan Agricultural Research Council, Islamabad, Pakistan, [5]Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI, United States

OPEN ACCESS

Edited by:
Ahmad M. Alqudah, Aarhus University, Denmark

Reviewed by:
Muhammad Azhar Nadeem, Sivas University of Science and Technology, Turkey
Mohsin Ali, Chinese Academy of Agricultural Sciences (CAAS), China
Pallavi Sinha, IRRI South Asia Hub, India

***Correspondence:**
Zubair Ahmed
zubairnarc15@gmail.com
Addie M. Thompson
thom1718@msu.edu

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *22 November 2021*
Accepted: *03 February 2022*
Published: *16 March 2022*

Citation:
Ahmed Z, Khalid M, Ghafoor A, Shah MKN, Raja GK, Rana RM, Mahmood T and Thompson AM (2022) SNP-Based Genome-Wide Association Mapping of Pollen Viability Under Heat Stress in Tropical Zea mays L. Inbred Lines. Front. Genet. 13:819849. doi: 10.3389/fgene.2022.819849

Global environmental changes with more extreme episodes of heat waves are major threats to agricultural productivity. Heat stress in spring affects the reproductive stage of maize, resulting in tassel blast, pollen abortion, poor pollination, reduced seed set, barren ears and ultimately yield loss. As an aneamophelous crop, maize has a propensity for pollen abortion under heat stress conditions. To overcome the existing challenges of heat stress and pollen abortion, this study utilized a broad genetic base of maize germplasm to identify superior alleles to be utilized in breeding programs. A panel of 375 inbred lines was morpho-physiologically screened under normal and heat stress conditions in two locations across two consecutive planting seasons, 2017 and 2018. The exposure of pollen to high temperature showed drastic decline in pollen germination percentage. The average pollen germination percentage (PGP) at 35 and 45°C was 40.3% and 9.7%, respectively, an average decline of 30.6%. A subset of 275 inbred lines were sequenced using tunable genotyping by sequencing, resulting in 170,098 single nucleotide polymorphisms (SNPs) after filtration. Genome wide association of PGP in a subset of 122 inbred lines resulted in ten SNPs associated with PGP35°C ($p \leq 10^{-5}$), nine with PGP45°C ($p \leq 10^{-6}$–10^{-8}) and ten SNPs associated with PGP ratio ($p \leq 10^{-5}$). No SNPs were found to be in common across PGP traits. The number of favorable alleles possessed by each inbred line for PGP35°C, PGP45°C, and the PGP ratio ranged between 4 and 10, 3–13 and 5–13, respectively. In contrast, the number of negative alleles for these traits ranged between 2 and 8, 3–13 and 3–13, respectively. Genetic mapping of yield (adjusted weight per plant, AWP^{-1}) and flowering time (anthesis-silking interval, ASI) in 275 lines revealed five common SNPs: three shared for AWP^{-1} between normal and heat stress conditions, one for ASI between conditions, and one SNP, CM007648.1-86615409, was associated with both ASI and AWP^{-1}. Variety selection can be performed based on these favorable alleles for various traits. These marker trait associations identified in the diversity panel can be utilized in breeding programs to improve heat stress tolerance in maize.

Keywords: maize, extreme temperature, genetic mapping, pollen sterility, yield, flowering time

INTRODUCTION

Global maize production exceeds 1,108 million tons, making it one of the most widely grown cereal crops around the world (FAOSTAT, 2020). Maize production is continuing to increase by 7.4% annually in Pakistan, but changing climate conditions and global warming pose serious threats to maize productivity around the world, particularly in the Indian subcontinent. Since 1850, each of the last 4 decades have been successively warmer, and global surface temperature has risen 1°C during 2011–2021 relative to 1850–1900 (IPCC, 2021). Each degree-Celsius increase in global temperature will reduce the global maize yield by 7.4% (Zhao et al., 2017). It is predicted that extreme heat stress during anthesis will reduce maize production by 45% from 1980 to 2080 (Deryng et al., 2014).

Maize plants are particularly sensitive to heat stress at the flowering stage, causing more yield reduction than during the grain filling stage. (Zhang et al., 2013). The reproductive tissues are the most sensitive parts of the plant, so a few degrees increase in temperature at the flowering stage can cause devastating losses in grain yield (Lobell et al., 2015) High temperatures (33–40°C) also have a negative effect on light capture, harvest index, and grain and biomass yields. Breeding for climate resilient maize is the only solution to overcome climatic adversities as predicted by climate change models. Evaluating maize germplasm for pollen viability is an effective approach for the development of climate resilient hybrids. Phenotypic studies along with genomic information can reduce breeding time by pinpointing suitable genomic regions and improving selection efficiency (Kulwal et al., 2014).

The genome wide association study (GWAS) has emerged as a powerful approach for identifying genes underlying complex morphological traits (Wang et al., 2016; Yano et al., 2016; Mwadzingeni et al., 2017; Nadeem et al., 2021), in which a diverse natural population is used to detect the statistical association between markers and traits. Next generation sequencing techniques like genotyping-by sequencing (GBS) scan the genome and generate millions of SNPs, providing dense genome coverage for the identification of desirable marker-trait associations in different plant species (Atwell et al., 2010; Huang et al., 2010; Tian et al., 2011). In maize, genome wide association studies have been employed for flowering time (Salvi et al., 2007), kernel size and weight (Li et al., 2010), kernel quality (Manicacci et al., 2009), drought tolerance (Thirunavukkarasu et al., 2014), and various other target traits (Xue et al., 2013), including root system architecture under drought stress (Zaidi et al., 2016; Darlene et al., 2018), and plant leaf angle and lodging and heat stress tolerance in subtropical maize (Longmei et al., 2021). Though mapping has been conducted for flowering time and yield, no GWAS is previously reported in maize considering pollen viability directly as the target trait.

The objectives of this study were: 1) evaluation of a maize diversity panel consisting of exotic and indigenous maize inbred lines for heat stress tolerance, considering pollen viability, yield, and ASI as the target traits, and 2) identification of marker-trait associations and quantification of their effect on the traits.

MATERIALS AND METHODS

Plant Material and Experimental Design

A maize diversity panel composed of 375 diverse inbred lines was used for this study. The diversity panel includes 103 exotic lines, 283 indigenous lines, and three check cultivars: Haq Nawaz Gold, FH-1898, and CML161. The diversity panel was formulated by collection of diverse maize genotypes from exotic sources (CIMMYT, Pakistan) and from plant genetic resources (PGRI), NARC, Pakistan. The details of the diversity panel are given in **Supplementary Table S1**. The trial was planted for two spring seasons in 2017 and 2018 each with normal (1st week of March) and late (1st week of April) sowing dates at both the National Agricultural Research Center (NARC), Islamabad, Pakistan, and Peshawar, Pakistan.

Two different sowing dates were used so that the flowering periods would be synchronized to both a normal temperature and a heat stress temperature in each year. Significant temperature differences at flowering time were observed in normal and stressed sowing trials. The normal sowing trials flowered during the end of May, and temperatures at the flowering stage (VT) were in the range of 29–35°C. The stressed sowing trials flowered during early to mid-July, and the temperatures at VT stage were in the range of 38–45°C.

Seeds were hand sown in single row plots of length 4 m using a dibbler at a depth of 3 cm. Row to row distance was maintained at 75 cm while plant to plant distance was maintained at 25 cm such that each plot contained 16 plants, and fertilizer was applied at the rate of 200:100:100 NPK. Manual thinning and weeding operations were performed to retain optimum plant numbers and weed control in the experiments. The same experimental plots were used for both years in both locations.

Phenotyping

Phenotypic data of various agronomic traits were collected according to published maize descriptors (CIMMYT/IBPGR 1991). To assess the relationship of other phenotypic traits to pollen viability, days to anthesis (DTA), days to silking (DTS), anthesis silking interval (ASI), and adjusted weight per plant (AWP^{-1}) were collected from 5 plants for each of the 375 varieties. At tasseling (VT) stage, DTA, DTS and ASI were recorded, while yield was measured at the R6 stage. Assessment of pollen germination was carried out on a subset of 122 lines using pollen germination media (PGM) proposed by Dresselhaus et al. (2011), using pollen collected from the first (non-stressed) planting of the NARC trail in 2017. Fresh pollen was collected between 7am with 70% relative humidity and 8am with 63% relative humidity. Anthers, insects and other contaminants were cleaned by sieving the collected pollen. Pollen from each inbred line was dusted on six disposable petri plates each containing 20 ml of PGM. Out of six, three petri plates were placed at 35°C and three at 45°C in a growth chamber under dark conditions with 70% relative humidity. The pollen was observed under a light microscope and photographed. The pollen germinated percentage (PGP) was recorded by counting viable pollen grains out of 100 pollen grains from each plate at two different temperature regimes i.e., 35 and

45°C. From these values, three traits were recorded: PGP 35°C, PGP 45°C, and the ratio between the two.

Statistical Analysis

Descriptive statistics of various morphological traits were calculated by using Microsoft Excel 365 data analysis tools. The recorded data for all the parameters were used for the analysis of variance (ANOVA) (Finney, 1964). Mean square values of each source of variation were used to calculate the genotypic variance σ^2G and phenotypic variance σ^2P, and Broad-sense heritability was calculated using formula $H^2 = \sigma^2G/\sigma^2P$. Field experiment data was analyzed as a randomized design in R software (R Core Team, 2018).

Genotypic Data

Genotypic data of 275 maize accessions were generated by Freedom Markers from fresh tissue via tunable genotyping by sequencing (Ott et al., 2017) with the restriction enzyme Bsp1286I. Briefly, genotypes were sequenced using an Illumina HiSeq X instrument, and reads were aligned to the *Zea mays* AGPv4 reference genome after de-barcoding and trimming of reads. SNP calling was conducted using only those reads that aligned to a single location in the reference genome. Initially, nearly 2.7 million SNPs were identified, with a mean missing data rate per SNP site of 64%. From these, a high-quality set of 496,740 SNPs in which each marker was genotyped in at least 50% of the samples was generated and referred to hereafter as MCR50 SNPs. Each of these SNPs were supported on average by 20 tGBS reads per SNP per genotyped sample. Finally, imputation was performed on the MCR50 SNPs using BeagleV4.1 (Browning and Browning, 2007; Browning and Browning, 2016) with 50 phasing iterations and other default parameters. This marker set was further filtered to exclude sites with a minor allele frequency <5% in the set of phenotyped varieties, yielding a final marker dataset of 170,098 sites across 262 varieties. These were relatively evenly distributed across the 10 chromosomes (**Figure 2**).

Population structure and kinship matrices of the sets of 122 (for pollen traits) and 275 (for other traits) maize accessions were generated within the rMVP program (Yin et al., 2021) in R software (R Core Team, 2018).

Association Mapping

Phenotypic data of all accessions were checked for outliers, and all high-quality data was used for association studies. The FarmCPU algorithm (Liu et al., 2016) within the rMVP package (Yin et al., 2021) of R (R Core Team, 2018) was used for association mapping, with the kinship matrix (k matrix) and an optimal number of principal components to account for population structure (Q matrix) and prevent false-positive associations. Quantile-quantile (Q-Q) plots of estimated versus observed *p*-values for marker-trait associations were produced to assess model fit and select the appropriate number of principle components for each trait. Bonferroni corrections at 0.05 yielded a significance threshold of 2.94×10^{-7} (0.05 threshold divided by 170,098 markers). Because Bonferroni can be overly stringent due to linkage, all SNPs with *p*-values lower than 10^{-5} are reported here as potential associations as a hypothesis-generating approach.

RESULTS

Morphological Analysis and Correlation Among Traits

Morphological data from the maize diversity panel was combined across the 2 years in the study. Phenotypic analysis showed considerable variation for all traits under both normal and heat stress conditions (**Table 1**). The diversity panel was evaluated based on five agro-morphological traits to inform the assessments of pollen viability, the major trait under study. Average pollen germination percentage (PGP) at 35°C temperature was 40.3% with a range of 0%–93.7%. In contrast, the average PGP at 45°C temperature was 9.7% with a range of 0%–57%. The average decline in PGP from 35 to 45°C was 30.6%. The mean AWP^{-1}, DTA, DTS, and ASI was 37.5 g, 62.5, 64.6, and 2.9 days, respectively, which were significantly reduced to 35.5 g, 47.4, 50, and 2.4 days under heat stress (**Table 1**). ANOVA results also indicated that significant variation was observed among the genotypes under both normal and heat stress conditions in both years.

Broad-sense heritability of all traits was estimated from the NARC location and ranged between 0.32 for ASI to 0.91 for DTS under normal sowing conditions which was significantly reduced to 0.21 for DTS to 0.44 for AWP^{-1} under heat stress (**Table 1**).

Relationships Among Traits

Correlation analysis was performed on the combined data of normal and stress trials. The PGP at 35°C showed positive correlation with AWP^{-1} (r = 0.54), DTA (r = 0.49) and DTS (r = 0.14), but PGP at 45°C showed negative correlation with AWP^{-1} (r = −0.3), DTA (r = −0.07) and DTS (r = −0.1). On the other hand, DTA and DTS showed negative correlation with yield under heat stress conditions, at (r = −0.64) and (r = −0.76), respectively (**Table 2**).

Genome Wide Association Study

Pollen Germination Percentage PGP 35°C, PGP 45°C and the ratio between the two were used for GWAS. In addition, GWAS was also performed for AWP^{-1} and ASI for two locations across two seasons and two sowing dates. Q-Q plots representing the distribution of SNP significance for each PGP trait are represented in **Figure 1**. A total of 29 potential SNPs were identified that may have association with the target traits (**Figure 2**; **Table 3**). Despite the smaller subset of varieties used to map PGP, GWAS identified 14 highly significant SNPs associated with PGP traits: one with PGP35°C on chromosome 2, nine with PGP45°C on chromosomes 1,4,5 & 9, and four with PGP ratio on chromosomes 2 and 9.

No common association of SNPs were found for the PGP traits under study. SNP CM007650.1-89641171 located on 89.64 MB position of chromosome 1 showed the highest (0.25) R^2 values, associated with PGP45°C. Among each trait various SNPs with highest R^2 values were observed.

TABLE 1 | Descriptive statistics of various geomorphological traits in subset of 122 diverse maize inbred lines planted under normal and heat stress conditions across 2 years (spring 2017 and 2018) at Islamabad (NARC).

Trait	Condition	Mean	SD	Min	Max	σ^2 P	σ^2 G	H^2
PGP35°C	**Normal**	40.30	29.60	0.00	93.70	1749.08***	851.79***	0.49
AWP^{-1}		37.50	9.10	8.00	85.00	1735.46***	1512.98***	0.87
DTA		62.50	6.90	51.00	80.70	161***	94.7***	0.59
DTS		64.60	7.70	38.30	84.00	194.60***	177.5***	0.91
ASI		2.90	0.70	1.70	6.70	0.95**	0.31**	0.32
PGP45°C	**Stress**	9.70	11.70	0.00	57.00	282.17**	112.67**	0.40
AWP^{-1}		35.50	20.00	0.00	122.30	123.71***	54.23***	0.44
DTA		47.40	0.71	30.00	7.60	42***	15***	0.36
DTS		50	0.70	34.00	72.60	50.5***	10.5***	0.21
ASI		2.90	0.04	1.33	7.20	0.85	0.30	0.35

** = 0.05% significant, *** = 0.001% significant.*

PGP, Pollen germination percentage at 35 and 45°C, AWP^{-1}: adjusted weight per plant, DTA, days to anthesis; DTS, days to silking; ASI, anthesis silking interval; SD, standard deviation; Min, Minimum; Max, Maximum; σ^2 P, phenotypic variance; σ^2 G, genotype variance; H^2, heritability.

TABLE 2 | Correlation of traits under normal (above diagonal) and heat stress (below diagonal) conditions; diagonal shows the correlation between normal and heat stressed conditions for each trait.

Traits	AWP^{-1}	DTA	DTS	ASI	PGP (35°C)
AWP^{-1}	0.87	–0.64	–0.76	0.62	0.54
DTA	0.96	0.58	–0.1	–0.03	0.49
DTS	0.52	0.7	0.59	0.06	0.14
ASI	0.11	0.9	0.93	0.23	–0.09
PGP (45°C)	–0.3	–0.07	-0.1	0.14	0.55

AWP^{-1}, adjusted weight per plant; DTA, days to anthesis; DTS, days to silking; ASI, anthesis silking interval; PGP, Pollen germination percentage at 35 and 45°C.

Trait PGP35°C was associated with SNP CM007648.1-75060024 located on chromosome 2 which also had an R^2 value of 0.25. PGP45°C also showed association with SNP CM007650.1-89641182located on chromosome 1 with the highest R^2 value of 0.25. In the case of PGP ratio, SNP CM007649.1-225376365 located on chromosome 3 showed the highest R^2 value of 0.24. PGP45°C showed the lowest *p*-value of ($p \leq 10^{-8}$) with SNP CM007650.1-89641171 located on chromosome 1 shown in **Table 4** and **Supplementary Table S2**.

A total of 59 SNPs were associated with AWP^{-1} (**Supplementary Figures S1, S2**; **Supplementary Table S1**). Of these, only one SNP, CM007647.1-160363841 located on chromosome 1, was associated with AWP^{-1} at NARC in 2017 on planting date 1 (normal sowing). Fourteen SNPs located on chromosomes 2 (4), 3 (1), 4 (3), 5 (1), 8 (4), and 10 (1) showed significant association with AWP^{-1} at NARC in 2018 on planting date 1. Similarly, 4 SNPs located on chromosomes 1 (1), 2 (1), and 10 (2), and 5 SNPs located on chromosomes 1 (1), 3 (1), 6 (2), and 8 (1) were found to be associated with AWP^{-1} at Peshawar in 2017 date 1 and 2018 date 1, respectively. Four SNPs located on chromosomes 1 (3) and 8 (1), and 9 SNPs located on chromosomes 1 (2), 3 (1), 4 (1), 6 (2), 8 (1), 9 (1), and 10 (1) showed association with AWP^{-1} at Peshawar in 2017 date 2 and 2018 date 2, respectively. Out of 59 SNPs, 18 were highly significant ($p \leq 10^{-6}$) (**Table 3**; **Supplementary Table S2**).

For anthesis silking interval (ASI), 67 SNPs were associated (**Supplementary Figures S3, S4**; **Supplementary Table S2**). Two SNPs located on chromosomes 5 (1) and 8 (1), and 9 SNPs

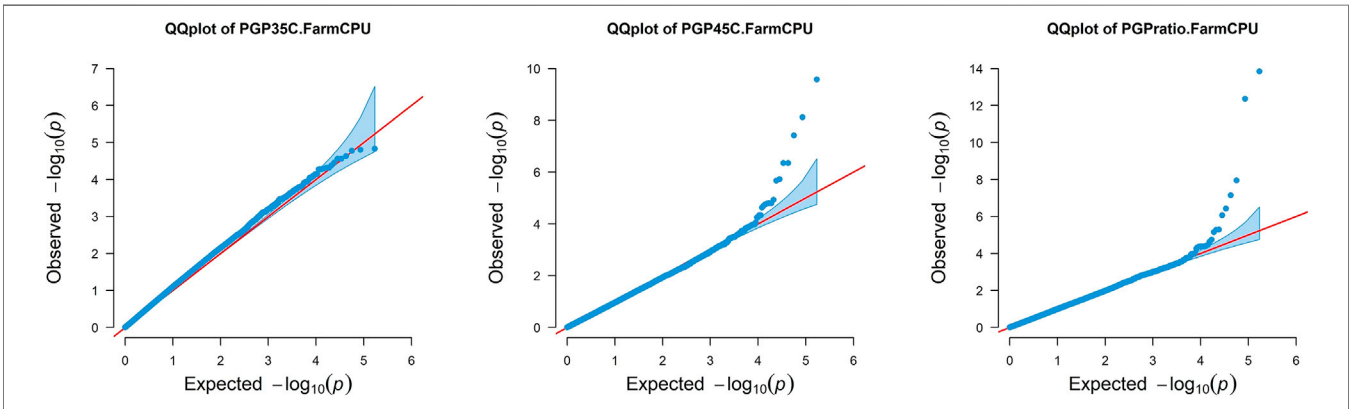


FIGURE 1 | Quantile-quantile (Q-Q) plots using Q + K FarmCPU model, showing comparative rise of expected $-\log_{10}$ *p*-values on x-axis against observed $-\log_{10}$ *p*-values on y-axis for each trait. PGP35°C: Pollen Germination percentage at 35°C; PGP45°C: Pollen Germination percentage at 45°C and PGP Ratio: Pollen Germination percentage ratio between 35 and 45°C.

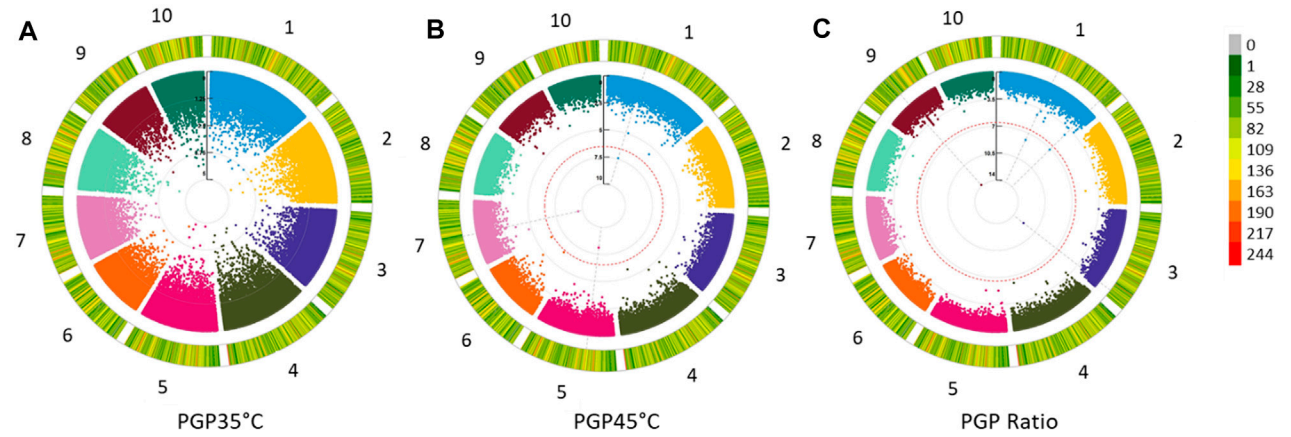


FIGURE 2 | Circular Manhattan plots revealing significant QTLs **(A)** PGP35°C: Pollen Germination percentage at 35°C; **(B)** PGP45°C: Pollen Germination percentage at 45°C and **(C)** PGP Ratio: Pollen Germination percentage ratio between 35 and 45°C in 122 maize inbred lines. The $-\log_{10} p$-values are used to plot against each of the 10 chromosomes represented in 10 different colors. The outer most circle indicates the marker density and side legend bar indicates the number of SNPs within a 1 MB window size.

TABLE 3 | Genome wide association studies (GWAS) results indicating SNPs associations with the target traits under various conditions using 122 (PGP) or 262 (AWP^{-1} and ASI) diverse maize inbred lines.

Traits	Total SNPs	1	2	3	4	5	6	7	8	9	10	*p* Value
PGP35°C	10		3	2	3	1					1	10^{-5}–10^{-6}
PGP45°C	9	4			1	3				1		10^{-6}–10^{-8}
PGP Ratio	9	1	2	3						3		10^{-5}–10^{-7}
AWP^{1}.NARC.2017.D1	1	1										10^{-5}–10^{-6}
AWP^{1}.NARC.2017.D2	9		4	1	2				1	1		10^{-6}–10^{-10}
AWP^{1}.NARC.2018.D1	14		4	1	3	1			4		1	10^{-6}–10^{-7}
AWP^{1}.NARC.2018.D2	13	5		1	1	2	2			2		10^{-6}–10^{-10}
AWP^{1}.Peshawer.2017.D1	4	1	1								2	10^{-5}–10^{-6}
AWP^{1}.Peshawer.2017.D2	4	3							1			10^{-5}–10^{-6}
AWP^{1}.Peshawer.2018.D1	5	1		1			2		1			10^{-5}–10^{-6}
AWP^{1}.Peshawer.2018.D2	9	2		1	1		2		1	1	1	10^{-5}–10^{-7}
ASI.NARC.2017.D1	2					1			1			10^{-5}–10^{-6}
ASI.NARC.2017.D2	11	4			1	1	1	1	2		1	10^{-6}–10^{-8}
ASI.NARC.2018.D1	9		1	1		4			2		1	10^{-6}–10^{-7}
ASI.NARC.2018.D2	3		2				1					10^{-6}–10^{-8}
ASI.Peshawer.2017.D1	8			3	1	2				1	1	10^{-6}–10^{-8}
ASI.Peshawer.2017.D2	13	5	2		3				1	1	1	10^{-6}–10^{-7}
ASI.Peshawer.2018.D1	11	3	1	1	5			1				10^{-6}–10^{-7}
ASI.Peshawer.2018.D2	10	1	1	3	1	1	1	1	1			10^{-6}–10^{-11}
Total	154	31	21	18	22	16	9	3	15	10	9	

located on chromosomes 2 (1), 3 (1), 5 (4), 8 (2), and 10 (1) showed significant association with ASI at NARC in 2017 for planting date 1, and NARC in 2018 data 1, respectively. At Peshawar in 2017 and 2018 date 1, 8 SNPs on chromosomes 3 (3), 4 (1), 5 (2), 9 (1), and 10 (1), and 11 SNPs on chromosomes 1 (3), 2 (1), 3 (1), 4 (1), 5 (1), 6 (1), 7 (1), and 8 (1), respectively, showed association with ASI. Under heat stress conditions (date 2) in Peshawar in 2017 and 2018, 13 SNPs located on chromosomes 1 (5), 2 (2), 4 (3), 8 (1), 9 (1), and 10 (1), and 10 SNPs on chromosomes 1 (1), 2 (1), 3 (3), 4 (1), 5 (1), 6 (1), 7 (1), and 1 (1), respectively, showed associations with ASI. Out of a total of 67 SNPs, 26 were found to be highly significant ($p \leq 10^{-6}$) (**Table 3**; **Supplementary Table S2**).

Cumulative Effect of Favorable Alleles on Traits

The number of favorable alleles for PGP traits in each inbred line was investigated. For PGP35°C, this number ranged between 6 and 95 favorable alleles for each inbred line (**Figures 3A,B**). Similarly, the number of favorable alleles per line for PGP45°C ranged between 2 and 98 (**Figures 3C,D**).

TABLE 4 | Common SNPs significantly associated within and between target traits in diverse maize inbred lines.

Trait	SNP	Chromosome	Position	Position (MB)	MAF	R^2	*p*-value
AWP[1].Peshawer.2017.D1	CM007647.1-19525844	1	19525844	19.525844	0.485	0.16	5.32E-06
AWP[1].Peshawer.2017.D2	CM007647.1-19525844	1	19525844	19.525844	0.23	0.23	4.46E-07
AWP[1].Peshawer.2018.D1	CM007650.1-26720973	3	26720973	26.720973	0.47	0.25	9.60E-06
AWP[1].Peshawer.2018.D2	CM007650.1-26720973	3	26720973	26.720973	0.49	0.24	7.38E-08
AWP[1].Peshawer.2018.D1	CM000786.4-148468397	10	148468397	148.468397	0.37	0.13	1.40E-07
AWP[1].Peshawer.2018.D2	CM000786.4-148468428	10	148468428	148.468428	0.43	0.23	6.65E-06
ASI.Peshawer.2018.D1	CM007647.1-258148529	1	258148529	258.148529	0.41	0.32	3.99E-06
ASI.Peshawer.2018.D2	CM007647.1-258192870	1	258192870	258.19287	0.455	0.12	1.08498E-06
ASI.Peshawer.2018.D2	CM007648.1-86615409	2	86615409	86.615409	0.47	0.24	2.13E-08
AWP[1].NARC.2017.D2	CM007648.1-86660602	2	86660602	86.660602	0.47	0.25	1.27E-07

Cumulative Effect of Favorable and Unfavorable Alleles on PGP35°C, PGP45°C and PGP, ratio.

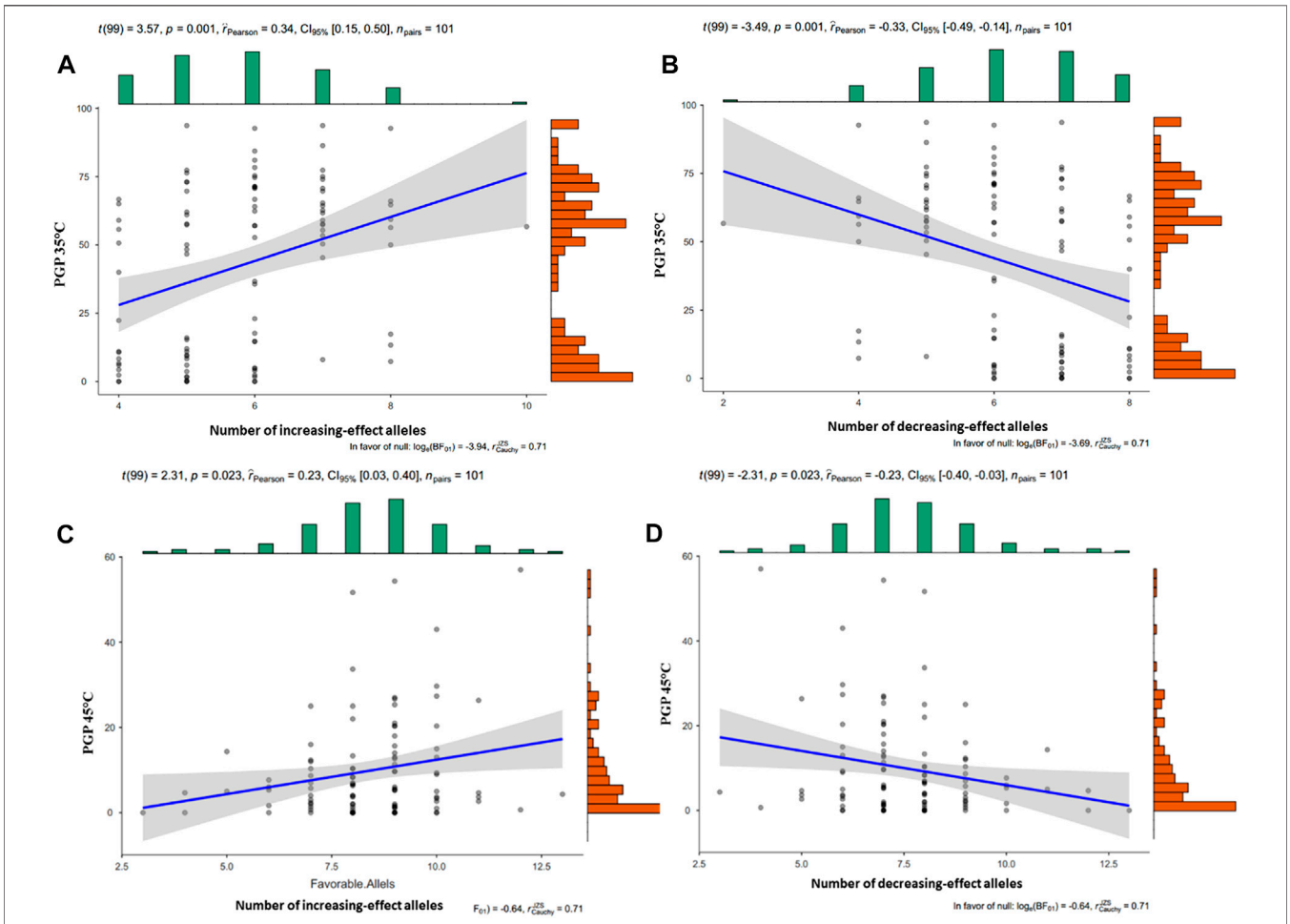


FIGURE 3 | Cummulative effect of the number of favorable **(A)** and unfavorable **(B)** alleles on PGP 35°C, and of the favorable **(C)** and unfavorable **(D)** alleles on PGP45°C. The dots represent the mean values. The scatter plot shows PGP on the y-axis and number of alles on the x-axis.

DISCUSSION

Agro-morphological screening showed that the diversity panel behaved differently under normal and heat stress conditions. The main trait under study was pollen viability, and as expected, it was adversely affected by heat stress. The pollen viability was estimated in terms of pollen germination percentage (PGP). At 35°C the mean value of PGP was 40.30% with minimum PGP 0% and maximum PGP 93.70%, which was significantly reduced by 9.70% mean value and 0 and 57% minimum and maximum PGP respectively at 45°C. This is due to the fact that pollen germination of maize is hindered above 38°C Sánchez et al.

(2014). The heat stress reduces the starch molecules in number and size within the pollen grain (Wang et al., 2019). Microspores exposed to heat stress during microsporogenesis leads to microspore abortion and pollen sterility (Begcy and Dresselhaus, 2018). Phenological traits like DTA, DTA and ASI were adversely affected by heat stress as significant reduction of mean DTA and DTS was observed in the diversity panel under heat stress as compared to normal sowing trials. The phenological development process of various crops is affected by high temperatures (Zhang et al., 2013).

In contrast, ASI is increased under heat stress compared to normal conditions; this was well explained by Cicchino et al. (2010). AWP^{-1} was the key indicator of plant response to heat stress in terms of major economic return. The mean AWP^{-1} was 37.5 g under normal conditions which was reduced to 35.50 g under heat stress conditions. Another important observation was that many of the inbred lines in the diversity panel produced baren plants under heat stress conditions. This is due to the fact that even a small increase (1.5°C) in temperature has a very significant negative effect on crop yield (Warland et al., 2006). An increase in average seasonal temperature by 1°C decreases the grain yield of cereal by 10 percent (Wang et al., 2012). Under heat stress conditions assimilatory capacity is reduced due to reduction in photosynthesis and increased maintenance respiration costs, resulting in productivity losses in wheat and maize crops (Reynolds et al., 2007).

In the case of maize, higher temperature (33–40°C) has a negative effect on light capture, harvest index, grain and biomass yield. At flowering, heat stress causes more yield reduction as compared to the grain filling stage (Zhang et al., 2013).

Broad-sense heritability estimates were moderate for traits under normal sowing trials, but were significantly reduced under heat stress conditions. This might be due to the fact that inbred lines were influenced by the adverse environment and affected the expression of traits. Our results were strongly in agreement with Noor et al., 2019 and Longmei et al., 2021. Correlation analysis revealed high positive correlation of AWP^{-1} with DTA (r = 0.96) and DTS (r = 0.52), and negative correlation with PGP 45°C (r = -0.3). DTA (r = 0.9) and DTS (r = 0.93) have high correlation with ASI under heat stress. PGP 35°C showed positive correlation with AWP^{-1} (r = 0.54), DTA (r = 0.49), and DTS (r = 0.14), and negative correlation with ASI (r = −0.09) under normal sowing conditions. The correlation analysis suggests that the traits under study might be influenced by the same genomic regions. Our studies are in coherence with previous studies (Xue et al., 2013; Wang et al., 2016; Zaidi et al., 2016). Further, overall time to flowering (DTA, DTS) seemed to be more indicative of yield (AWP^{-1}) than PGP under heat stress. It is interesting that the association of DTA and DTS with yield (AWP^{-1}) reverses directions under normal and heat stress conditions, perhaps emphasizing the importance of developmental stage on the impact of extreme weather conditions.

Association studies of PGP traits was carried out, and we considered all SNPs below 10^{-5} as candidates for future studies. Though this portion of the experiment was conducted on a reduced subset of 122 diverse inbred lines, 29 putative SNPs were found to be associated with the target traits at this significance level. Several studies have reported the association of traits with SNPs in maize using GWAS analysis (Xue et al., 2013; Wang et al., 2016; Carlos et al., 2019; Longmei et al., 2021). Because pollen germination percentage is a complex and time-consuming trait to score, this study was the first to our knowledge to map this trait in diverse exotic and indigenous maize inbred lines. Unique loci were identified that did not correspond with other physiological traits, introducing the possibility to investigate different genetic loci for heat stress adaptation traits in maize improvement.

GWAS of AWP^{-1} at two locations, years, and sowing dates in 275 lines revealed 59 MTAs. Of these, 24 were identified under normal sowing conditions (date 1), and 35 under heat stress conditions (date 2). Most of the SNPs that were identified were unique to a specific location and year, once again emphasizing the role of the environment and highlighting the need for validation of significant loci prior to their application in a breeding program.

In this study, three SNPs were identified as associated with AWP^{-1} in more than one condition, validating our results. In each case, these SNPs were associated with AWP^{-1} at Peshawar in 2017 on both date 1 and 2:

- SNP CM007647.1-19525844 on chromosome 1 at 19.5 MB
- SNP CM007650.1-26720973 on chromosome 3 at 26.7 MB
- SNP CM000786.4-148468397 on chromosome 10 at 148.5 MB

Though these SNPs were specific to location and year, they significantly impacted yield in both the normal and heat stress planting dates, indicating an overall contribution to fitness in that environment as opposed to a mechanism of heat stress tolerance specifically.

Association studies of ASI at two locations, years, and sowing dates identified a total of 67 SNPs distributed across nine chromosomes. Under normal sowing conditions, 30 MTAs were identified, while 37 SNPs showed significant association with ASI under heat stress conditions. Distribution of these traits across locations, years, and chromosomes are in **Supplementary Table S2**. As with AWP^{-1}, most SNPs identified were associated with only one environment. However, one SNP, CM007647.1-258148529 on chromosome 1 at 258 MB was found to be commonly associated with ASI at Peshawar in 2018 at both sowing dates, validating its association with ASI.

Finally, one SNP, CM007648.1-86615409 on chromosome 2 at 86.6 MB was found to be associated with two different traits: ASI at Peshawar in 2018, and AWP-1 at NARC in 2017, both under heat stress conditions. This locus may be of interest to pursue in future studies or for breeding heat tolerant varieties, as it was validated across traits, locations, and years as significantly impacting both yield and ASI in heat stress conditions specifically.

The results indicate that different regions on chromosomes are responsible for the expression of various traits under study. This might be due to a difference in gene regulatory mechanisms under heat stress conditions compared to normal conditions, or perhaps different subsets of a greater regulatory network become critical in

different specific environments. It is also possible that a more powerful experiment (more lines, environments, replications) may have identified additional shared loci. Still, SNPs were identified as significantly associated with the traits of interest, and a subset were validated. Based on these associations, candidate genes can also be identified and investigated for their roles in stress signaling. In previous studies, 206 significant SNPs were found associated with 115 candidate genes for drought tolerance and related traits (Wang et al., 2016). Sixty-seven SNPs were significantly associated with root structural traits as reported by Zaidi et al. (2016). Similarly, another study reported by Frey et al. (2015) revealed 607 heat responsive genes. Longmei et al. (2021) reported 46 SNPs associated with target traits under heat stress conditions in maize, and none of the SNPs were colocalized with multiple traits under heat stress conditions.

To understand the effect of combined favorable and unfavorable alleles on PGP35°C, PGP45°C, and PGP Ratio, the number of increasing effect alleles were investigated. Alleles corresponding to higher average PGP were considered favorable alleles, and those corresponding to lower average PGP were considered unfavorable alleles. The number of PGP 35°C increasing effect alleles ranged from 4 to 10, and the number of PGP 35°C decreasing effect alleles ranged between 2 and 8, **Figures 3A,B**. In the case of PGP 45°C, increasing effect alleles ranged from 3 to 13 and decreasing effect alleles ranged from 6 to 13, **Figures 3C,D**. The PGP ratio increasing and decreasing effect alleles ranged from 5 to 12 and 6–13, respectively.

CONCLUSION

Identification of maize genomic regions responsible for better crop performance under heat stress conditions is essential for the selection of inbred lines and development of heat tolerant hybrids. In this study, pollen viablility is considered as a major factor, reponsible for yield losses under heat stress conditions. Importantly, the exotic genotypes were not very adaptive, and the local germplasm showed resiliance aginst heat stress conditions. There was significant reduction in PGP from 40.30% at 35°C to 9.70% at 45°C. This reduction in PGP showed negative correlation with yield contributing traits as well. GWAS indicated 14 significant SNPs associated with PGP35°C, PGP45°C and PGP ratio. Further work in identification of candidate genes using these significant SNPs is required to fully elucidate the role of these genomic regions in heat stress tolerance.

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

ZA: Conducted experiment, collected data, data analysis, initial draft writeup; MK: Data cleaning, analysis and writeup; AG: DNA extraction, supervised the trials and helped in paper review; MS: Supervised the trials and helped in paper review; GR: Supervised the trials and helped in paper review; RR: Supervised the trials and helped in paper review; TM: Assisted in data recoding at both locations; AT: Lead for GBS, data cleaning and analysis, GWAS analysis.

FUNDING

This work was supported in part by: the maize, sorghum, and millet program, CSI, NARC for trails; the Higher Education Commission, Pakistan travel award to ZA; and startup funds to AT for sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.819849/full#supplementary-material

REFERENCES

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide Association Study of 107 Phenotypes in *Arabidopsis thaliana* Inbred Lines. *Nature* 465, 627–631. doi:10.1038/nature08800

Begcy, K., and Dresselhaus, T. (2018). Epigenetic Responses to Abiotic Stresses during Reproductive Development in Cereals. *Plant Reprod.* 31, 343–355. doi:10.1007/s00497-018-0343-4

Browning, B. L., and Browning, S. R. (2016). Genotype Imputation with Millions of Reference Samples. *Am. J. Hum. Genet.* 98, 116–126. doi:10.1016/j.ajhg.2015.11.020

Browning, S. R., and Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies by Use of Localized Haplotype Clustering. *Am. J. Hum. Genet.* 81 (5), 1084–1097. doi:10.1086/521987

Cicchino, M., Edreira, J. I. R., Uribelarrea, M., and Otegui, M. E. (2010). Heat Stress in Field-Grown maize: Response of Physiological Determinants of Grain Yield. *Crop Sci.* 50 (4), 1438–1448. doi:10.2135/cropsci2009.10.0574

CIMMYT/IBPGR (1991). Descriptors for Maize. Rome, 100p. Available at: https://www.bioversityinternational.org/e-library/publications/detail/descriptors-for maizedescriptores-para-maizdescripteurs-pour-le-mais/.

Deryng, D., Conway, D., Ramankutty, N., Price, J., and Warren, R. (2014). Global Crop Yield Response to Extreme Heat Stress under Multiple Climate Change Futures. *Environ. Res. Lett.* 9, 034011. doi:10.1088/1748-9326/9/3/034011

Dresselhaus, T., Lausser, A., and Márton, M. L. (2011). Using maize as a Model to Study Pollen Tube Growth and Guidance, Cross-Incompatibility and Sperm Delivery in Grasses. *Ann. Bot.* 108 (4), 727–737. doi:10.1093/aob/mcr017

FAOSTAT (2020). Faostat Dataset. Retrieved from .

Finney, D. J. (1964). *Statistical Method in Biological Assay*. New York: Hafner, 21–57. doi:10.1002/bimj.4710210714

Frey, F. P., Claude, U., Bruno, H., Richard, R., and Benjamin, S. (2015). Genome-wide Expression Profiling and Phenotypic Evaluation of European maize Inbreds at Seedling Stage in Response to Heat Stress. *BMC Genom.* 16, 123. doi:10.1186/s12864-015-1282-1

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide Association Studies of 14 Agronomic Traits in rice Landraces. *Nat. Genet.* 42, 961–967. doi:10.1038/ng.695

IPCC (2021). "Summary for Policymakers," in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.Editors Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, S., Berger, N., et al. In Press.

Kulwal, P. L., Thudi, M., and Varshney, R. K. (2014). ""Genomics Interventions in Crop Breeding for Sustainable Agriculture" in Encyclopedia of Sustainability,". Editor R. A Meyers (New York, NY, USA: Sci Tech), 2527–2540.

Li, Q., Li, L., Yang, X., Warburton, M. L., Bai, G., Dai, J., et al. (2010). Relationship, Evolutionary Fate and Function of Two maize Co-orthologs of rice GW2 Associated with Kernel Size and Weightsize and Weight. *BMC Plant Biol.* 10 (10), 143. doi:10.1186/1471-2229-10-143

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-wide Association Studies. *Plos Genet.* 12, e1005767. doi:10.1371/journal.pgen.1005767

Lobell, D. B., Hammer, G. L., Chenu, K., Zheng, B., McLean, G., and Chapman, S. C. (2015). The Shifting Influence of Drought and Heat Stress for Crops in Northeast Australia. *Glob. Change Biol.* 21, 4115–4127. doi:10.1111/gcb.13022

Longmei, N., Gill, G. K., Zaidi, P. H., Kumar, R., Nair, S. K., Hindu, V., et al. (2021). Genome Wide Association Mapping for Heat Tolerance in Sub-tropical maize. *BMC Genomics* 22, 154. doi:10.1186/s12864-021-07463-y

Maldonado, C., Mora, F., Scapim, C. A., and Coan, M. (2019). Genome-wide Haplotype-Based Association Analysis of Key Traits of Plant Lodging and Architecture of maize Identifies Major Determinants for Leaf Angle: hapLA4. *PLoS ONE* 14, e0212925. doi:10.1371/journal.pone.0212925

Manicacci, D., Camus-kulandaivelu, L., Fourmann, M., Arar, C., Barrault, S., Rousselet, A., et al. (2009). Epistatic Interactions betweenOpaque2Transcriptional Activator and its Target GeneCyPPDK1Control Kernel Trait Variation in Maize Transcriptional Activator and its Target Gene CyPPDK1 Control Kernel Traitvariation in maize. *Plant Physiol.150* 150, 506–520. doi:10.1104/pp.108.131888

Mwadzingeni, L., Shimelis, H., Rees, D. J. G., and Tsilo, T. J. (2017). Genome-wide Association Analysis of Agronomic Traits in Wheat under Drought-Stressed and Non-stressed Conditions. *Plos One* 12, e0171692. doi:10.1371/journal.pone.0171692

Nadeem, M. A., Habyarimana, E., Karaköy, T., and Baloch, F. S. (2021). Genetic Dissection of Days to Flowering via Genome-wide Association Studies in Turkish Common Bean Germplasm. *Physiol. Mol. Biol. Plants* 27 (7), 1609–1622. doi:10.1007/s12298-021-01029-8

Noor, J. J., Vinayan, M. T., Umar, S., Devi, P., Iqbal, M., Seetharam, K., et al. (2019). Morpho-physiological Traits Associated with Heat Stress Tolerance in Tropical maize (*Zea mays* L.) at Reproductive Stage. *Aust. J Crop Sci.* 13 (04), 536–545. doi:10.21475/ajcs.19.13.04

Ott, A., Liu, S., Schnable, J. C., Yeh, C.-T. E., Wang, K.-S., and Schnable, P. S. (2017). tGBS Genotyping-By-Sequencing Enables Reliable Genotyping of Heterozygous Loci. *Nucleic Acids Res.* 45, e178. doi:10.1093/nar/gkx853

R Core Team (2018). "R: A Language and Environment for Statistical Computing," in *R Foundation for Statistical Computing* (Vienna: Austria). Available online at https://www.R-project.org/.

Reynolds, M. P., Pierre, C. S., Saad, A. S. I., Vargas, M., and Condon, A. G. (2007). Evaluating Potential Genetic Gains in Wheat Associated with Stress-Adaptive Trait Expression in Elite Genetic Resources under Drought and Heat Stress. *Crop Sci.* 47, S–172. doi:10.2135/cropsci2007.10.0022IPBS

Salvi, S., Sponza, G., Morgante, M., Tomes, D., Niu, X., Fengler, K. A., et al. (2007). Conserved Noncoding Genomic Sequences Associated with a Flowering-Time Quantitative Trait Locus in maize. *Proc. Natl. Acad. Sci.* 104, 11376–11381. doi:10.1073/pnas.0704145104

Sánchez, B., Rasmussen, A., and Porter, J. R. (2014). Temperatures and the Growth and Development of maize and rice: a Review. *Glob. Change Biol.* 20, 408–417. doi:10.1111/gcb.12389

Sanchez, D. L., Liu, S., Ibrahim, R., Blanco, M., and Lübberstedt, T. (2018). Genome-wide Association Studies of Doubled Haploid Exotic Introgression Lines for Root System Architecture Traits in maize (*Zea mays* L.). *Plant Sci.* 268, 30–38. doi:10.1016/j.plantsci.2017.12.004

Thirunavukkarasu, N., Hossain, F., Arora, K., Sharma, R., Shiriga, K., Mittal, S., et al. (2014). Functional Mechanisms of Drought Tolerance in Subtropical maize (Zea mays L.) Identified Using Genome-wide Association Mapping. *BMC Genomics* 15, 1182. doi:10.1186/1471-2164-15-1182

Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., et al. (2011). Genome-wide Association Study of Leaf Architecture in the maize Nested Association Mapping Population. *Nat. Genet.* 43, 159–162. doi:10.1038/ng.746

Wang, H., Xu, X., Vieira, F. G., Xiao, Y., Li, Z., Wang, J., et al. (2016). The Power of Inbreeding: NGS-Based GWAS of rice Reveals Convergent Evolution during rice Domestication. *Mol. Plant* 9, 975–985. doi:10.1016/j.molp.2016.04.018

Wang, X., Cai, J., Liu, F., Jin, M., Yu, H., Jiang, D., et al. (2012). Pre-anthesis High Temperature Acclimation Alleviates the Negative Effects of post-anthesis Heat Stress on Stem Stored Carbohydrates Remobilization and Grain Starch Accumulation in Wheat. *J. Cereal Sci.* 55, 331–336. doi:10.1016/j.jcs.2012.01.004

Wang, Y., Tao, H., Tian, B., Sheng, D., Xu, C., Zhou, H., et al. (2019). Flowering Dynamics, Pollen, and Pistil Contribution to Grain Yield in Response to High Temperature during maize Flowering. *Environ. Exp. Bot.* 158, 80–88. doi:10.1016/j.envexpbot.2018.11.007

Warland, J., McKeown, A. W., and McDonald, M. R. (2006). Impact of High Air Temperatures on Brassicacae Crops in Southern Ontario. *Can. J. Plant Sci.* 86, 1209–1215. doi:10.17582/journal.pjar/2021/34.3.479.48610.4141/p05-067

Xue, Y., Warburton, M. L., Sawkins, M., Zhang, X., Setter, T., Xu, Y., et al. (2013). Genome-wide Association Analysis for Nine Agronomic Traits in maize under Well-Watered and Water-Stressed Conditions. *Theor. Appl. Genet.* 126, 2587–2596. doi:10.1007/s00122-0130215810.1007/s00122-013-2158-x

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.-c., Hu, L., et al. (2016). Genome-wide Association Study Using Whole-Genome Sequencing Rapidly Identifies New Genes Influencing Agronomic Traits in rice. *Nat. Genet.* 48, 927–934. doi:10.1038/ng.3596

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: A Memory-Efficient, Visualization-Enhanced, and Parallel-Accelerated Tool for Genome-wide Association Study. *Genomics, Proteomics & Bioinformatics* S1672–0229 (21), 00050–00054. doi:10.1016/j.gpb.2020.10.007(

Zaidi, P. H., Seetharam, K., Krishna, G., Krishnamurthy, L., Gajanan, S., Babu, R., et al. (2016). Genomic Regions Associated with Root Traits under Drought Stress in Tropical Maize (Zea mays L.). *PLoS ONE* 11, e0164340. doi:10.1371/journal.pone.0164340

Zhang, X., Cai, J., Wollenweber, B., Liu, F., Dai, T., Cao, W., et al. (2013). Multiple Heat and Drought Events Affect Grain Yield and Accumulations of High Molecular Weight Glutenin Subunits and Glutenin Macropolymers in Wheat. *J. Cereal Sci.* 57, 134–140. doi:10.1016/j.jcs.2012.10.010

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., et al. (2017). Temperature Increase Reduces Global Yields of Major Crops in Four Independent Estimates. *Proc. Natl. Acad. Sci. USA* 114, 9326–9331. doi:10.1073/pnas.1701762114

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 March 2022
doi: 10.3389/fgene.2022.830644

Genome-wide Association Study for Starch Pasting Properties in Chinese Spring Wheat

Yousheng Tian[1,2†], Wei Sang[3†], Pengpeng Liu[3], Jindong Liu[4], Jishan Xiang[3], Fengjuan Cui[3], Hongjun Xu[3], Xinnian Han[3], Yingbin Nie[3], Dezhen Kong[3], Weihua Li[1] and Peiyuan Mu[3]**

[1]The Key Laboratory of the Oasis Ecological Agriculture, College of Agriculture, Shihezi University, Shihezi, China, [2]Department of Administrative Management, Xinjiang Academy of Agri-reclamation Sciences, Shihezi, China, [3]Institute of Crop Science, Xinjiang Academy of Agri-reclamation Sciences/Key Lab of Xinjiang Production and Construction Corps for Cereal Quality Research and Genetic Improvement, Shihezi, China, [4]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

In order to understand the genetic basis of starch pasting viscosity characteristics of Chinese spring wheat, we assessed the genetic variation of RVA parameters determined by the Rapid Visco Analyser in a panel of 192 Chinese spring wheat accessions grown in Er'shi, Shihezi and Zhaosu during 2012 and 2013 cropping seasons. A genome-wide association study with 47,362 single nucleotide polymorphism (SNP) markers was conducted to detect marker-trait associations using mixed linear model. Phenotypic variations of RVA parameters ranged from 1.6 to 30.7% and broad-sense heritabilities ranged from 0.62 to 0.91. Forty-one SNP markers at 25 loci were significantly associated with seven RVA traits in at least two environments; among these, 20 SNPs were located in coding sequences (CDS) of 18 annotation genes, which can lead to discovering novel genes underpinning starch gelatinization in spring wheat. Haplotype analysis revealed one block for breakdown (BD) on chromosome 3B and two blocks for pasting temperature (T) on chromosome 7B. Cultivars with superior haplotypes at these loci showed better starch pasting viscosity than the average of all cultivars surveyed. The identified loci and associated markers provide valuable sources for future functional characterization and genetic improvement of starch quality in wheat.

Keywords: candidate genes, GWAS, haplotype analysis, RVA parameters, *Triticum aestivum*

OPEN ACCESS

Edited by:
Awais Rasheed,
Quaid-i-Azam University, Pakistan

Reviewed by:
Deepmala Sehgal,
International Maize and Wheat Improvement Center, Mexico
Anjan Hazra,
University of Calcutta, India

***Correspondence:**
Weihua Li
lwh_agr@shzu.edu.cn
Peiyuan Mu
mupy@163.com

[†]*These authors share first authorship*

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *07 December 2021*
Accepted: *25 February 2022*
Published: *25 March 2022*

Citation:
Tian Y, Sang W, Liu P, Liu J, Xiang J, Cui F, Xu H, Han X, Nie Y, Kong D, Li W and Mu P (2022) Genome-wide Association Study for Starch Pasting Properties in Chinese Spring Wheat. Front. Genet. 13:830644. doi: 10.3389/fgene.2022.830644

INTRODUCTION

Wheat (*Triticum aestivum* L.) is one of the most important staple food crops worldwide. With the improvement of living standards, people pay more attention to the quality of end-use products of wheat. Improvement of quality traits has become a major objective in wheat breeding (Kong et al., 2013). The gelatinization characteristic of wheat flour is a main index to evaluate the processing quality of food products (Crosbie, 1991; Panozzo et al., 1993; Liu et al., 2003; Kaur et al., 2016; Amiri et al., 2018; Moiraghi et al., 2019). Rapid Visco Analyser (RVA) profile has proven useful in wheat breeding programs to assess the eating and cooking quality of wheat (Konik et al., 1992; He et al., 2003; He et al., 2004; Zhang et al., 2004; Zhang et al., 2005; León et al., 2006). Starch pasting viscosity is controlled by multiple genes and often influenced by environments (Zhang et al., 2009), the traditional methods for assessing RVA parameters are laborious and need expensive equipment. The use of molecular markers for an indirect marker-assisted selection (MAS) is effective in selection for quality traits in breeding process. Therefore, it is important to study the genetic basis of starch gelatinization for wheat quality improvement using MAS.

Previous studies have been heading for localizing genes and QTL for starch gelatinization characteristics to expedite MAS in wheat breeding (Udall et al., 1999; Araki et al., 2000; Deng et al., 2014). However, QTL identified by bi-parent populations cannot explain the variation of starch gelatinization characteristics in complex genetic population due to relatively simple genetic background and lower allele variability. The genome-wide association study (GWAS) can identify genomic regions associated with variations in a given trait by combining phenotypic with genotypic data (AL-Maskri et al., 2012). Compared with conventional bi-parental QTL mapping, GWAS has the advantage of surveying a larger range of allelic variations and avoiding a time-consuming process for establishing a customized mapping population (Tadesse et al., 2015). Because of having more genetic diversity and historical recombination of alleles among associated panels, GWAS can get more accurate results (Muhu-Din Ahmed et al., 2020). With development of sequencing technology, high-density wheat SNP arrays have been developed, which combined with GWAS were widely used to identify genetic loci for important traits in hexaploid wheat (Sukumaran et al., 2015; Sun et al., 2017; Rimbert et al., 2018; Zhang et al., 2018; Yan et al., 2019; Lv et al., 2020; Shi et al., 2020), especially for quality-related traits, such as grain protein content, wet gluten content, grain starch content, SDS-sedimentation volume, dough rheological properties, and so on (Muqaddasi et al., 2020; Yang et al., 2020; Muhu-Din Ahmed et al., 2020), but none for starch pasting properties.

Herein, we performed a GWAS to identify genetic loci for RVA parameters using the wheat 90K SNP array and multi-environment field data in a panel of 192 Chinese spring wheat genotypes. Markers significantly associated with RVA parameters and candidate genes were identified. The results of this study can enhance our understanding of the genetic basis of wheat starch gelatinization and provide valuable information for MAS in wheat breeding.

MATERIALS AND METHODS

Plant Materials and Field Trials

A set of 192 genetically diverse spring wheat accessions, representing cultivars and breeding lines from different provinces of China, was grown at Er'shi, Shihezi and Zhaosu in Xinjiang province in randomized complete blocks with three replications during 2012 and 2013 cropping seasons (hereafter referred as 2012_ES, 2012_SHZ, 2012_ZS, 2013_ES, 2013_SHZ and 2013_ZS, respectively). Each genotype was sown in ten rows, with a row length of 3 m, a row-to-row distance of 25 cm and plant-to-plant distance of 10 cm. There were differences in climate and soil conditions among Er'shi, Shihezi and Zhaosu, and different temperatures between the years 2012 and 2013. Planting and harvest dates and trial management varied according to the recommendations of each location. The accessions were harvested at maturity and cleaned prior to quality test.

Milling

Flour milling was performed in a mill (MLU202, Wuxi, China) to flour extraction rates of around 65%. Prior to milling, the hard, medium hard (mixtures of hard and soft wheat) and soft wheats were tempered overnight to moisture contents of around 16, 15, and 14%, respectively.

Measurements of RVA Parameters

Pasting properties of flour were determined with a Rapid Visco Analyser (RVA-Techmaster, Newport Scientific, Australia). The 3 g flour was suspended in 25 ml of distilled water before the solution was placed inside RVA instrument. The programs of temperature in the following order: held at 50°C for 60 s, heated from 50 to 95°C at a rate of 1°C/5 s and held at 95°C for 150 s, then cooled to 50°C at a rate of 1°C/5 s and held at 50°C for 120 s. RVA parameters including peak viscosity (PV), trough viscosity (TV), breakdown (BD), final viscosity (FV), setback (SB), peak time (PT) and pasting temperature (T) were determined.

Genotyping

Fresh leaf samples were collected from 10-day old seedlings and sent to the CapitalBio Technology company[1] in Beijing for genotyping with the high-density illumina wheat 90K SNP array. After excluding the low-quality SNP markers with minor allele frequency (MAF) ≤0.02 and missing data ≥10%, 47,362 SNPs were used for GWAS. All SNP markers were anchored on the wheat genome (IWGSC RefSeq v1.0) using BLASTN by 50 bp SNP flanking sequences on both sides of the SNP.

Structure Analysis

The Bayesian clustering technique was used with 3400 SNP markers to classify groups of genotypically same individuals using the statistical software STRUCTURE v.2.3.4 (Pritchard et al., 2000). Burn-in iterations of 10^4 cycles were used, followed by a simulation runs of 10^5 cycles with an admixture model. The K values of 1–10 and 3 independent runs were selected to attain reliable results. Web-based analysis "Structure Harvester v0.6.93"[2] was applied to obtain maximum value or peak of "K" for validation to understand the STRUCTURE v.2.3.4 results using ad-hoc techniques (Earl and VonHoldt, 2011). ΔK was plotted against the number of sub-group K following Evanno et al. (2005).

A principal component analysis (PCA) using filtered SNPs was performed with the TASSEL v.5.2.43 and the first three PCA values were plotted in three dimensions.

Genome-Wide Association Study

Phenotypic data in different environments and best linear unbiased prediction (BLUP) values were analyzed, respectively, for association analysis to identify the marker-trait associations (MTAs) employing mixed linear model (MLM) in TASSEL v.5.2.43 (Bradbury et al., 2007). The MLM option requires population structure (Q-matrix) and kinship matrix (K-matrix) as covariates for GWAS to avoid false positives. The Q-matrix was generated through STRUCTURE v.2.3.4, whereas the K-matrix was generated by TASSEL v.5.2.43. The Bonferroni multiple testing correction was used to identify

[1]https://www.capitalbiotech.com/.

[2]http://taylor0.biology.ucla.edu/structureHarvester/.

significant markers. Significant SNPs associated with RVA traits were claimed when the significance test reached $p < 4.18E\text{-}4$ (20/47,362 = 0.000418). Linkage disequilibrium (LD) was calculated by TASSEL 5.0 with the markers whose positions were known. The LD decay plot was generated by intra-chromosomal r^2 and base pair distance using R package ggplot2.

Mapping SNPs and Prediction of Candidate Genes

The Chinese spring reference genome (IWGSC RefSeq v1.0) and gene annotations in GFF3 format were retrieved from the Ensemble database release 44[3]. The SNP marker sequences were mapped to the wheat genome using BLASTN program with a stringent E-value of 0.0001. For each SNP only the best scoring hit was retained. Each aligned genomic position was annotated into 5′-UTR, 3′-UTR, CDS, intron and intergenic region according to the genomic regions provided in the GFF3 file. The intergenic region was the genomic region with no annotated genes. The protein functions of candidate genes were predicted in the Uniprot Protein database[4].

Haplotype Analysis

For the genomic regions harboring SNPs with -log10 (p) above the threshold, and the phenotypic values of accessions with different alleles reached significant level ($p < 0.05$) in multi-environment, haplotype analysis was carried out by Haploview version 4.2 software (Barrett et al., 2005), and candidate loci were determined by testing the significant differences on phenotypes among major haplotypes through analysis of variance (ANOVA).

Statistical Analysis

ANOVA and correlation analysis were carried out using SPSS version 22.0. The coefficient of variation (CV) was calculated by dividing the standard deviation by the average of trait values. The BLUP values of all traits over 2 years across environments were calculated by the R package Lme4 (Bates et al., 2015). Broad-sense heritabilities (h^2) of RVA traits were calculated as $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{gl}^2/l + \sigma_{gs}^2/s + \sigma_{gls}^2/ls)$, where σ_g^2, σ_{gl}^2, σ_{gs}^2, σ_{gls}^2 variances for genotypic, genotype by location interaction, genotype by season interaction and genotype by location by season interaction, respectively, whereas l and s were the numbers of locations and seasons, respectively. SNP density plots, Manhattan and quantile-quantile (Q-Q) plots were generated in R package CMplot while histograms were performed in Origin 8.0.

RESULTS

Phenotypic Distributions and Correlations of RVA Parameters

The variation coefficients of PV, TV, BD, FV, SB, PT and T were in the ranges of 8.4–13.6%, 7.0–12.9%, 20.4–30.7%, 5.4–12.6%, 7.6–12.9%, 1.6–3.0%, 1.9–13.6%, respectively, across environments, and h^2 estimates were 0.91, 0.77, 0.86, 0.89, 0.82, 0.62 and 0.85, respectively (**Supplementary Table S1**). Except the T, the frequency distributions of BLUP values of RVA parameters were nearly symmetrically distributed (**Supplementary Figure S1**). RVA parameters exhibited wide phenotypic variations and high broad-sense heritabilities, which were imperative for an efficient GWAS.

TABLE 1 | Pearson's correlation coefficients (r) between RVA parameters.

Trait	PV	TV	BD	FV	SB	PT
TV	0.68[a]					
BD	0.81[a]	0.13				
FV	0.63[a]	0.94[a]	0.10			
SB	0.36[a]	0.72[a]	−0.09	0.89[a]		
PT	0.03	0.63[a]	−0.43[a]	0.48[a]	0.31[a]	
T	−0.24[a]	−0.14	−0.22[a]	−0.13	−0.04	0.01

PV, peak viscosity; TV, trough viscosity; BD, breakdown; FV, final viscosity; SB, setback; PT, peak time; T, pasting temperature.
[a]*Indicates significance levels at* p < *0.01.*

Based on the BLUP values across six environments, PV was significantly and positively correlated with TV (r = 0.68**), BD (r = 0.81**) and FV (r = 0.63**); TV was significantly and positively correlated with FV (r = 0.94**), SB (r = 0.72**) and PT (r = 0.63**); FV was significantly and positively correlated with SB (r = 0.89**) and PT (r = 0.48**); BD was significantly and negatively correlated with PT (r = −0.43**, **Table 1**).

Genome-wide Associations

Based on whole-genome genotyping data, PCA showed that the association panel could be divided into two subgroups (**Figure 1A**), the same with the results calculated by the STRUCTURE software with 3,400 markers, in which the peak of the broken line graph was observed at k = 2 (**Figure 1B**), indicating the association panel can be divided into two sub-populations. After filtering low-quality markers, 47,362 SNPs were used for GWAS analysis with BLUP values and individual data in six environments. The number of SNPs within 1 Mb window size in chromosomes indicated that these SNPs were almost evenly distributed across the whole genome (**Figure 1C**). To avoid multiple significances within individual LD blocks, the support interval was determined when the decay distance of LD reached $r^2 = 0.2$, and LD was estimated to decay at about 7 Mb for whole genome (**Supplementary Figure S2**). Manhattan and Q-Q plots based on BLUP values identified 92 significant SNPs at 51 loci (**Figure 2**; **Supplementary Table S2**), of which 13, 7, 5, 16, 10, 7 and 2 were significantly associated with PV, TV, BD, FV, SB, PT and T, respectively (**Figure 1D**). Phenotypic variations explained by these SNPs ranged from 6.9 to 11.1% for PV, 6.6–8.4% for TV, 7.2–8.8% for BD, 5.9–11.9% for FV, 5.2–7.3% for SB, 6.7–7.7% for PT, and 6.8–9.1% for T (**Supplementary Table S2**).

Significant and Stable SNPs Associated With RVA Parameters and Candidate Genes

Forty-one SNPs at 25 loci were stably detected in at least two environments (BLUP is considered as an environment,

[3]ftp://ftp.ensemblgenomes.org/pub/plants/release-44.
[4]https://www.uniprot.org/.

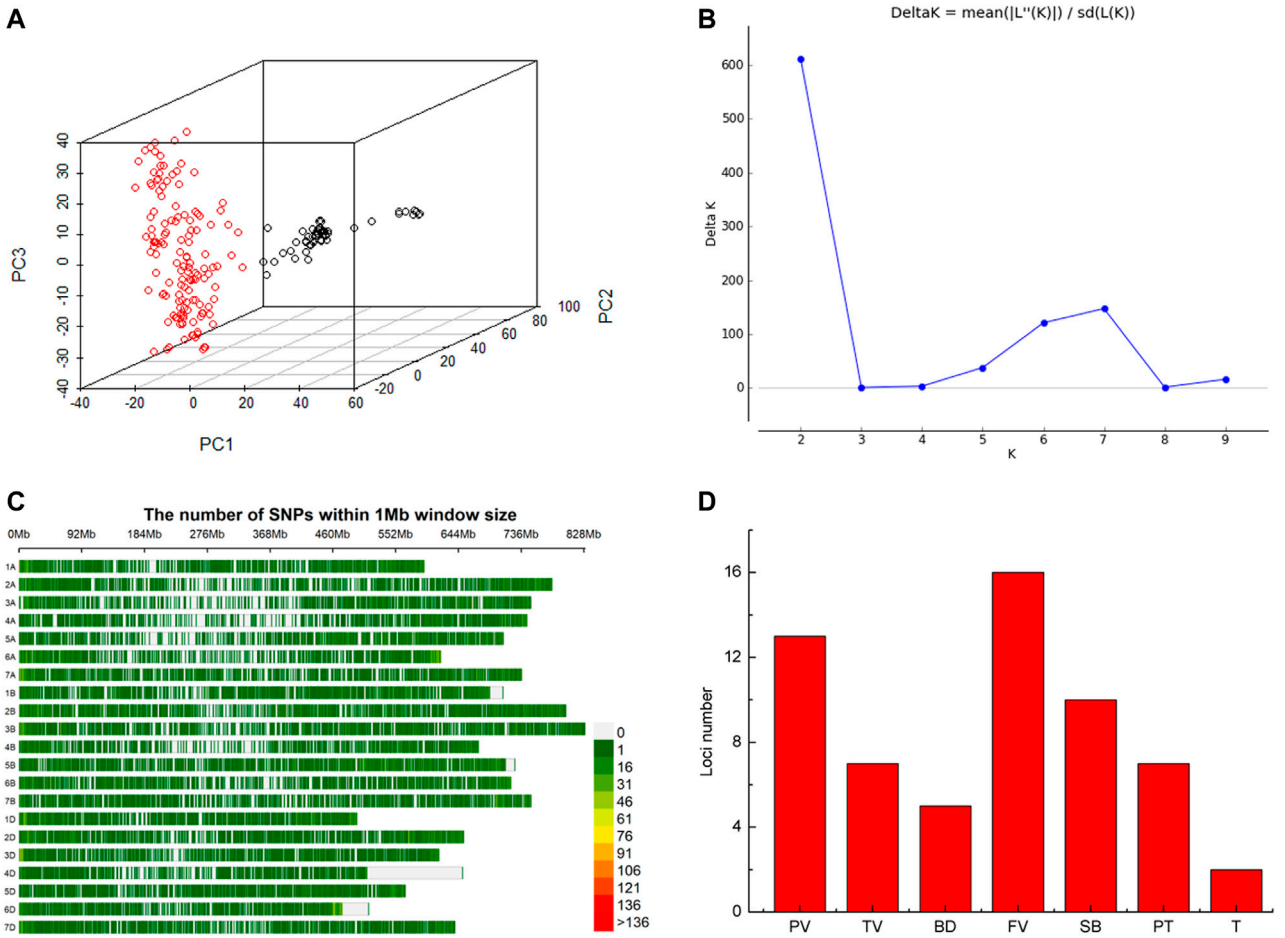


FIGURE 1 | Structure of the association panel of 192 Chinese spring wheat varieties for genome-wide association study (GWAS), distribution of SNPs and significant loci revealed by GWAS. **(A)** Structure of association panel. **(B)** Broken line of delta K. **(C)** Distribution of SNPs on chromosomes. **(D)** Histogram plot of loci number significantly associated with RVA parameters *via* GWAS based on BLUP values.

Supplementary Table S3), and four SNPs (*GENE-4428_113, BobWhite_c13098_670, IAAV4275, Kukri_rep_c69088_774*) on chromosomes 7A (1), 7B (1) and 7D (2) at 534, 489 and 463 Mb, respectively, were significantly associated with both PV and TV, while one SNP (*Excalibur_c9183_1397*) on chromosome 7D at 13 Mb was significantly associated with both PV and BD. Thirty-four SNP markers associated with RVA parameters were mapped in the annotated genes; 20 SNPs among these were located in the CDS of 18 annotation genes, which were considered as candidate genes (**Supplementary Table S4**).

Haplotype Analysis of Genetic Loci Related to RVA Parameters

To test the effect of different genotypes on RVA parameters, stable SNPs were selected to group the populations according to their genotypes, and *t*-test was used to test the significance of genotypic effects on the traits. Three SNPs (*BS00023017_51, Excalibur_c10124_361, Kukri_c4560_110*) revealed significant differences ($p < 0.05$) of the traits between two alleles in at least four environments (**Supplementary Table S5**), indicating that these loci had a great influence on phenotypic variation.

The haplotype analysis for regions harboring the SNP marker (*BS00023017_51*) associated with BD showed that the 776–782 Mb interval on chromosome 3B had an 862-kb block, and six SNPs including *BS00023017_51* were clustered in the block (**Figures 3A,D**). Comparison of BD values indicated that 37 cultivars with the TT/AC/CC/CC/GG/AA (*BS00024883_51/BS00023017_51/IAAV8892/RFL_Contig2578_862/wsnp_Ex_c33879_42293206/wsnp_CAP11_c59_99263*) haplotypes showed significantly higher BD ($p < 0.05$) than the average of all cultivars surveyed in four environments (2012_SHZ, 2013_ES, 2013_SHZ, 2013_ZS, **Figures 3B,C**). Haplotype analysis was also conducted for region harboring the SNPs (*Excalibur_c10124_361, Kukri_c4560_110*) associated with T, and two blocks in 442–454 Mb interval on chromosome 7B were detected. Two cultivars with CC/AA/CC (*CAP12_c8025_110/IAAV 2037/Excalibur_c49622_60*), 17 cultivars with AA/TT/GG/GG (*wsnp_Ex_c64815_63464750/*

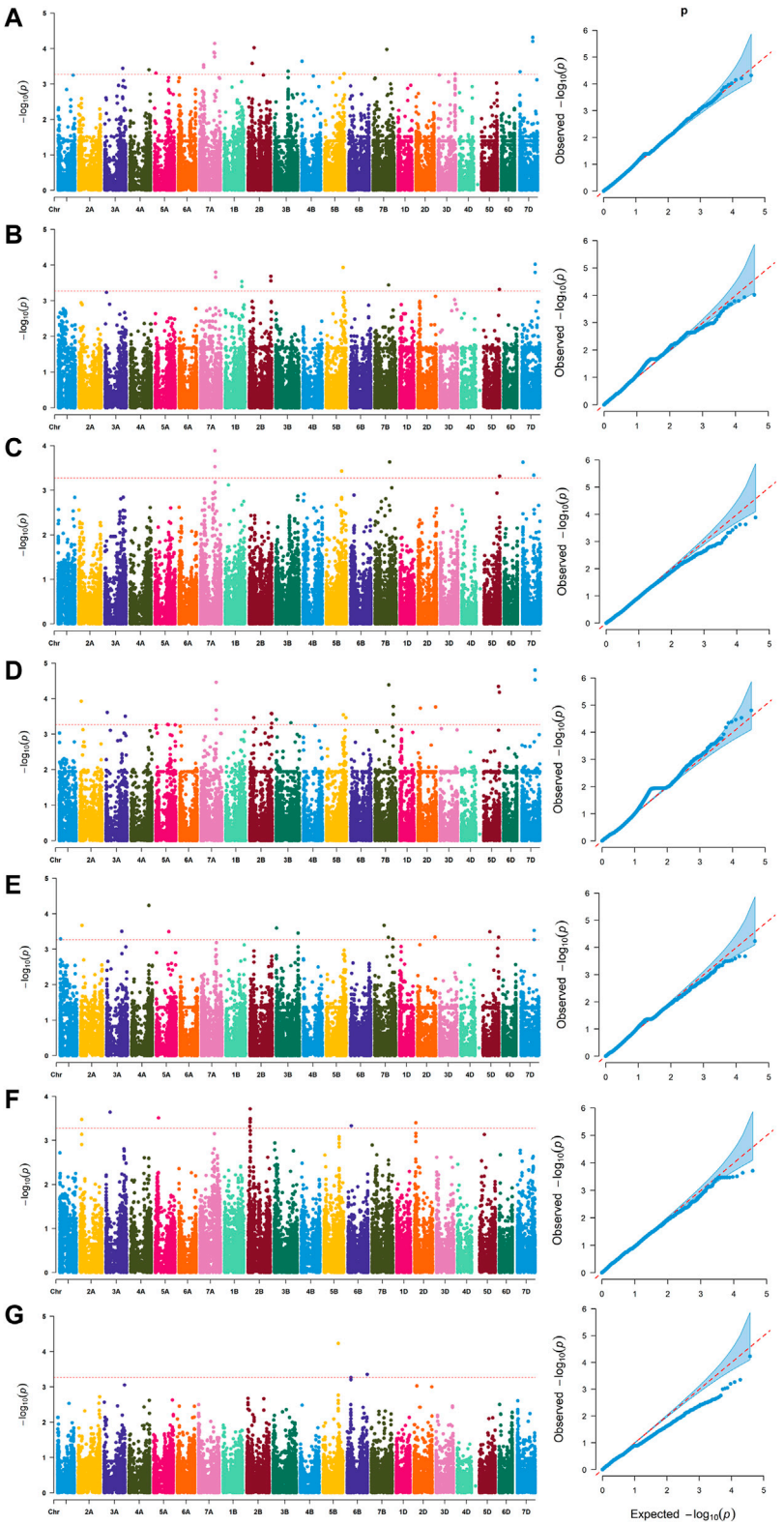


FIGURE 2 | Manhattan and quantile-quantile **(Q–Q)** plots for RVA parameters identified by genome-wide association study based on BLUP values. **(A)** peak viscosity (PV). **(B)** trough viscosity (TV). **(C)** breakdown (BD). **(D)** final viscosity (FV). **(E)** setback (SB). **(F)** peak time (PT). **(G)** pasting temperature (T).

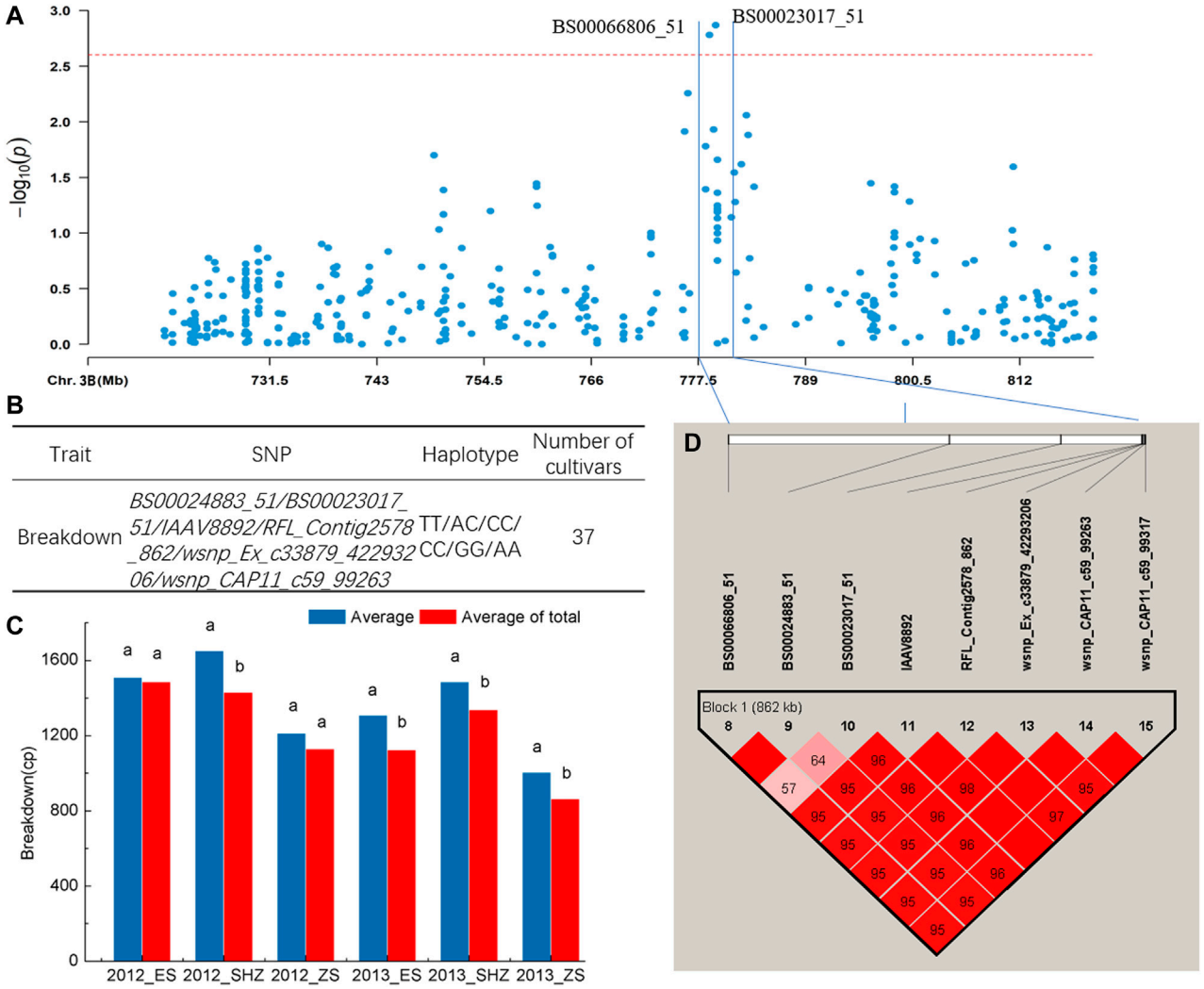


FIGURE 3 | Haplotype analysis. **(A)** Local Manhattan plot. **(B)** Haplotypes with different SNP alleles in the block. **(C)** Phenotypic effects of haplotypes in different environments, Different lowercase letters indicate significant differences at $p < 0.05$. **(D)** Haplotype analysis of multi-environment significant SNPs associated with breakdown on chromosome 3B, the color represents the linkage between SNPs, and the deeper color means the higher linkage between SNPs.

RAC875_c4438_419/IAAV3414/TA001679-0356) haplotypes showed higher T than the average of all cultivars surveyed, exhibiting significant differences in some locations ($p < 0.05$, **Supplementary Table S6**, **Supplementary Figure S3**).

DISCUSSION

Phenotypic Variations and Heritabilities in RVA Parameters

Identification of genetic loci controlling wheat quality parameters is useful in breeding programs (Matus and Hayes, 2002; Mourad et al., 2020). In the association panel evaluated in this study, we observed significant phenotypic variations among accessions in RVA parameters (**Supplementary Table S1**). Earlier researches were in line with current results that starch pasting properties exhibited large variations (He et al., 2006; Ram and Sharma, 2013). Successive phenotypic distributions indicated polygenic inheritance of RVA parameters (**Supplementary Figure S1**). Heritability estimation provides information about the extent of a particular genetic character to be transmitted to offspring (Muhu-Din Ahmed et al., 2020). In this experiment, RVA parameters had high heritabilities (**Supplementary Table S1**), similar with a previous study, which also reported a high heritability in RVA parameters (Rahim et al., 2020). Our results demonstrated that this panel of wheat varieties had high levels of intra-species genetic flow, which made it suitable for the genetic study of RVA parameters by GWAS.

GWAS for RVA Parameters

Marker-trait association study established the relationship between specific phenotypic and genetic variability within a

genome, which ultimately detected loci underpinning corresponding traits. The ability to capture significant associations between polymorphic loci and phenotypic variance depend on the extent of LD along the genome (Pritchard and Przeworski, 2001; Remington et al., 2001). LD decay was about 7 Mb in our association panel in the present study, which consistent with previous studies (Appels et al., 2018; Yang et al., 2020), but inconsistent with most studies (Sukumaran et al., 2015; Jamil et al., 2019; Roncallo et al., 2019; Muhammad et al., 2020; Hu et al., 2021; Sharma, 2021), because of the differences in the type of markers used for genotyping and the sample size variation in the different studies (Chao et al., 2010). In the past, many studies for protein quality in wheat have been reported (Kristensen et al., 2019; Yang et al., 2020), whereas, few reports are about QTL for RVA parameters. In this study, we identified SNP markers associated with RVA traits using MLM. Besides the BLUP value, GWAS was also conducted with data of individual environments as a reference for locating SNPs that were relatively stable across different experimental environments. We highlighted those SNPs detected in at least two environments as stable SNPs and 41 significant MTAs were identified (**Supplementary Table S3**).

Loci for RVA Parameters

Significant and stable MTAs for PV were mainly clustered to chromosomes 2B, 3A, 3B, 4A, 4B, 5A, 7A, 7B and 7D (**Supplementary Table S3**); several were consistent with previous studies (Udall et al., 1999; Sun et al., 2008; Zhang et al., 2009; Deng et al., 2014). *Wx-B1* gene (*TRAESCS4A02G418200*, chr4A, 688,097,145–688,100,962 bp) encodes granule-bound starch synthase, generating a higher PV and BD (Briney et al., 1998; Araki et al., 2000; Batey et al., 2002; Miura et al., 2002; Ram and Sharma, 2013). Some studies found the QTL for starch pasting properties near *Wx-B1* on chromosome 4A (Batey et al., 2002; Mccartney et al., 2006). In the present study, we also identified a SNP marker (*BobWhite_c17731_56*, Chr.4A: 689,849,795 bp, **Supplementary Table S3**) associated with PV, its physical position was very close to *Wx-B1* gene. *TRAESCS7D02G365900* on chromosome 7D at 473,617,760 to 473,624,545 bp, encodes phosphorylase which probably contributes to starch synthesis and degradation (Tetlow et al., 2004; Tickle et al., 2009; Mishra et al., 2016). In the current study, we found two SNP markers (*IAAV4275* and *Kukri_rep_c69088_774*) around there.

Our results corroborated other QTL studies in wheat, where QTL associated with BD was detected on chromosome 3B (**Supplementary Table S3**). However, we did not find QTL for BD on chromosomes 1A, 1B, 3A, 4A, 4B, 6A and 6D (Zhang et al., 2009; Deng et al., 2014). We detected new QTL (*GENE-4993_69* on chromosome 7B, *Excalibur_c9183_1397* and *Tdurum_contig69003_459* on chromosome 7D, **Supplementary Table S3**) were not reported in previous studies. One stable SNP (*BS00067650_51*, **Supplementary Table S3**) associated with FV was detected at 526,464,187 bp on chromosome 5D in this study which was never reported before, while several FV QTL were identified on chromosomes 1A, 1B, 1D, 3A, 3D, 4A, 4B, 5B, 5D, 6B, 6D and 7A previously (Batey et al., 2002; Mccartney et al., 2006; Sun et al., 2008; Zhang et al., 2009; Deng et al., 2014). Using two DH populations, Batey et al. (2002) found QTL for FV near *Wx-B1* on chromosome 4A, where no locus associated with FV was detected in this research. Nine SNPs in seven loci on chromosomes 2B (2), 5B (1), 6B (1), 7B (2) and 6D (1) were significantly associated with T (**Supplementary Table S3**), whereas QTL for T were reported on chromosomes 2D, 3A, 3B, 4A, 4B, 5D, 6D in previous studies (Zhao et al., 2009; Deng et al., 2014). The MTA at *BS00022437_51* (chr.6B, 715,775,218 bp) significantly associated with T was very close to *TRAESCS6B02G418100* (chr.6B, 690,447,513–690,452,026 bp) which was the homolog of phosphorylase (PHO, Tetlow et al., 2004; Li et al., 2018).

We also detected eight stable SNPs at five loci for TV on chromosomes 7A, 2B, 5B, 7B and 7D, one SNP for SB on chromosome 5A, and one SNP for PT on chromosome 2A (**Supplementary Table S3**). Whereas, previous studies have reported more loci on other chromosomes (Sun et al., 2008; Deng et al., 2014), which were not detected in this study, because alleles at these loci might be fixed in this association panel, or rare alleles cannot be detected.

Candidate Genes

The annotation conducted on 41 stable SNPs identified by GWAS showed that seven (17%) SNPs were mapped in intergenic regions, and 34 (83%) were mapped in genic regions (**Supplementary Table S4**). Among these, 20 SNPs were located in the CDS of 18 annotation genes which will result in discovering new genes controlling RVA parameters in wheat. Furthermore, *GENE-4428_113*, *BobWhite_c13098_670*, *IAAV4275* and *Kukri_rep_c69088_774* had effects on both PV and TV, and *Excalibur_c9183_1397* had effect on both PV and BD; these SNPs were all mapped in the CDS of annotation genes, which may be key candidate genes that participate in regulating wheat starch quality. *Excalibur_c9183_1397* was located in the CDS region of *TraesCS7D02G026700* encoding 1,3-beta-glucan synthase that is present mainly in the cell walls of starchy endosperm (Moravčíková et al., 2016). Other candidate genes that were not reported in previous research need to be paid more attention for study in the future.

Haplotype Analysis

In the present study, SNPs significantly associated with RVA parameters were identified on almost all wheat chromosomes. To validate the effect of each SNP, estimations of SNP effects were used to predict the observed phenotypic performance (**Supplementary Table S5**). *BS00023017_51*, *Excalibur_c10124_361* and *Kukri_c4560_110* had significant ($p < 0.05$) effects on traits in at least four environments, and could be used as optimal loci in marker-assisted breeding and quality improvement. The haplotype analysis was conducted on these loci and three blocks were detected (**Figure 3**, **Supplementary Figure S3**; **Supplementary Table S6**). Cultivars with superior haplotypes showed relatively better phenotypes than the average of all cultivars surveyed.

Therefore, these loci can be considered to improve cultivars for starch quality.

CONCLUSION

It is not feasible that phenotypic selection for improving wheat starch quality at the early stages in wheat breeding. Even at later stages in the breeding process, starch quality was also impacted greatly by environments. Therefore, genetic selection through MAS is a desirable way to improve wheat starch quality in wheat breeding programs. Based on dense SNPs across the whole genome, the GWAS has become a common approach to uncover genetic components of agronomic traits, which provides us with insightful information into genetic architecture of complex traits. In this study, GWAS analysis were performed for RVA parameters with 47,362 SNPs in 192 Chinese spring wheat accessions among six environments. Forty-one SNPs at 25 loci were stably detected in at least two environments, of which 20 SNPs were located in the CDS of 18 annotation genes. Haplotype analysis for regions harboring the SNPs (*BS00023017_51, Excalibur_c10124_361, Kukri_c4560_110*) revealed one block for BD on chromosome 3B and two blocks for T on chromosome 7B, cultivars with superior haplotypes at these loci showed better starch pasting viscosity than the average of all cultivars surveyed. Validation studies for SNPs in the candidate genes and detected loci will be conducted in the future by designing KASP assays, which can be further used for marker-assisted breeding for improvement of grain starch quality in wheat.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in Figshare DOI: 10.6084/m9.figshare.19297601.

AUTHOR CONTRIBUTIONS

WL and PM conceived and planned the research; WS, PL, JX, FC, HX, XH, YN, and DK performed the research; YT analyzed the data and wrote the manuscript; JL edited and reviewed the final manuscript. All authors read and approved the final manuscript.

FUNDING

This work was funded by the Key Science and Technology Project of Xinjiang Production and Construction Corps (2019AB021), the National Natural Science Foundation of China (U1178306), the international science and technology cooperation program project of Xinjiang Production and Construction Corps (2019BC003).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.830644/full#supplementary-material

REFERENCES

Al-Maskri, A. Y., Sajjad, M., and Khan, S. H. (2012). Association Mapping: a Step Forward to Discovering New Alleles for Crop Improvement. *Int. J. Agric. Biol.* 14, 153–160. doi:10.13140/2.1.1925.9524

Amiri, R., Sasani, S., Jalali-Honarmand, S., Rasaei, A., Seifolahpour, B., and Bahraminejad, S. (2018). Genetic Diversity of Bread Wheat Genotypes in Iran for Some Nutritional Value and Baking Quality Traits. *Physiol. Mol. Biol. Plants* 24, 147–157. doi:10.1007/s12298-017-0481-4

Appels, R., Eversole, K., Appels, R., Eversole, K., Feuillet, C., Keller, B., et al. (2018). Shifting the Limits in Wheat Research and Breeding Using a Fully Annotated Reference Genome. *Science* 361 (6403), 661–674. doi:10.1126/science.aar7191

Araki, E., Miura, H., and Sawada, S. (2000). Differential Effects of the Null Alleles at the Three Wx Loci on the Starch-Pasting Properties of Wheat. *Theor. Appl. Genet.* 100, 1113–1120. doi:10.1007/s001220051394

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: Analysis and Visualization of LD and Haplotype Maps. *Bioinformatics* 21, 263–265. doi:10.1093/bioinformatics/bth457

Bates, D. M., Mäechler, M., Bolker, B. M., and Walker, S. (2015). Lme4: Linear Mixed-Effects Models Using "Eigen" and S4. *J. Stat. Softw.* 67. doi:10.18637/jss.v067.i01

Batey, I. L., Hayden, M. J., Cai, S., Sharp, P. J., Cornish, G. B., Morell, M. K., et al. (2002). Genetic Mapping of Commercially Significant Starch Characteristics in Wheat Crosses. *Aust. J. Agric. Res.* 52, 1287–1296. doi:10.1071/AR01053

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

Briney, A., Wilson, R., Potter, R. H., Barclay, I., Crosbie, G., Appels, R., et al. (1998). A PCR-Based Marker for Selection of Starch and Potential Noodle Quality in Wheat. *Mol. Breed.* 4, 427–433. doi:10.1023/A:1009664917998

Chao, S., Dubcovsky, J., Dvorak, J., Luo, M.-C., Baenziger, S. P., Matnyazov, R., et al. (2010). Population- and Genome-specific Patterns of Linkage Disequilibrium and SNP Variation in spring and winter Wheat (Triticum aestivum L.). *BMC Genomics* 11, 727. doi:10.1186/1471-2164-11-727

Crosbie, G. B. (1991). The Relationship between Starch Swelling Properties, Paste Viscosity and Boiled Noodle Quality in Wheat Flours. *J. Cereal Sci.* 13, 145–150. doi:10.1016/S0733-5210(09)80031-3

Deng, Z., Tian, J., Chen, F., Li, W., Zheng, F., Chen, J., et al. (2014). Genetic Dissection on Wheat Flour Quality Traits in Two Related Populations. *Euphytica* 203, 221–235. doi:10.1007/s10681-014-1318-7

Earl, D. A., and VonHoldt, B. M. (2011). STRUCTURE HARVESTER: a Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conservation Genet. Resour.* 4, 359–361. doi:10.1007/s12686-011-9548-7

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, 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

He, Z. H., Liu, A. H., Peña, R. J., and Rajaram, S. (2003). Suitability of Chinese Wheat Cultivars for Production of Northern Style Chinese Steamed Bread. *Euphytica* 131, 155–163. doi:10.1023/A:1023929513167

He, Z. H., Yang, J., Zhang, Y., Quail, K. J., and Peña, R. J. (2004). Pan Bread and Dry white Chinese Noodle Quality in Chinese winter Wheats. *Euphytica* 139, 257–267. doi:10.1007/s10681-004-3283-z

He, Z., Xu, Z., Xia, L., Xia, X., Yan, J., Zhang, Y., et al. (2006). Genetic Variation for Waxy Proteins and Starch Properties in Chinese winter Wheats. *Cereal Res. Commun.* 34, 1145–1151. doi:10.1556/crc.34.2006.2-3.254

Hu, P., Zheng, Q., Luo, Q., Teng, W., Li, H., Li, B., et al. (2021). Genome-wide Association Study of Yield and Related Traits in Common Wheat under Salt-Stress Conditions. *BMC Plant Biol.* 21 (1), 1–20. doi:10.1186/s12870-020-02799-1

Jamil, M., Ali, A., Gul, A., Ghafoor, A., Napar, A. A., Ibrahim, A. M. H., et al. (2019). Genome-wide Association Studies of Seven Agronomic Traits under Two Sowing Conditions in Bread Wheat. *BMC Plant Biol.* 19 (1), 1–18. doi:10.1186/s12870-019-1754-6

Kaur, A., Shevkani, K., Katyal, M., Singh, N., Ahlawat, A. K., and Singh, A. M. (2016). Physicochemical and Rheological Properties of Starch and Flour from Different Durum Wheat Varieties and Their Relationships with Noodle Quality. *J. Food Sci. Technol.* 53, 2127–2138. doi:10.1007/s13197-016-2202-3

Kong, L., Si, J., Zhang, B., Feng, B., Li, S., and Wang, F. (2013). Environmental Modification of Wheat Grain Protein Accumulation and Associated Processing Quality: a Case Study of China. *Aust. J. Crop Sci.* 7, 173–181.

Konik, C. M., Miskelly, D. M., and Gras, P. W. (1992). Contribution of Starch and Non-starch Parameters to the Eating Quality of Japanese white Salted Noodles. *J. Sci. Food Agric.* 58, 403–406. doi:10.1002/jsfa.2740580315

Kristensen, P. S., Jensen, J., Andersen, J. R., Guzmán, C., Orabi, J., and Jahoor, A. (2019). Genomic Prediction and Genome-wide Association Studies of Flour Yield and Alveograph Quality Traits Using Advanced winter Wheat Breeding Material. *Genes* 10, 669–688. doi:10.3390/genes10090669

Kumar, D., Sharma, S., Sharma, R., Pundir, S., Singh, V. K., Chaturvedi, D., et al. (2021). Genome-wide Association Study in Hexaploid Wheat Identifies Novel Genomic Regions Associated with Resistance to Root Lesion Nematode (*Pratylenchus Thornei*). *Sci. Rep.* 11 (1), 3572. doi:10.1038/s41598-021-80996-0

León, A. E., Barrera, G. N., Pérez, G. T., Ribotta, P. D., and Rosell, C. M. (2006). Effect of Damaged Starch Levels on Flour-thermal Behaviour and Bread Staling. *Eur. Food Res. Technol.* 224, 187–192. doi:10.1007/s00217-006-0297-x

Li, W., Chen, Z. Y., Li, Z., Zhao, X. F., Pu, Z. E., Chen, G. Y., et al. (2018). Characterization of Starch Synthetic Genes and Starch Granule during Seeds Development between Synthetic Hexaploid Wheat and its Parents. *Cereal Res. Commun.* 46, 275–286. doi:10.1556/0806.46.2018.09

Liu, J. J., He, Z. H., Zhao, Z. D., Peña, R. J., and Rajaram, S. (2003). Wheat Quality Traits and Quality Parameters of Cooked Dry white Chinese Noodles. *Euphytica* 131, 147–154. doi:10.1023/A:1023972032592

Lv, G., Dong, Z., Wang, Y., Geng, J., Li, J., Lv, X., et al. (2020). Identification of Genetic Loci of Black Point in Chinese Common Wheat by Genome-wide Association Study and Linkage Mapping. *Plant Dis.* 104 (7), 2005–2013. doi:10.1094/PDIS-12-19-2733-RE

Matus, I. A., and Hayes, P. M. (2002). Genetic Diversity in Three Groups of Barley Germplasm Assessed by Simple Sequence Repeats. *Genome* 45, 1095–1106. doi:10.1139/g02-071

Mccartney, C. A., Somers, D. J., Lukow, O., Ames, N., Noll, J., Cloutier, S., et al. (2006). QTL Analysis of Quality Traits in the spring Wheat Cross RL4452×'AC Domain'. *Plant Breed.* 125, 565–575. doi:10.1111/j.1439-0523.2006.01256.x

Mishra, A., Singh, A., Sharma, M., Kumar, P., and Roy, J. (2016). Development of EMS-Induced Mutation Population for Amylose and Resistant Starch Variation in Bread Wheat (*Triticum aestivum*) and Identification of Candidate Genes Responsible for Amylose Variation. *BMC Plant Biol.* 16, 217–232. doi:10.1186/s12870-016-0896-z

Miura, H., Wickramasinghe, M. H. A., Subasinghe, R. M., Araki, E., and Komae, K. (2002). Development of Near-Isogenic Lines of Wheat Carrying Different Null Wx Alleles and Their Starch Properties. *Euphytica* 123, 353–359. doi:10.1023/A:1015042322687

Moiraghi, M., Sciarini, L. S., Paesani, C., León, A. E., and Pérez, G. T. (2019). Flour and Starch Characteristics of Soft Wheat Cultivars and Their Effect on Cookie Quality. *J. Food Sci. Technol.* 56, 4474–4481. doi:10.1007/s13197-019-03954-9

Moravčíková, J., Margetínyová, D., Gálová, Z., Žur, I., Gregorová, Z., Zimová, M., et al. (2016). Beta-1,3-glucanase Activities in Wheat and Relative Species. *Nova Biotechnologica Chim.* 15, 122–132. doi:10.1515/nbec-2016-0013

Mourad, A. M. I., Belamkar, V., and Baenziger, P. S. (2020). Molecular Genetic Analysis of spring Wheat Core Collection Using Genetic Diversity, Population Structure, and Linkage Disequilibrium. *BMC Genomics* 21, 434–456. doi:10.1186/s12864-020-06835-0

Muhammad, A., Hu, W., Li, Z., Li, J., Xie, G., Wang, J., et al. (2020). Appraising the Genetic Architecture of Kernel Traits in Hexaploid Wheat Using GWAS. *Ijms* 21, 5649. doi:10.3390/ijms21165649

Muhu-Din Ahmed, H. G., Sajjad, M., Zeng, Y., Iqbal, M., Habibullah Khan, S., Ullah, A., et al. (2020). Genome-wide Association Mapping through 90K SNP Array for Quality and Yield Attributes in Bread Wheat against Water-Deficit Conditions. *Agriculture* 10, 392–425. doi:10.3390/agriculture10090392

Muqaddasi, Q. H., Brassac, J., Ebmeyer, E., Kollers, S., Korzun, V., and Argillier, O. (2020). Prospects of GWAS and Predictive Breeding for European winter Wheat's Grain Protein Content, Grain Starch Content, and Grain Hardness. *Sci. Rep.* 10, 12541. doi:10.1038/s41598-020-69381-5

Panozzo, J. F., Plieske, J., Ganal, M. W., and Röder, M. S. (1993). The Rapid Viscoanalyser as a Method of Testing for Noodle Quality in a Wheat Breeding Programme. *J. Cereal Sci.* 17, 25–32. doi:10.1006/jcrs.1993.1004

Pritchard, J. K., and Przeworski, M. (2001). Linkage Disequilibrium in Humans: Models and Data. *Am. J. Hum. Genet.* 69, 1–14. doi:10.1086/321275

Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association Mapping in Structured Populations. *Am. J. Hum. Genet.* 67, 170–181. doi:10.1086/302959

Rahim, M. S., Mishra, A., Katyal, M., Thakur, S., Sharma, M., Kumar, P., et al. (2020). Marker-trait Association Identified Candidate Starch Biosynthesis Pathway Genes for Starch and Amylose-Lipid Complex Gelatinization in Wheat (*Triticum aestivum* L.). *Euphytica* 216, 151–173. doi:10.1007/s10681-020-02688-6

Ram, S., and Sharma, I. (2013). Allelic Diversity in Granule Bound Starch Synthase Genes in Indian Wheats and Their Relationship with Starch Pasting Properties. *Cereal Res. Commun.* 41, 141–149. doi:10.1556/CRC.2012.0024

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., et al. (2001). Structure of Linkage Disequilibrium and Phenotypic Associations in the maize Genome. *Proc. Natl. Acad. Sci.* 98, 11479–11484. doi:10.1073/pnas.201394398

Rimbert, H., Darrier, B., Navarro, J., Kitt, J., Choulet, F., Leveugle, M., et al. (2018). High Throughput SNP Discovery and Genotyping in Hexaploid Wheat. *PLoS One* 13, e0186329. doi:10.1371/journal.pone.0186329

Roncallo, P. F., Beaufort, V., Larsen, A. O., Dreisigacker, S., and Echenique, V. (2019). Genetic Diversity and Linkage Disequilibrium Using SNP (KASP) and AFLP Markers in a Worldwide Durum Wheat (*Triticum Turgidum* L. Var *Durum*) Collection. *PLoS One* 14 (6), e0218562. doi:10.1371/journal.pone.0218562

Shi, C., Zheng, Y., Geng, J., Liu, C., Pei, H., Ren, Y., et al. (2020). Identification of Herbicide Resistance Loci Using a Genome-wide Association Study and Linkage Mapping in Chinese Common Wheat. *Crop J.* 8, 666–675. doi:10.1016/j.cj.2020.02.004

Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., and Reynolds, M. P. (2015). Genome-wide Association Study for Grain Yield and Related Traits in an Elite spring Wheat Population Grown in Temperate Irrigated Environments. *Theor. Appl. Genet.* 128, 353–363. doi:10.1007/s00122-014-2435-3

Sun, C., Zhang, F., Yan, X., Zhang, X., Dong, Z., Cui, D., et al. (2017). Genome-wide Association Study for 13 Agronomic Traits Reveals Distribution of superior Alleles in Bread Wheat from the Yellow and Huai Valley of China. *Plant Biotechnol. J.* 15, 953–969. doi:10.1111/pbi.12690

Sun, H., Lü, J., Fan, Y., Zhao, Y., Kong, F., Li, R., et al. (2008). Quantitative Trait Loci (QTLs) for Quality Traits Related to Protein and Starch in Wheat. *Prog. Nat. Sci.* 18, 825–831. doi:10.1016/j.pnsc.2007.12.013

Tadesse, W., Ogbonnaya, F. C., Jighly, A., Sanchez-Garcia, M., Sohail, Q., Rajaram, S., et al. (2015). Genome-wide Association Mapping of Yield and Grain Quality Traits in winter Wheat Genotypes. *PLoS One* 10, e0141339. doi:10.1371/journal.pone.0141339

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, 2131–2145. doi:10.1093/jxb/erh248

Tickle, P., Burrell, M. M., Coates, S. A., Emes, M. J., Tetlow, I. J., and Bowsher, C. G. (2009). Characterization of Plastidial Starch Phosphorylase in *Triticum aestivum* L. Endosperm. *J. Plant Physiol.* 166, 1465–1478. doi:10.1016/j.jplph.2009.05.004

Udall, J. A., Souza, E., Anderson, J., Sorrells, M. E., and Zemetra, R. S. (1999). Quantitative Trait Loci for Flour Viscosity in winter Wheat. *Crop Sci.* 39, 238–242. doi:10.2135/cropsci1999.0011183X003900010036x

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, 2702. doi:10.1038/s41598-019-38570-2

Yang, Y., Chai, Y., Zhang, X., Lu, S., Zhao, Z., Wei, D., et al. (2020). Multi-locus GWAS of Quality Traits in Bread Wheat: Mining More Candidate Genes and Possible Regulatory Network. *Front. Plant Sci.* 11, 1091–1110. doi:10.3389/fpls.2020.01091

Zhang, Y., He, Z. H., Ye, G. Y., Zhang, A. M., and Ginkel, M. V. (2004). Effect of Environment and Genotype on Bread-Making Quality of spring-sown spring Wheat Cultivars in China. *Euphytica* 139, 75–83. doi:10.1007/s10681-004-2131-5

Zhang, X., Chen, J., Yan, Y., Yan, X., Shi, C., Zhao, L., et al. (2018). Genome-wide Association Study of Heading and Flowering Dates and Construction of its Prediction Equation in Chinese Common Wheat. *Theor. Appl. Genet.* 131, 2271–2285. doi:10.1007/s00122-018-3181-8

Zhang, Y., Nagamine, T., He, Z. H., Ge, X. X., Yoshida, H., and Peña, R. J. (2005). Variation in Quality Traits in Common Wheat as Related to Chinese Fresh white Noodle Quality. *Euphytica* 141, 113–120. doi:10.1007/s10681-005-6335-0

Zhang, Y., Wu, Y., Xiao, Y., Yan, J., Zhang, Y., Zhang, Y., et al. (2009). QTL Mapping for Milling, Gluten Quality, and Flour Pasting Properties in a Recombinant Inbred Line Population Derived from a Chinese Soft×hard Wheat Cross. *Crop Pasture Sci.* 60, 587–597. doi:10.1071/CP08392

Zhao, L., Zhang, K., Liu, B., and Tian, J. (2009). Detection of Quantitative Trait Loci for Paste Viscosity Characteristics Based on the Doubled Haploid Progeny from a Cross between Two Chinese Wheat Varieties. *Can. J. Plant Sci.* 89, 837–844. doi:10.4141/cjps08201

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.

frontiers | Frontiers in Genetics

MINI REVIEW
published: 18 July 2022
doi: 10.3389/fgene.2022.866121

New Hope for Genome Editing in Cultivated Grasses: CRISPR Variants and Application

Asad Riaz[1†], Farah Kanwal[1†], Iqrar Ahmad[2], Shakeel Ahmad[2], Ayesha Farooq[2], Claus Krogh Madsen[3], Henrik Brinch-Pedersen[3], Zelalem Eshetu Bekalu[3], Fei Dai[1], Guoping Zhang[1] and Ahmad M. Alqudah[3]**

[1]College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China, [2]Centre for Advanced Studies in Agriculture and Food Security, University of Agriculture Faisalabad, Faisalabad, Pakistan, [3]Department of Agroecology, Research Center Flakkebjerg, Aarhus University, Slagelse, Denmark

OPEN ACCESS

Edited by:
Mahendar Thudi,
Dr. Rajendra Prasad Central Agricultural University, India

Reviewed by:
Muntazir Mushtaq,
National Bureau of Plant Genetic Resources (ICAR), India
Kumari Anjani,
Dr. Rajendra Prasad Central Agricultural University, India

***Correspondence:**
Guoping Zhang
zhanggp@zju.edu.cn
Ahmad M. Alqudah
ama@agro.au.dk
ahqudah@gmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 30 January 2022
Accepted: 23 June 2022
Published: 18 July 2022

Citation:
Riaz A, Kanwal F, Ahmad I, Ahmad S, Farooq A, Madsen CK, Brinch-Pedersen H, Bekalu ZE, Dai F, Zhang G and Alqudah AM (2022) New Hope for Genome Editing in Cultivated Grasses: CRISPR Variants and Application. Front. Genet. 13:866121. doi: 10.3389/fgene.2022.866121

With the advent of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated protein (Cas) mediated genome editing, crop improvement has progressed significantly in recent years. In this genome editing tool, CRISPR-associated Cas nucleases are restricted to their target of DNA by their preferred protospacer adjacent motifs (PAMs). A number of CRISPR-Cas variants have been developed e.g. CRISPR-Cas9, -Cas12a and -Cas12b, with different PAM requirements. In this mini-review, we briefly explain the components of the CRISPR-based genome editing tool for crop improvement. Moreover, we intend to highlight the information on the latest development and breakthrough in CRISPR technology, with a focus on a comparison of major variants (CRISPR-Cas9, -Cas12a, and -Cas12b) to the newly developed CRISPR-SpRY that have nearly PAM-less genome editing ability. Additionally, we briefly explain the application of CRISPR technology in the improvement of cultivated grasses with regard to biotic and abiotic stress tolerance as well as improving the quality and yield.

Keywords: plant genome editing, CRISPR, Cas9, Cas12, SPRY, cultivated-grasses, stress tolerance

INTRODUCTION

Since the advent of agriculture, plants have been cultivated and utilized as a source of food and energy to feed humans and livestock. The world's population is expected to increase by 40% in 2050 and the demand for food will be increased by 50% (Tilman et al., 2011; van Dijk et al., 2021). To meet the demand for food, crop production needs to be significantly improved in the near coming decades (Ronald 2011). With time, the evolutionary drive, domestication, and breeding of cultivated plants have transformed due to scientific advancement and environmental conditions. However, there is the utmost need for future crop improvement with better adapted to harsh weather and enhanced agronomical traits of yield (Koeppel et al., 2019).

Previously, the crops were improved by conventional breeding methods which are time-consuming and laborious. Those traditional methods are reinforced by modern molecular and genomic-based breeding techniques to meet future food demand (Riaz et al., 2021). Recently, molecular biology has progressed with several great discoveries including genome sequencing and genetic engineering. Many molecular approaches were introduced for genome editing which relied on site-specific recognition of DNA sequences via zinc finger nucleases (ZFNs) and

transcription activator-like effector nucleases (TALENS) which rely on DNA-protein interaction and require case-by-case protein re-design (Urnov et al., 2010; Miller et al., 2011). The discovery of the bacterial Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR associated protein (Cas) system (Gasiunas et al., 2012; Jinek et al., 2012) and its adaptation to genome editing in various organisms including plants provided a much more accessible method that has initiated a revolution in crop improvement (Shan et al., 2013). A wide range of plants including agronomic crops such as rice, wheat, maize, and barley are being subjected to molecular improvement using the new breeding technology CRISPR, which facilitates genome editing by gene deletion or/and insertion, and replacement (Liu HJ. et al., 2020; Usman et al., 2020; Han et al., 2021; Zhang et al., 2021). The CRISPR-Cas 9 system was originally discovered in bacteria and archaea immune systems where it detects and degrades invasive DNA from bacteriophages and plasmids. The active component is a ribonucleoprotein complex consisting of a Cas RNA-guided endonuclease and guide RNA that recognizes target DNA through the variable protospacer motif (Fineran and Charpentier 2012). In order to effectively introduce double-strand breaks, CRISPR requires a PAM (protospacer adjacent motif) sequence to be present in the target DNA adjacent to the protospacer complementary sequence. This poses a limitation to genome editing designs and has motivated the search for variants of CRISPR tools with alternative PAM requirements. PAM is a short 2–6 bp sequence preceded by the targeted DNA sequence. Cas9 nuclease from the type II CRISPR-Cas 9 system of *Streptococcus pyogenes* is the most commonly used system and it requires an NGG (N, any nucleotide; G, guanine) PAM sequence for DNA targeting (Jinek et al., 2012). The CRISPR-based genome editing tool is now extensively embraced with a low-cost, fast, and easy-to-use targeted gene editing system to cultivated grasses (Cram et al., 2019; Kim et al., 2019; Zhang et al., 2020; Lawrenson et al., 2021). New variants of CRISPR with different PAM sequences such as CRISPR-Cas 12a (formerly known as Cpf1) (Chen et al., 2018) and CRISPR-Cas 12b (formerly known as C2c1) (Ming et al., 2020) have been developed. At the beginning of the year 2021, the application of the next generation of CRISPR-Cas variant known as CRISPR-SpRY has been developed without any PAM restriction that improves gene editing resolution (Walton et al., 2020). The progress in CRISPR technology increases the efficacy and specificity which significant improvement editing outcomes and widen the applications in crop improvement.

This review discusses the components of CRISPR-based genome editing tools in detail and presents a quick comparison of previous main CRISPR variants with the latest one based on PAM sequence requirements. Moreover, an overview of different applications of CRISPR-based genome editing tools is also covered in a schematic diagram and its application for the improvement of agronomic traits as well as biotic and abiotic stresses in cultivated grasses such as rice (*Oryzae sativa*), maize (*Zea mays*), barley (*Hordeum vulgare*) and wheat (*Triticum aestivum*) is explained briefly. Therefore, this review showed the application and usefulness of genome editing as a new breeding technology for crop improvement.

CRISPR-BASED GENOME EDITING TOOL: COMPONENTS AND MECHANISM

Components

CRISPR/Cas9 has been widely studied, well understood, and extensively used (Liu ZQ. et al., 2020). Two major components are essential for typical engineered CRISPR-Cas systems, a Cas endonuclease protein, and a single guide RNA (sgRNA) of 20 nucleotide sequences that guide the Cas enzyme to the target sequence for introducing the double-stranded break (DSB) (Mahfouz et al., 2014). Multiple variants of Cas9 and gRNA are available according to their novel application in the field of genetic engineering in plants (Chen et al., 2019; Li et al., 2019).

Cas-9 consists of two regions, called the recognition (REC) lobe and the nuclease (NUC) lobe. The REC lobe has two multi-helix domains, named REC1 and REC2, essential components to bind with the guide RNA and target DNA (Makarova et al., 2017). REC1 is comprised of an extended α-helical structure of 25 alpha helices and 2 β-sheets, whereas REC2 has a six-helix structure and is embedded within the REC1 domain (Nishimasu et al., 2014). The NUC lobe has three - domains: RuvC, HNH, and PAM-interacting domains. In order to cut a double-stranded DNA, the REC lobe initiates the binding of sgRNA and DNA, whereas the RuvC and HNH domains properly execute the cleavage of the complementary and non-complementary strand of the target DNA, respectively. In the meantime, the carboxy-terminal residing PAM-interacting domain confers PAM interaction and specificity to the target DNA (Yamano et al., 2016).

Guide RNA is composed of two elements, CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA). The crRNA is a long sequence of 18–20 base pairs that recognize and specify the target DNA by binding with it (Barrangou 2013). The tracrRNA is a long, twisted structure that serves as a binding scaffold for Cas-9 nuclease (Li 2015). The sequence of tracrRNA is partially complementary to a segment of crRNA. And, the base pairing of complementary sequences results in an RNA-duplex (tracrRNA–crRNA) and activates the Cas9 to form the Cas9-crRNA-tracrRNA editing complex (Mojica et al., 2009). The structural engineering of tracrRNA–crRNA duplex into a single guide RNA (sgRNA) creates a dual component Cas9-sgRNA system that simplifies the editing of genomic regions (Jinek et al., 2012).

Mechanism

Biological systems of CRISPR/Cas are a part of the adaptive immune system of bacteria and archaea, protecting them from nucleic acid invaders such as viruses by cleaving the alien DNA in a sequence-dependent fashion (Bortesi and Fischer, 2015). At the proximal end of a CRISPR locus, short fragments of the invading DNA (spacers) are integrated between two adjacent repeats to

confer immunity to the invading cells. Upon subsequent encounters with invasive DNA, the CRISPR arrays and spacers are transcribed to produce the 40 nt small interfering crRNAs, which bind together with tracrRNAs to activate and guide the Cas9 nuclease (Karvelis et al., 2013). Cas9 activation tempts the REC lobe to undertake conformational changes and forms a central channel to accommodate the negatively charged guide-RNA and target-DNA heteroduplex in a positively charged interface between the REC and NUC lobes (Nishimasu et al., 2014). The RuvC and the PAM-interacting domain create a positively charged surface to interact with the 3′end of the sgRNA, and the catalytic domain HNH comes closer to the DNA cleavage site.

Upon finding the appropriate PAM for the target site, Cas-9 triggers local DNA melting followed by hybridization of RNA with DNA leading to the activation of Cas-9 protein to cleave target DNA (Jiang and Doudna, 2017). This process eventually results in the cleavage of homologous double-stranded DNA sequences known as protospacers in the invading DNA (Barrangou et al., 2007). Usually, the 5′-NGG-3′ PAM sequence is more preferred and frequently used than 5′-NAG-3' (Hsu et al., 2013).

After DSB, the cellular DNA repair pathways commence. The repair mechanism induces the cell to undertake homology-directed repair (HDR), microhomology-mediated end joining (MMEJ) or non-homologous end joining (NHEJ). HDR occurrence leads towards specific editing by a repair template-specific desired genomic modification (Vu et al., 2017). HDR-based strategies have proven difficult in plants because of a high preference for the NHEJ pathway. Various strategies to overcome this plant-specific limitation have been proposed (Vu et al., 2020). MMEJ/NHEJ is an error-prone repair system that involves the arrangement of micro homologous sequences internal to broken ends prior to joining and is coupled with insertions and deletions (Afzal et al., 2020). Polymerase theta-mediated end joining (TMEJ) is an advanced form of MMEJ/NHEJ, used as break repair with the homology of >1bp sequences (Schimmel et al., 2019). In the case of NHEJ, no DNA repair template is provided, and its error-prone nature often leads to inactivating mutations such as small deletions (Trenner and Sartori, 2019). In the case of a perfect repair, the target may simply undergo a new cycle of DSB and repair. Some other repair mechanisms also exist like the single-stranded annealing (SSA) pathway of HDR, which requires only a single DNA duplex and uses the repeat sequences as the identical sequences as in HDR (Cejka and Symington, 2021). Thus, all these methods are efficient tools for genome editing that might be insertion/deletion, replacement or knockout of desired genes in the cultivated grasses for improvement.

CRISPR-VARIANTS: OLD VS. LATEST

Target specificity is provided by the base complementarity of the protospacer motif of the guide RNA. However, the DNA region targeted for cleavage by the enzyme has to be followed by the appropriate PAM sequence. The prototypical Cas9 derived from *Streptococcus pyogenes* (SpCas9) needs a GC-enriched site PAM in the form of NGG (N = A/T/G/C) which limits the targeting flexibility. The presence of PAM restrains access to some potential sites due to which many precision edits of targeting sites encoding non-canonical PAM remained inaccessible (**Figure 1**). Notwithstanding, several efforts were made to increase the flexibility of target site recognition by Cas enzymes (Nishimasu et al., 2018; Miller et al., 2020), and new endonuclease enzyme variants were developed such as Cas12a (Chen et al., 2018) and Cas12b (Ming et al., 2020). Similar to Cas9, these variants also retain the limitation of PAM requirement by relying on T-enriched at the 5′-end of PAM in the form of TTTV (V = A/G/C) (**Figure 1**). Recently, Walton et al. (2020) successfully overcame the limitation by developing the structure-guided engineered variant of the SpCas9 enzyme, named SpRY targeting the genomic DNA with the independence of PAM restriction (nearly PAM-less qualities) (NRN > NYN, where R is A or G and where Y is C or T) in human cells. Very recently, Ren et al. (2021) explored the versatility of this improved genome-editing tool in plants for the first time and proved that SpRY targeted a total of 59 NNN PAM sites (NAN/NGN/NCN/NTN) in rice. Cas9 was demonstrated for being unable to edit relaxed PAM sites and less efficient in non-canonical PAM sites than SpRY, which possibly achieved larger deletions up to five base pairs at relaxed PAM sites which is impossible using Cas9. Elimination of the PAM requirement exposes the CRISPR-Cas T-DNA to self-editing thus introducing a risk of gRNA inactivation or modification (Ren et al., 2021).

Likewise, single base editing using a CRISPR-mediated genome editing tool has also been developed and applied in different cultivated grasses. For example, cytosine base editing (CBE) and Adenine Base editing (ABE) have been optimized for base editing in rice, wheat, and maize (Shimatani et al., 2017; Zong et al., 2017; Li C. et al., 2018; Zong et al., 2018). But, these were inefficient at some targets and several strategies have been used to enhance their efficiency monocots. Fortunately, it was proved that the SpRY-PmCDA1 (PAM-less C to T nucleotide editor) successfully converted the C-to-T base in rice (Ren et al., 2021). Thus, the expanded target range of this CRISPR-associated–SpRY enzyme harnessed the high accuracy of base (nucleotide-level) editing by using SpRY-based cytosine base editors (CBEs) in relaxed PAM (first to sixth base of protospacer). This was not possible in the traditional C-to-T base editors due to the specific distance requirement in the editing windows (Manghwar et al., 2019). On the other hand, the SpRY-based adenine base editor (ABE8e) also showed higher efficiency of A-to-G conversion with an editing window of fourth to eighth bases of the protospacer (Ren et al., 2021). Hence, a novel choice of base number edits is now possible in plants using SpRY-based CBEs and ABEs. In the toolbox of a CRISPR-based system, PAMs play a vital role as a specific uniform for Cas enzymes by differentiating them from non-self DNA sequences (Westra et al., 2013). Therefore, the application of the SpRY base CRISPR tool for PAM-less targeting raised a significant

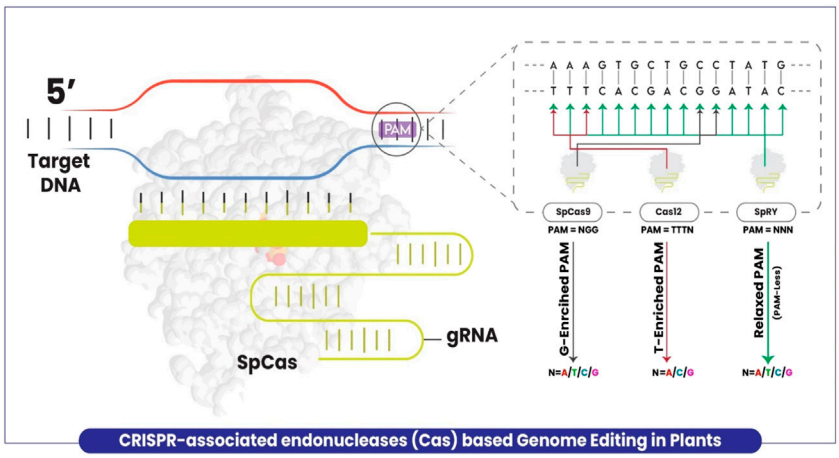


FIGURE 1 | A schematic model explaining the plant genome editing using CRISPR-associated enzymes with limitations and extra feature of SpRY as compared to others. Cas9 targets a G enriched site with PAM = NGG (Black arrow), Cas12 targets T enriched site with PAM = TTTV (Red arrow), whereas SpRY with no PAM restriction (Green arrow).

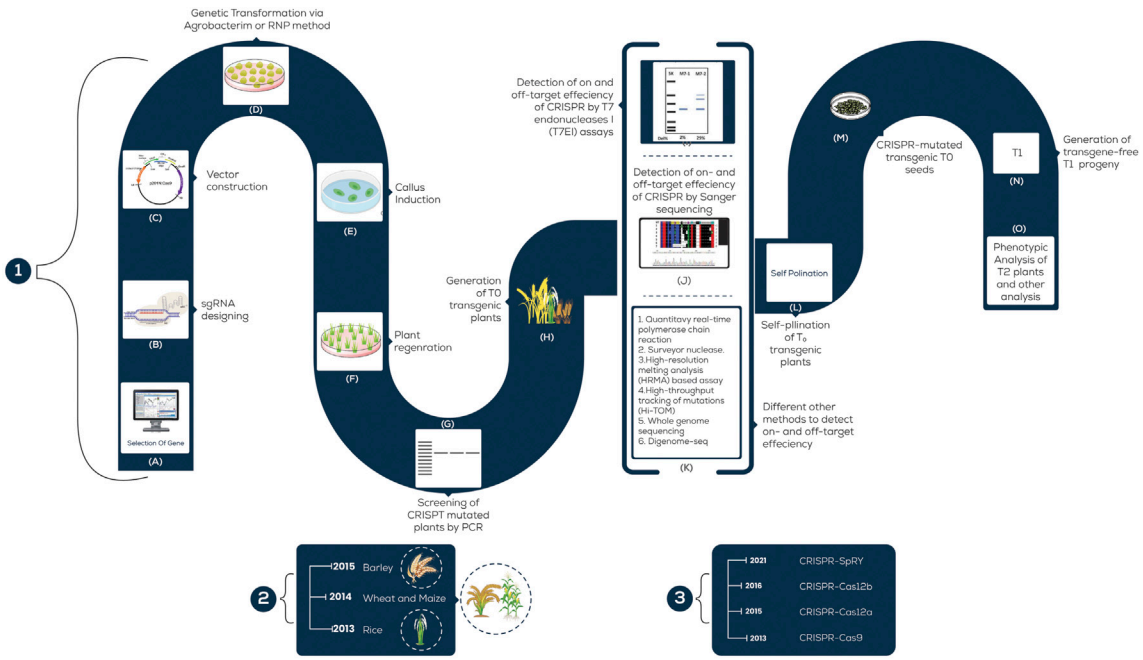


FIGURE 2 | A schematic illustration of the steps involved in CRISPR/Cas9 Genetic Transformation; 1; **(A)** Specific gene-targeted; **(B)** Designing sgRNA for the desired gene; **(C)** Vector; **(D)** Transformation of the CRISPR/Cas9 system; **(E)** Callus formation; **(F)** Regeneration of shoots from callus; **(G)** T0- Mutated plants; **(H)** Transgenic plants testing by PCR; **(I)** Identification of mutated plants by T7E1; **(J)** Screening of mutants by sequencing; **(K)** Various techniques to detect edited plants; **(L)** Self-pollination of T0 transgenic plants; **(M)** Mutated T0 seeds; **(N)** T1 progeny; **(O)** Phenotypic analysis of T2 plants. (2) The scale mentions the year in which each grass was employed for CRISPR-based genome editing. (3) The scale mentions the year in which each (major) CRISPR tool was developed and used in agriculture.

limitation of self-editing, which was suggested to utilize for secondary off-targeting of novel and unaccounted edits. The only one identified off-target in multiple T0 transgenic rice lines could be taken as promiscuity of *de novo* spacer from a self-targeting gRNA vector (Ren et al., 2021).

These outcomes and unclear shortcomings compel further investigation of structural engineering and application in different systems notably such self-editing was not reported/observed for single base editing in human cells using SpRY-ABEs (Walton et al., 2020).

TABLE 1 | Application of CRISPR associated genome editing in major cultivated grasses.

	Crop	Gene	Function	References
Abiotic Stress	Wheat	*TaDREB2*	Drought tolerance	Bhowmik et al. (2018)
	Rice	*OsMYB1*	Tolerance to various environmental stresses	Mao et al. (2013)
	Barley	*HvPM19*	Regulator of grain dormancy under stress	Lawrenson et al. (2015)
	Maize	*ZmHKT1*	Salt tolerance	Xing et al. (2014)
	Barley	*inositol-tetrakisphosphate 1-kinase*	Tolerance to salinity stress	VlckoOhnoutkova (2020)
	Rice	*OsARM1 and OsNramp5*	Heavy metal (cadmium and arsenic) resistance	Tang et al. (2017)
	Rice	*OsPYL*	Enhanced high-temperature tolerance	Miao et al. (2018)
Biotic stress	Wheat	*TaABCC6 ABC*	ABC transporter (ABCC6) associated with Fusarium head blight (FHB) susceptibility	Cui. (2017)
	Rice	*OsSWEET11*	Resistance against pathogens and enhance yield	Xing et al. (2014)
	Sorghum	*DsRED2*	Biotic and abiotic stresses	Xing et al. (2014)
	Barley	*HvMORC6a*	Oomycetes resistance	Galli et al. (2022)
	Rice	*OsWRKY93 and OsMORE1a*	Resistant to viral diseases such as tungro disease and fungal disease (Magnaporthe oryzae)	Macovei et al. (2018), Kim et al. (2022), Li et al. (2021)
	Maize	*ALS*	Herbicide resistance	Svitashev et al. (2015)
	Barley	*HvMORC1*	Resistance against *Fusarium graminearum*	Kumar et al. (2018)
	Rice	*eif4g*	Resistant to tungro disease	Macovei et al. (2018)
	Wheat	*TaNFXL1*	Resistance against *Fusarium graminearum*	Brauer et al. (2020)
Quality Yield	Wheat	*Alpha-gliadin*	Gluten protein	Brandt et al. (2017)
	Barley	*GST* and *IPI*	Recombinant protein accumulation	Panting et al. (2021)
	Maize	*ZmIPK*	Phytic acid biosynthesis	Liang et al. (2014), Svitashev et al. (2015)
	Sorghum	*Alpha-kafirin*	Improving lysine and digestibility in sorghum	Li et al. (2018a)
	Rice	*SBEIIb*	High levels of amylose content	Sun et al. (2017)
	Maize	*Wx1*	Higher yield	Waltz. (2016)
	Barley	*HcCKK1*	Higher number of grains	Holubova et al. (2018)
	Barley	*HvCKX1*	Convert hulled into naked grains leading to higher grain yield, improved brewing quality	(Gasparis et al. (2018), Meints and Hayes (2019)
	Maize	*LIG, MS26, MS45*	Male-sterility	Svitashev et al. (2015)

APPLICATION OF CRISPR TECHNOLOGY IN CULTIVATED GRASSES

Plants are exposed to different environmental stresses such as microbes or climatic changes which are referred to as biotic and abiotic stresses, respectively. Both of these (biotic and abiotic) cause almost 50% yield loss globally (Thabet and Alqudah, 2019; Alqudah et al., 2020). The forthcoming discussion presents a brief account of CRISPR application in cultivated grasses with some recent examples listed in **Figure 2** and **Table 1** and its potential to increase the quality as well as to combat the losses caused by stress.

Abiotic Stress Resistance

It is important to note that abiotic stresses, such as drought (water shortage), flooding (hypoxia), salinity, heavy metals, temperature (hot and cold), and their interactions are the major factors hindering agricultural production. Using a variety of breeding approaches, different genes/pathways and regulatory networks involved in stress responses have been determined. To epitome, CRISPR-based genome editing has broadened the target of the biologists to activate or suppress the targeted genes involved in plant abiotic stress resistance (Osakabe et al., 2016). For example, heavy metal (cadmium and arsenic) resistance in rice plants was developed by knocking out *OsARM1* and *OsNramp5* genes using CRISPR-based gene editing (Tang et al., 2017; Wang et al., 2017). In wheat, two drought-associated genes, *TaDREB2* (dehydration responsive element binding protein 2) and *TaERF3* (ethylene-responsive factor 3) have been successfully targeted in the protoplast using CRISPR-Cas (Kim et al., 2018). Similarly, in comparison with wild-type, rice with CRISPR-Cas editing of the OsPYL abscisic acid receptor gene family showed enhanced high-temperature tolerance (Miao et al., 2018). An elevated expression of ARGOS8 (an auxin-related gene involved in organ size 8) in maize plants using CRISPR, helped to improve the drought tolerance in maize (Shi et al., 2017). Likewise, barley itpk1 (inositol-tetrakisphosphate 1-kinase) mutants were developed using a CRISPR-based genome editing tool to validate the gene function of *HvITPK1*, and insertion mutant lines revealed a higher tolerance to salinity stress than deletion mutants (Vlcko and Ohnoutkova, 2020). CRISPR-Cas 9-mediated *OsRR22* gene editing was applied to improve salt stress tolerance in rice (Zhang et al., 2019). Based on previously reported studies, precise base editing is one of the most suitable techniques which can be employed to develop mutants with loss or gain of function to develop stress-tolerant varieties. Similarly, some reports revealed the regulation of abiotic stress in plants by cis-regulatory sequences (Liu et al., 2014), so novel promoter variants can also be created to produce useful novel

phenotypic variation and new quantitative trait loci (QTLs) by 'gain-of-function' mutation for various traits associated with abiotic stress tolerance.

Biotic Stress Resistance

Plants are affected to varying degrees by biotic stress, among other environmental stresses. It has been possible with CRISPR-Cas -based genome editing to engineer crops resistant to bacterial, fungal, and viral diseases as well as oomycetes. For instance, rice, wheat, maize, and barley have seen great success in the ability to increasing resistance to powdery mildews, bacterial blights, and blast diseases (Chen et al., 2019; Mushtaq et al., 2021; Ali et al., 2022). A study conducted by Wang et al. (2014) developed a powdery mildew resistant wheat by disrupting the *TAMLOA1*, *TAMLOA2*, and *TAMLOA3* genes in the wheat genome using the CRISPR-Cas 9 system. In barley plants, the HvMORC1 gene is silenced via CRISPR-Cas 9 which led to an increase in the resistance against *Fusarium graminearum* (Kumar et al., 2018). Recently, CRISPR-mediated genome editing of the TaNFXL1 (Resistant; R) gene in wheat led to the enhanced resistance against *Fusarium graminearum* (Brauer et al., 2020). CRISPR-mediated genome engineering has also helped to produce eif4g rice that is resistant to viral diseases such as tungro disease (Macovei et al., 2018) and fungal disease (Magnaporthe oryzae) by genetic functional validation of OsWRKY93 and OsMORE1a gene (Li et al., 2021; Kim et al., 2022). A study about oomycetes resistance in barley investigated the functional genomics of *HvMORC6a* gene in barley by using the application of CRISPR technology (Galli et al., 2022). Although, genome editing has excelled with the application of CRISPR-based tools. But, there is still a huge gap to create disease-resistant germplasm which could be easier by targeting R and S-genes as well as their orthologues in other species. The optimization and reprogramming of CRISPR components could be useful to establish resistance against such biotic stresses for which no natural resistance is found.

Improving Quality and Yield

Until now, the use of genome editing has had a positive effect on improving quality attributes such as starch content, fragrance, nutritional value, and storage durability in crop plants. The CRISPR-Cas9 system has been used to modify the starch branching enzyme gene SBEIIb to develop rice with high levels of amylose content with such nutritional properties of starch that benefit patients suffering from noninfectious chronic diseases related to diet and carbohydrates. (Sun et al., 2017). New ways of altering traits regulated by large, redundant gene families are being explored by CRISPR-Cas9 e.g., α-gliadin gene family, the major gluten encoding gene family in wheat that consists of more than 100 genes. Researchers have created low-gluten wheat by simultaneously knocking out the most conserved domains of α-gliadin family members that ultimately help to avoid celiac disease (Sanchez-Leon et al., 2018).

As a result of CRISPR-Cas technology, a company in the United States was able to mutate the waxy gene Wx1 to give higher yield maize for commercial use (Waltz 2016). In barley, the role of the cytokinin dehydrogenase enzyme (CKX1) was explored by silencing the HcCKK1 gene through CRISPR-Cas 9 which resulted in a higher number of grains in transgenic lines (Holubova et al., 2018). Similarly, CRISPR-mediated genome editing enabled the researchers to convert hulled into naked barley grains leading to higher grain yield and improved brewing quality (Gasparis et al., 2018; Meints and Hayes, 2019). The expression level of pro-nutritional phytase was modulated by targeting the barley PAPhy_a promoter (Holme et al., 2017). In cultivated grasses, many studies reported that the presence of various compounds with anticancer, anti-inflammatory, and anti-microbial properties such as anthocyanin in rice promotes growth and enhances environmental stress (Mackon et al., 2021), and lunasin peptide in barley, wheat and rye seeds possess chemo-preventive properties (Hernandez-Ledesmade et al., 2008; Nakurte et al., 2013). The application of CRISPR/Cas technology can be employed to target transcription factors regulating such compounds leading to quality improvement as well as yield.

CRISPR-CAS BASED miRNA-EDITING FOR CROP IMPROVEMENT

Micro RNA (miRNA) are small RNA molecules with complementary binding sites in target mRNAs, representing a promising avenue to control complex traits since miRNAs can precisely down-regulate any number of co-expressed target transcripts and their respective pathways (Tang and Chu, 2017). Interestingly, most of the miRNA-targeted transcripts encode transcription factors (TFs), which themselves often act as crucial hubs of developmental regulation. It should be noted that using SpRY, knocking out miRNAs regulated genes or engineering quantitative trait variation can be more efficient than Cas9 (Zhou et al., 2017). Recent research has revealed key functions for miRNAs in controlling crop plant agronomic traits including cereals. For instance, in rice, increased expression of the miR156-targeted *OsSPL14* is associated with improved grain yield, characterized by reduced vegetative branching (tillering) and increased panicle branching (Jiao et al., 2010; Miura et al., 2010). Further, the free threshing trait in wheat is caused by SNPs in the miRNA172 binding site of the major domestication gene Q (Liu et al., 2018). Taken together, integrating knowledge about miRNAs with advanced cereal molecular genetics techniques like CRISPR is a promising strategy for crop improvement. The small sequence size of miRNA sites makes loss-of-function mutation production is difficult to achieve. Thus, precise mutation developed by CRISPR-Cas9 is an excellent option to knockout miRNA small sequence in rice (Basso et al., 2019; Bi et al., 2020). Moreover, targeting the sequence prior to the miRNA region by CRISPR-Cas maximizes the chance of miRNA knockout. Chung et al. (2020) demonstrated that such an approach to creating a deletion within the pre-miRNA regions by CRISPR-rCas9 was efficient in knockout miRNAs in rice. Therefore, CRISPR can be applied to elucidate the function of miRNA-regulated genes in cultivated grasses and their relevance towards the improvement of

agronomic traits. Thus, based on the given proofs-of-concept, we propose broadening the scope of CRISPR in miRNA-regulated genes in cultivated grasses for a deeper understanding of their function and role in crop improvement.

CONCLUSION AND FUTURE PERSPECTIVE

CRISPR-Cas represents the most recent development in genome engineering which has revolutionized crop breeding since 2013. With CRISPR-Cas, genome editing has become a relatively simple, low-cost, and robust process, resulting in huge advances in crop improvement.

Regulation of GMOs

Producing disease-resistant and environment-adapted crops, as well as improving yields and quality, are the main applications of CRISPR technology in agriculture. One noteworthy example is that preassembled CRISPR-Cas9 ribonucleoproteins were delivered DNA-free into plant protoplasts of rice and wheat (Woo et al., 2015; Zhang et al., 2016). Thus, crops developed by DNA-free technology may be considered non-GM crops in some countries where GM crops are not allowed or are under prohibitive approval requirements. This would allow the development of higher-quality grains with better phenotypes that could be commercialized and sold. The European Union, however, does not permit ribonucleoprotein mutated crops without GMO approval (Gelinsky and Hilbeck, 2018).

Well Suited Cas9 and SpRY Options

CRISPR/Cas as a powerful tool led to tremendous advances in crop improvement through precise knockout, knock-in, replacement, point mutations, and fine-tuning of any gene. A potential method to attenuate on-target editing and circumvent the vector self-editing, which gives rise to in activation, off-targeting, should be further explored with a clearer understanding of the characteristics underlying SpRY. Although Cas9 is not capable of PAM-less editing, however, it proved itself better than SpRY for canonical NGG PAM site editing, which reveals the unterminated importance of Cas9 (Ren et al., 2021). To date, SpRY is a choice for more fine exploration of plant genome and its application in rice plants will be rolled as the first confident presentation of unconstrained targeting with nearly PAM-less editing in monocots (rice) and dicots also as Ren et al. (2021) proved that SpRY can be used in Dahurian larch. It will inspire many exciting investigations such as the expansion of *in vivo* directed evolution efforts to improve other plant characteristics against resistance for high yielding of important agronomic traits to ensure sustainable food security.

AUTHOR CONTRIBUTIONS

Conceptualization; AR, FK, and AMA, writing—original draft preparation; AR, FK, AMA, and AF, writing—review and editing, AR, CM, HB-P, ZB, AMA and SK, visualization, AR and SA, All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was part of the Novo Nordisk Foundation projects GRAINY "NNF20OC0064295" for AMA and NovoCrops "NNF19OC0056580" for HB-P and CM. ZB was funded by the Innovation Fund Denmark grant 8055-00038B, ReTraQue.

ACKNOWLEDGMENTS

We are thankful to Javaria Tabusam to help in the compilation of literature for the first draft, Marek Marzek for his valuable comments to improve the first draft of the manuscript.

REFERENCES

Afzal, S., Sirohi, P., and Singh, N. K. (2020). A Review of Crispr Associated Genome Engineering: Application, Advances and Future Prospects of Genome Targeting Tool for Crop Improvement. *Biotechnol. Lett.* 42, 1611–1632. doi:10.1007/s10529-020-02950-w

Ali, Q., Yu, C. J., Hussain, A., Ali, M., Ahmar, S., Sohail, M. A., et al. (2022). Genome Engineering Technology for Durable Disease Resistance: Recent Progress and Future Outlooks for Sustainable Agriculture. *Front. Plant Sci.* 13, 860281. doi:10.3389/fpls.2022.860281

Alqudah, A. M., Sallam, A., Stephen Baenziger, P. S., and Börner, A. (2020). Gwas: Fast-Forwarding Gene Identification and Characterization in Temperate Cereals: Lessons from Barley - a Review. *J. Adv. Res.* 22, 119–135. doi:10.1016/j.jare.2019.10.013

Barrangou, R. (2013). CRISPR-Cas Systems and RNA-Guided Interference. *WIREs RNA* 4, 267–278. doi:10.1002/wrna.1159

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). Crispr Provides Acquired Resistance against Viruses in Prokaryotes. *Science* 315, 1709–1712. doi:10.1126/science.1138140

Basso, M. F., Ferreira, P. C. G., Kobayashi, A. K., Harmon, F. G., Nepomuceno, A. L., Molinari, H. B. C., et al. (2019). Micro Rna S and New Biotechnological Tools for its Modulation and Improving Stress Tolerance in Plants. *Plant Biotechnol. J.* 17, 1482–1500. doi:10.1111/pbi.13116

Bhowmik, P., Ellison, E., Polley, B., Bollina, V., Kulkarni, M., Ghanbarnia, K., et al. (2018). Targeted Mutagenesis in Wheat Microspores Using CRISPR/Cas9. *Sci. Rep.* 8, 6502. doi:10.1038/s41598-018-24690-8

Bi, H., Fei, Q., Li, R., Liu, B., Xia, R., Char, S. N., et al. (2020). Disruption of Mirna Sequences by Talens and Crispr/cas9 Induces Varied Lengths of Mirna Production. *Plant Biotechnol. J.* 18, 1526–1536. doi:10.1111/pbi.13315

Bortesi, L., and Fischer, R. (2015). The Crispr/cas9 System for Plant Genome Editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi:10.1016/j.biotechadv.2014.12.006

Brandt, K. M., Gunn, H., Buschke, B. L., Heesacker, A., Moretti, N., Karasev, A., et al. (2017). *Testing Non-transgenic Crispr Technology for Wheat Improvement*. Austria: Presentation. 13th IWGS–Tulln.

Brauer, E. K., Balcerzak, M., Rocheleau, H., Leung, W., Schernthaner, J., Subramaniam, R., et al. (2020). Genome Editing of a Deoxynivalenol-Induced Transcription Factor Confers Resistance to fusarium Graminearum in Wheat. *Mpmi* 33, 553–560. doi:10.1094/mpmi-11-19-0332-r

Cejka, P., and Symington, L. S. (2021). DNA End Resection: Mechanism and Control. *Annu. Rev. Genet.* 55, 285–307. doi:10.1146/annurev-genet-071719-020312

Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., et al. (2018). Crispr-cas12a Target Binding Unleashes Indiscriminate Single-Stranded Dnase Activity. *Science* 360, 436–439. doi:10.1126/science.aar6245

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). Crispr/cas Genome Editing and Precision Plant Breeding in Agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi:10.1146/annurev-arplant-050718-100049

Chung, P. J., Chung, H., Oh, N., Choi, J., Bang, S. W., Jung, S. E., et al. (2020). Efficiency of Recombinant Crispr/rcas9-Mediated Mirna Gene Editing in Rice. *Ijms* 21, 9606. doi:10.3390/ijms21249606

Cram, D., Kulkarni, M., Buchwaldt, M., Rajagopalan, N., Bhowmik, P., Rozwadowski, K., et al. (2019). Wheatcrispr: A Web-Based Guide Rna Design Tool for Crispr/cas9-Mediated Genome Editing in Wheat. *BMC Plant Biol.* 19, 474. doi:10.1186/s12870-019-2097-z

Cui, X. (2017). Ottawa, Canada: Université d'Ottawa. Master of Science in Biology.Targeted Gene Editing Using Crispr/cas9 in a Wheat Protoplast System

Fineran, P. C., and Charpentier, E. (2012). Memory of Viral Infections by Crispr-Cas Adaptive Immune Systems: Acquisition of New Information. *Virology* 434, 202–209. doi:10.1016/j.virol.2012.10.003

Galli, M., Martiny, E., Imani, J., Kumar, N., Koch, A., Steinbrenner, J., et al. (2022). CRISPR/Sp Cas9-mediated Double Knockout of Barley Microrchidia MORC1 and MORC6a Reveals Their Strong Involvement in Plant Immunity, Transcriptional Gene Silencing and Plant Growth. *Plant Biotechnol. J.* 20, 89–102. doi:10.1111/pbi.13697

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crrna Ribonucleoprotein Complex Mediates Specific DNA Cleavage for Adaptive Immunity in Bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2579–E2586. doi:10.1073/pnas.1208507109

Gasparis, S., Kała, M., Przyborowski, M., Łyżnik, L. A., Orczyk, W., and Nadolska-Orczyk, A. (2018). A Simple and Efficient Crispr/cas9 Platform for Induction of Single and Multiple, Heritable Mutations in Barley (Hordeum Vulgare l.). *Plant Methods* 14, 111. doi:10.1186/s13007-018-0382-8

Gelinsky, E., and Hilbeck, A. (2018). European Court of Justice Ruling Regarding New Genetic Engineering Methods Scientifically Justified: A Commentary on the Biased Reporting About the Recent Ruling. *Environ. Sci. Eur.* 30, 52. doi:10.1186/s12302-018-0182-9

Han, Y., Broughton, S., Liu, L., Zhang, X. Q., Zeng, J., He, X., et al. (2021). Highly Efficient and Genotype-independent Barley Gene Editing Based on Anther Culture. *Plant Commun.* 2, 100082. doi:10.1016/j.xplc.2020.100082

Hernández-Ledesma, B., de Lumen, C-C., and de Lumen, B. O. (2008). Lunasin: A Novel Cancer Preventive Seed Peptide. *Perspect. Med. Chem.* 2, 75–80. doi:10.4137/pmc.s372

Holme, I. B., Wendt, T., Gil-Humanes, J., Deleuran, L. C., Starker, C. G., Voytas, D. F., et al. (2017). Evaluation of the Mature Grain Phytase Candidate Hvpaphy_a Gene in Barley (Hordeum Vulgare l.) Using Crispr/cas9 and Talens. *Plant Mol. Biol.* 95, 111–121. doi:10.1007/s11103-017-0640-6

Holubová, K., Hensel, G., Vojta, P., Tarkowski, P., Bergougnoux, V., and Galuszka, P. (2018). Modification of Barley Plant Productivity through Regulation of Cytokinin Content by Reverse-Genetics Approaches. *Front. Plant Sci.* 9, 1676. doi:10.3389/fpls.2018.01676

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA Targeting Specificity of Rna-Guided Cas9 Nucleases. *Nat. Biotechnol.* 31, 827–832. doi:10.1038/nbt.2647

Jiang, F., and Doudna, J. A. (2017). Crispr-cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* 46, 505–529. doi:10.1146/annurev-biophys-062215-010822

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., et al. (2010). Regulation of Ossp14 by Osmir156 Defines Ideal Plant Architecture in Rice. *Nat. Genet.* 42, 541–544. doi:10.1038/ng.591

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A Programmable Dual-Rna-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337, 816–821. doi:10.1126/science.1225829

Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P., and Siksnys, V. (2013). crRNA and tracrRNA Guide Cas9-Mediated DNA Interference inStreptococcus Thermophilus. *RNA Biol.* 10, 841–851. doi:10.4161/rna.24203

Kim, C. Y., Park, J. Y., Choi, G., Kim, S., Vo, K. T. X., Jeon, J. S., et al. (2022). A Rice Gene Encoding Glycosyl Hydrolase Plays Contrasting Roles in Immunity Depending on the Type of Pathogens. *Mol. Plant Pathol.* 23, 400–416. doi:10.1111/mpp.13167

Kim, D., Alptekin, B., and Budak, H. (2018). Crispr/cas9 Genome Editing in Wheat. *Funct. Integr. Genomics* 18, 31–41. doi:10.1007/s10142-017-0572-x

Kim, Y-A., Moon, H., and Park, C-J. (2019). CRISPR/Cas9-targeted Mutagenesis of Os8N3 in Rice to Confer Resistance to Xanthomonas Oryzae Pv. Oryzae. *Rice (N Y)* 12, 67–13. doi:10.1186/s12284-019-0325-7

Koeppel, I., Hertig, C., Hoffie, R., and Kumlehn, J. (2019). Cas Endonuclease Technology-A Quantum Leap in the Advancement of Barley and Wheat Genetic Engineering. *Ijms* 20, 2647. doi:10.3390/ijms20112647

Kumar, N., Galli, M., Ordon, J., Stuttmann, J., Kogel, K. H., and Imani, J. (2018). Further Analysis of Barley Morc1 Using a Highly Efficient Rna-Guided Cas9 Gene-Editing System. *Plant Biotechnol. J.* 16, 1892–1903. doi:10.1111/pbi.12924

Lawrenson, T., Hinchliffe, A., Clarke, M., Morgan, Y., and Harwood, W. (2021). In-planta Gene Targeting in Barley Using Cas9 with and without Geminiviral Replicons. *Front. Genome Ed.* 3, 663380. doi:10.3389/fgeed.2021.663380

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., et al. (2015). Induction of Targeted, Heritable Mutations in Barley and brassica Oleracea Using Rna-Guided Cas9 Nuclease. *Genome Biol.* 16, 258–313. doi:10.1186/s13059-015-0826-7

Li, A., Jia, S., Yobi, A., Ge, Z., Sato, S. J., Zhang, C., et al. (2018a). Editing of an Alpha-Kafirin Gene Family Increases, Digestibility and Protein Quality in Sorghum. *Plant Physiol.* 177, 1425–1438. doi:10.1104/pp.18.00200

Li, C., Zong, Y., Wang, Y. P., Jin, S., Zhang, D. B., Song, Q. N., et al. (2018b). Expanded Base Editing in Rice and Wheat Using a Cas9-Adenosine Deaminase Fusion. *Genome Biol.* 19, 59–9. doi:10.1186/s13059-018-1443-z

Li, H. (2015). Structural Principles of Crispr Rna Processing. *Structure* 23, 13–20. doi:10.1016/j.str.2014.10.006

Li, J., Li, Y., and Ma, L. G. (2019). Crispr/cas9-based Genome Editing and its Applications for Functional Genomic Analyses in Plants. *Small Methods* 3, 1800473. doi:10.1002/smtd.201800473

Li, Y. Y., Liao, S. T., Mei, P. Y., Pan, Y. Y., Zhang, Y., Zheng, X. Z., et al. (2021). Oswrky93 Dually Functions between Leaf Senescence and in Response to Biotic Stress in Rice. *Front. Plant Sci.* 12, 643011. doi:10.3389/fpls.2021.643011

Liang, Z., Zhang, K., Chen, K., and Gao, C. (2014). Targeted Mutagenesis in Zea Mays Using Talens and the Crispr/cas System. *J. Genet. Genomics* 41, 63–68. doi:10.1016/j.jgg.2013.12.001

Liu, H. J., Jian, L., Xu, J., Zhang, Q., Zhang, M., Jin, M., et al. (2020a). High-throughput Crispr/cas9 Mutagenesis Streamlines Trait Gene Identification in Maize. *Plant Cell.* 32, 1397–1413. doi:10.1105/tpc.19.00934

Liu, J-H., Peng, T., and Dai, W. (2014). Critical Cis-Acting Elements and Interacting Transcription Factors: Key Players Associated with Abiotic Stress Responses in Plants. *Plant Mol. Biol. Rep.* 32, 303–317. doi:10.1007/s11105-013-0667-z

Liu, P., Liu, J., Dong, H., and Sun, J. (2018). Functional Regulation of Q by Microrna172 and Transcriptional Co-repressor Topless in Controlling Bread Wheat Spikelet Density. *Plant Biotechnol. J.* 16, 495–506. doi:10.1111/pbi.12790

Liu, Z. Q., Dong, H. N., Cui, Y. L., Cong, L. N., and Zhang, D. W. (2020b). Application of Different Types of Crispr/cas-Based Systems in Bacteria. *Microb. Cell. Fact.* 19, 172–214. doi:10.1186/s12934-020-01431-z

Mackon, E., Mackon, G. C. J. D. E., Ma, Y. F., Kashif, M. H., Ali, N., Usman, B., et al. (2021). Recent Insights into Anthocyanin Pigmentation, Synthesis, Trafficking, and Regulatory Mechanisms in Rice (Oryza Sativa l.) Caryopsis. *Biomolecules* 11, 394. doi:10.3390/biom11030394

Macovei, A., Sevilla, N. R., Cantos, C., Jonson, G. B., Slamet-Loedin, I., Čermák, T., et al. (2018). Novel Alleles of Rice eIF4G Generated by CRISPR/Cas9-targeted Mutagenesis Confer Resistance to Rice Tungro Spherical Virus. *Plant Biotechnol. J.* 16, 1918–1927. doi:10.1111/pbi.12927

Mahfouz, M. M., Piatek, A., and Stewart, C. N., Jr (2014). Genome Engineering via Talens and Crispr/cas9 Systems: Challenges and Perspectives. *Plant Biotechnol. J.* 12, 1006–1014. doi:10.1111/pbi.12256

Makarova, K. S., Zhang, F., and Koonin, E. V. (2017). Snapshot: Class 1 Crispr-Cas Systems. *Cell.* 168, 946. doi:10.1016/j.cell.2017.02.018

Manghwar, H., Lindsey, K., Zhang, X., and Jin, S. (2019). Crispr/cas System: Recent Advances and Future Prospects for Genome Editing. *Trends Plant Sci.* 24, 1102–1125. doi:10.1016/j.tplants.2019.09.006

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F., and Zhu, J. K. (2013). Application of the Crispr-Cas System for Efficient Genome Engineering in Plants. *Mol. Plant* 6, 2008–2011. doi:10.1093/mp/sst121

Meints, B., and Hayes, P. M. (2019). Breeding Naked Barley for Food, Feed, and Malt. *Plant Breed. Rev.* 43, 95–119. doi:10.1002/9781119616801.ch4

Miao, C., Xiao, L., Hua, K., Zou, C., Zhao, Y., Bressan, R. A., et al. (2018). Mutations in a Subfamily of Abscisic Acid Receptor Genes Promote Rice Growth and Productivity. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6058–6063. doi:10.1073/pnas.1804774115

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., et al. (2011). A Tale Nuclease Architecture for Efficient Genome Editing. *Nat. Biotechnol.* 29, 143–148. doi:10.1038/nbt.1755

Miller, S. M., Wang, T., Randolph, P. B., Arbab, M., Shen, M. W., Huang, T. P., et al. (2020). Continuous Evolution of Spcas9 Variants Compatible with Non-g Pams. *Nat. Biotechnol.* 38, 471–481. doi:10.1038/s41587-020-0412-8

Ming, M., Ren, Q., Pan, C., He, Y., Zhang, Y., Liu, S., et al. (2020). Crispr-cas12b Enables Efficient Plant Genome Engineering. *Nat. Plants* 6, 202–208. doi:10.1038/s41477-020-0614-6

Miura, K., Ikeda, M., Matsubara, A., Song, X. J., Ito, M., Asano, K., et al. (2010). Osspl14 Promotes Panicle Branching and Higher Grain Productivity in Rice. *Nat. Genet.* 42, 545–549. doi:10.1038/ng.592

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short Motif Sequences Determine the Targets of the Prokaryotic Crispr Defence System. *Microbiol. Read.* 155, 733–740. doi:10.1099/mic.0.023960-0

Mushtaq, M., Dar, A. A., Basu, U., Bhat, B. A., Mir, R. A., Vats, S., et al. (2021). Integrating Crispr-Cas and Next Generation Sequencing in Plant Virology. *Front. Genet.* 12, 735489. doi:10.3389/fgene.2021.735489

Nakurte, I., Kirhnere, I., Namniece, J., Saleniece, K., Krigere, L., Mekss, P., et al. (2013). Detection of the Lunasin Peptide in Oats (Avena Sativa L). *J. cereal Sci.* 57, 319–324. doi:10.1016/j.jcs.2012.12.008

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., et al. (2014). Crystal Structure of Cas9 in Complex with Guide Rna and Target DNA. *Cell.* 156, 935–949. doi:10.1016/j.cell.2014.02.001

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., et al. (2018). Engineered Crispr-Cas9 Nuclease with Expanded Targeting Space. *Science* 361, 1259–1262. doi:10.1126/science.aas9129

Osakabe, Y., Watanabe, T., Sugano, S. S., Ueta, R., Ishihara, R., Shinozaki, K., et al. (2016). Optimization of Crispr/cas9 Genome Editing to Modify Abiotic Stress Responses in Plants. *Sci. Rep.* 6, 26685. doi:10.1038/srep26685

Panting, M., Holme, I. B., Björnsson, J. M., Zhong, Y., and Brinch-Pedersen, H. (2021). Crispr/cas9 and Transgene Verification of Gene Involvement in Unfolded Protein Response and Recombinant Protein Production in Barley Grain. *Front. Plant Sci.* 12, 755788. doi:10.3389/fpls.2021.755788

Ren, Q., Sretenovic, S., Liu, S., Tang, X., Huang, L., He, Y., et al. (2021). Pam-less Plant Genome Editing Using a Crispr-Spry Toolbox. *Nat. Plants* 7, 25–33. doi:10.1038/s41477-020-00827-4

Riaz, A., Kanwal, F., Börner, A., Pillen, K., Dai, F., and Alqudah, A. M. (2021). Advances in Genomics-Based Breeding of Barley: Molecular Tools and Genomic Databases. *Agronomy* 11, 894. doi:10.3390/agronomy11050894

Ronald, P. (2011). Plant Genetics, Sustainable Agriculture and Global Food Security. *Genetics* 188, 11–20. doi:10.1534/genetics.111.128553

Sánchez-León, S., Gil-Humanes, J., Ozuna, C. V., Giménez, M. J., Sousa, C., Voytas, D. F., et al. (2018). Low-gluten, Nontransgenic Wheat Engineered with Crispr/cas9. *Plant Biotechnol. J.* 16, 902–910. doi:10.1111/pbi.12837

Schimmel, J., van Schendel, R., den Dunnen, J. T., and Tijsterman, M. (2019). Templated Insertions: A Smoking Gun for Polymerase Theta-Mediated End Joining. *Trends Genet.* 35, 632–644. doi:10.1016/j.tig.2019.06.001

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted Genome Modification of Crop Plants Using a Crispr-Cas System. *Nat. Biotechnol.* 31, 686–688. doi:10.1038/nbt.2650

Shi, J., Gao, H., Wang, H., Lafitte, H. R., Archibald, R. L., Yang, M., et al. (2017). ARGOS8 Variants Generated by CRISPR-Cas9 Improve Maize Grain Yield under Field Drought Stress Conditions. *Plant Biotechnol. J.* 15, 207–216. doi:10.1111/pbi.12603

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., et al. (2017). Targeted Base Editing in Rice and Tomato Using a Crispr-Cas9 Cytidine Deaminase Fusion. *Nat. Biotechnol.* 35, 441–443. doi:10.1038/nbt.3833

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

Svitashev, S., Young, J. K., Schwartz, C., Gao, H., Falco, S. C., and Cigan, A. M. (2015). Targeted Mutagenesis, Precise Gene Editing, and Site-specific Gene Insertion in Maize Using Cas9 and Guide Rna. *Plant Physiol.* 169, 931–945. doi:10.1104/pp.15.00793

Tang, J., and Chu, C. (2017). Micrornas in Crop Improvement: Fine-tuners for Complex Traits. *Nat. Plants* 3, 17077. doi:10.1038/nplants.2017.77

Tang, L., Mao, B., Li, Y., Lv, Q., Zhang, L., Chen, C., et al. (2017). Knockout of Osnramp5 Using the Crispr/cas9 System Produces Low Cd-Accumulating Indica Rice without Compromising Yield. *Sci. Rep.* 7, 14438. doi:10.1038/s41598-017-14832-9

Thabet, S. G., and Alqudah, A. M. (2019). Crops and Drought. *eLS*, 1–8. doi:10.1002/9780470015902.a0025265

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global Food Demand and the Sustainable Intensification of Agriculture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20260–20264. doi:10.1073/pnas.1116437108

Trenner, A., and Sartori, A. A. (2019). Harnessing DNA Double-Strand Break Repair for Cancer Treatment. *Front. Oncol.* 9, 1388. doi:10.3389/fonc.2019.01388

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome Editing with Engineered Zinc Finger Nucleases. *Nat. Rev. Genet.* 11, 636–646. doi:10.1038/nrg2842

Usman, B., Nawaz, G., Zhao, N., Liao, S., Qin, B., Liu, F., et al. (2020). Programmed Editing of Rice (Oryza Sativa l.) Osspl16 Gene Using Crispr/cas9 Improves Grain Yield by Modulating the Expression of Pyruvate Enzymes and Cell Cycle Proteins. *Ijms* 22, 249. doi:10.3390/ijms22010249

van Dijk, M., Morley, T., Rau, M. L., and Saghai, Y. (2021). A Meta-Analysis of Projected Global Food Demand and Population at Risk of Hunger for the Period 2010-2050. *Nat. Food* 2, 494–501. doi:10.1038/s43016-021-00322-9

Vlcko, T., and Ohnoutkova, L. (2020). Allelic Variants of Crispr/cas9 Induced Mutation in an Inositol Trisphosphate 5/6 Kinase Gene Manifest Different Phenotypes in Barley. *Plants (Basel)* 9, 195. doi:10.3390/plants9020195

Vu, G. T. H., Cao, H. X., Fauser, F., Reiss, B., Puchta, H., and Schubert, I. (2017). Endogenous Sequence Patterns Predispose the Repair Modes of CRISPR/Cas9-induced DNA Double-stranded Breaks in *Arabidopsis thaliana*. *Plant J.* 92, 57–67. doi:10.1111/tpj.13634

Vu, T. V., Sivankalyani, V., Kim, E. J., Doan, D. T. H., Tran, M. T., Kim, J., et al. (2020). Highly Efficient Homology-directed Repair Using CRISPR/Cpf1-geminiviral Replicon in Tomato. *Plant Biotechnol. J.* 18, 2133–2143. doi:10.1111/pbi.13373

Walton, R. T., Christie, K. A., Whittaker, M. N., and Kleinstiver, B. P. (2020). Unconstrained Genome Targeting with Near-Pamless Engineered Crispr-Cas9 Variants. *Science* 368, 290–296. doi:10.1126/science.aba8853

Waltz, E. (2016). Crispr-edited Crops Free to Enter Market, Skip Regulation. *Nat. Biotechnol.* 34, 582. doi:10.1038/nbt0616-582

Wang, F. Z., Chen, M. X., Yu, L. J., Xie, L. J., Yuan, L. B., Qi, H., et al. (2017). Osarm1, an R2r3 Myb Transcription Factor, Is Involved in Regulation of the Response to Arsenic Stress in Rice. *Front. Plant Sci.* 8, 1868. doi:10.3389/fpls.2017.01868

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous Editing of Three Homoeoalleles in Hexaploid Bread Wheat Confers Heritable Resistance to Powdery Mildew. *Nat. Biotechnol.* 32, 947–951. doi:10.1038/nbt.2969

Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., et al. (2013). Type I-E CRISPR-Cas Systems Discriminate Target from Non-target DNA through Base Pairing-Independent PAM Recognition. *PLoS Genet.* 9, e1003742. doi:10.1371/journal.pgen.1003742

Woo, J. W., Kim, J., Kwon, S. I., Corvalán, C., Cho, S. W., Kim, H., et al. (2015). DNA-Free Genome Editing in Plants with Preassembled Crispr-Cas9 Ribonucleoproteins. *Nat. Biotechnol.* 33, 1162–1164. doi:10.1038/nbt.3389

Xing, H. L., Dong, L., Wang, Z. P., Zhang, H. Y., Han, C. Y., Liu, B., et al. (2014). A Crispr/cas9 Toolkit for Multiplex Genome Editing in Plants. *BMC Plant Biol.* 14, 327. doi:10.1186/s12870-014-0327-y

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I. M., Li, Y., et al. (2016). Crystal Structure of Cpf1 in Complex with Guide Rna and Target DNA. *Cell.* 165, 949–962. doi:10.1016/j.cell.2016.04.003

Zhang, A., Liu, Y., Wang, F., Li, T., Chen, Z., Kong, D., et al. (2019). Enhanced Rice Salinity Tolerance via Crispr/cas9-Targeted Mutagenesis of the Osrr22 Gene. *Mol. Breed.* 39, 47. doi:10.1007/s11032-019-0954-y

Zhang, J., Zhang, X., Chen, R., Yang, L., Fan, K., Liu, Y., et al. (2020). Generation of Transgene-free Semidwarf Maize Plants by Gene Editing of Gibberellin-Oxidase20-3 Using Crispr/cas9. *Front. Plant Sci.* 11, 1048. doi:10.3389/fpls.2020.01048

Zhang, S., Zhang, R., Gao, J., Song, G., Li, J., Li, W., et al. (2021). CRISPR/Cas9-mediated Genome Editing for Wheat Grain Quality Improvement. *Plant Biotechnol. J.* 19, 1684–1686. doi:10.1111/pbi.13647

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and Transgene-free Genome Editing in Wheat through Transient Expression of Crispr/cas9 DNA or Rna. *Nat. Commun.* 7, 12617–12618. doi:10.1038/ncomms12617

Zhou, J., Deng, K., Cheng, Y., Zhong, Z., Tian, L., Tang, X., et al. (2017). Crispr-cas9 Based Genome Editing Reveals New Insights into Microrna Function and Regulation in Rice. *Front. Plant Sci.* 8, 1598. doi:10.3389/fpls.2017.01598

Zong, Y., Song, Q. N., Li, C., Jin, S., Zhang, D. B., Wang, Y. P., et al. (2018). Efficient C-To-T Base Editing in Plants Using a Fusion of Ncas9 and Human Apobec3a. *Nat. Biotechnol.* 36, 950–953. doi:10.1038/nbt.4261

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise Base Editing in Rice, Wheat and Maize with a Cas9-Cytidine Deaminase Fusion. *Nat. Biotechnol.* 35, 438–440. doi:10.1038/nbt.3811

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership