Research & Reviews: Journal of Microbiology and Biotechnology e-ISSN: 2320-3528

Plant Specific Homeobox *TaWUS-like* Gene Associated with TGW and GN in Common Wheat

Xuemei Si[1#], Shoaib Ur Rehman[1,2#], Yuquan Wang[1], Jian Hou[1], Chenyang Hao[1], Tian Li[1], Xueyong Zhang[1*], Hongxia Liu[1*]

[1]Key Laboratory of Crop Gene Resources and Germplasm Enhancement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

[2]Institute of Plant Breeding and Biotechnology, MNS University of Agriculture, Multan, Pakistan

[#]These authors have equal contributions

Research Article

Received: 05-Dec-2022, Manuscript No. JMB-22-82370; **Editor assigned:** 07-Dec-2022, PreQC No. JMB-22-82370(PQ); **Published:** 04-Jan-2023, DOI: 10.4172/2320-3528.12.1.001.

***For Correspondence:** Hongxia Liu and Xueyong Zhang, Key Laboratory of Crop Gene Resources and Germplasm Enhancement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

E-mail: liuhongxia02@caas.cn zhangxueyong@caas.cn

Keywords: Wheat; *WOX*; Haplotype; Allelic variation; MTA; Yield

ABSTRACT

WUSCHEL-related homeobox (*WOX*) gene is a plant-specific homeobox containing transcription factor (TF). *WUSCHEL* play critical roles in plant developmental events and crop yield. In present study, three copies of *TaWUS-like* were isolated from chromosomes 5A, 5B and 5D of common wheat. *TaWUS-likes* were investigated for sequence polymorphism to conduct marker trait association (MTA). Twenty single nucleotide polymorphisms (SNPs) variation and six inserts and deletions (InDels) were identified in the promoter (2.0 kb-3.0 Kb) and coding regions among the three copies of *TaWUS-like* homeologs. For high throughput genotyping, Kompetitive Allele-Specific PCR (KASP) markers were developed. Favored haplotypes were significantly associated with thousand grain weight (TGW) (*5B-HapI* and *5D-HapI*) and grain number (GN) (*5A-HapI* and *5B-HapI*) in wheat population collected from China. More sequence polymorphisms presented in *TaWUS-like-5B* genome than those of *TaWUS-like-5A* and *TaWUS-like-5D* genome. Absence of sequence polymorphisms in studied diploid accessions for *TaWUS-likes* and presence in studied tetra and hexaploid wheat accessions along with polymorphic information content and genetic diversity values suggest that polyploidization events resulted in increased genetic diversity for *TaWUS-like-5A* while decreased diversity for *TaWUS-like-5B* and *TaWUS-like-5D*. Geographic distribution and allelic frequency indicated that the favored haplotypes experienced positive but less intensive selection point towards underutilization of *TaWUS-likes* in wheat breeding. The developed KASP markers can be instrumental in grain yield improvement by marker assisted selection in wheat breeding programs.

INTRODUCTION

Breeding for higher grain yield has been a major objective in wheat breeding programs. A comprehensive understanding of genes conferring grain yield would benefit breeding for new cultivars. Wheat plays crucial role in ensuring global food security and doubling the wheat production by 2050 due to population surge is a big challenge [1,2]. Thousand grain weight (TGW) and grain number (GN) are considered as important yield contributing parameters. Variable correlations between TGW and GN have been reported, which mainly depends upon population [3,4]. In Chinese modern wheat cultivars, GN and TGW have positive correlation between them [5]. Several genes related to grain yield have been cloned in wheat and their functional markers (FMs) have been developed.

Marker assisted selection (MAS) has accelerated wheat breeding process [6]. As breeding science progressed, a better understanding of causal loci has increased breeders control over complex traits [7]. In comparison with conventional markers such as RFLP, AFLP, RAPD, ISSR and SSR, Single nucleotide polymorphism (SNP) based markers are more valuable due to its stability, cost effectiveness, high abundance and high-throughput scoring [8]. Moreover, SNP based markers are reported to be very effective in association analysis, genetic diversity, FMs development and MAS breeding [9-13]. Association analysis is an efficient tool used to identify marker trait association of target gene and has been widely reported in many plants including wheat and rice [11,14].

WUSCHEL-related homeobox gene are plant specific clade of homeobox-containing transcription factors (TFs). Increasing evidence reveal that *WUSCHEL* TF plays an important role in the whole process of plant growth and development from maintenance of stem cells at meristem to the formation of embryo [15,16]. The role of *WUSCHEL* genes in plant development has been studied in detail in *Arabidopsis thaliana*, corn and rice. Till now, only few reports are available on the regulation of yield-related traits by *WUSCHEL-related* TF family in gramineous crops. Zhang, et al. found that two functionally redundant TF genes, *WUSCHEL-related* homeobox 6 (*WOX6*) and *WOX11*, are expressed asymmetrically in response to auxin to connect gravitropism responses with the control of rice tiller angle. These approaches will identify genes to improve grain yields by facilitating the optimization of plant architecture [17]. Map-based cloning revealed that *DWT1* encodes a *WUSCHEL-related* homeobox TF homologous to the *Arabidopsis WOX8* and *WOX9* and highly expressed in young panicles. The *dwt1* mutant plants develop main shoots with normal height and larger panicles, but dwarf tillers bear smaller panicles as compared to those of the wild-type [18]. In case of wheat, gene sequence, structure and function of *TaWUS-like* genes are still limited.

In this study, we reported one *WUSCHEL-related* homeobox TF gene *TaWUS-like*, which were associated with TGW and GN in common wheat. The objectives of this study were to (i) identify polymorphism sites and develop breeder friendly high throughput KASP marker(s) for *TaWUS-like-5A*, *TaWUS-like-5B* and *TaWUS-like-5D*, (ii) identify favorable allelic variation(s), (iii) study the frequency and geographic distribution of allelic variation(s) among Chinese landraces and Chinese modern cultivars, European, Pakistani, CIMMYT, Russian and North American wheat cultivars. The purpose of this study is to provide a scalable, flexible, high throughput and gel free molecular marker to assist ongoing marker assisted breeding in wheat.

MATERIALS AND METHODS

Plant materials and growth conditions

Eight wheat germplasm populations were used as plant material to analyze the allelic variation in *TaWUS-like-5A*, *TaWUS-like-5B* and *TaWUS-like-5D* (Supplementary Table 1). A set of 348 Chinese modern wheat cultivars (Population-I) was phenotypes for grain-related traits including TGW, Grain Thickness (GT), Grain Width (GW) and Grain Length (GL) and spike-related traits including GN, Spikelet Number per spike (SN) and Spike Length (SL), Plant Height (PH), Heading Date (HD), Maturity Date (MD), Effective Tiller Number (ETN) were recorded in three environments in 2002, 2005 at Luoyang (112°45'E; 36°61'N) in Henan province and 2010 at Shunyi (116°56'E; 40°23'N) in Beijing. Randomized complete block design was followed in triplicates. Each accession was planted in four rows (plot size 2 meters). Forty seeds were planted in each row (row distance 25 centimetres). Ten plants from the middle of each plot were used to phenotype the aforementioned traits.

Another a set of 157 Chinese wheat landraces (Population-II) from Chinese Mini-core Collection (MCC), 336 European wheat cultivars (Population-III), 153 Pakistani wheat accessions (Population-IV), 53 CIMMYT wheat accessions (Population-V), 83 Russian wheat accessions (Population-VI) and 429 North American wheat accessions (Population-VII) were used to investigate the global geographic distribution of *TaWUS-like-5A*, *TaWUS-like-5B* and *TaWUS-like-5D* allelic variations and to evaluate the utilization value of the newly developed molecular markers in wheat breeding programs.

Chromosomal localization of *TaWUS-like*s gene

To identify the chromosomal location of target gene, the genomic sequences were used as queries to perform BLAST online (https://urgi.versailles.inra.fr/blast/blast.php). Chromosomal location was further confirmed by PCR amplification in diploid, tetraploid, hexaploid wheat species, nulli-tetrasomic, and ditelosomic lines of Chinese Spring (CS) wheat using *TaWUS-like-5A*-F/R, *TaWUS-like-5B*-F/R and *TaWUS-like-5D*-F/R primers (Supplementary Table 2).

A set of 16 diploid accessions for A subgenome, 12 diploid accessions for B subgenome, 11 diploid accessions for D subgenome and 21 tetraploid accessions for AABB genome were used to study genetic diversity for *TaWUS-like*s (Supplementary Table 3).

DNA isolation and sequencing of *TaWUS-like-5A/TaWUS-like-5B/TaWUS-like-5D*

DNA of 34 highly diverse polymorphic wheat accessions (Population-VIII) was extracted from young leaves with a DNA quick plant system (TIANGEN Biotech, Beijing, China) and dissolved in sterile water. A pair of primers for each *TaWUS-like* copy was selected to amplify the genomic sequences of *TaWUS-like-5A*, *TaWUS-like-5B* and *TaWUS-like-5D* (Supplementary Table 2).

KOD DNA polymerase was used for PCR amplification. PCR was performed in a total volume of 20 µL having 10 µL of KOD buffer, 1.2 µL of dNTPs (2.5 mM for each nucleotide), 0.2 µL of KOD FX enzyme, 0.4/0.4 µL forward and reverse primers (10 µM), 1 µL DNA (25 ng μl^{-1}) and 6.8 µL of ddH_2O. PCR conditions are given in Supplementary Table 2. PCR product was checked on agarose gel (1.2%) and desired bands were isolated and purified using AXYGEN DNA purifying kit, followed by cloning into pEASY®-Blunt cloning vector (TransGen Biotech, Beijing, China), then was transformed to 33 µL *trans1*-T1 phage resistant chemically competent cells (TransGen Biotech, Beijing, China) by heat shock. Ten positive clones for each sample were selected for sequencing by DNA analyzer 3730xl. To get full length desired sequences of *TaWUS-like-5A*, *TaWUS-like-5B* and *TaWUS-like-5D*, M13 forward/reverse primers and respective sequence walking primers were used (Supplementary Table 2).

The sequence of each clone was obtained by assembling with SeqMan program in DNASTAR Lasergene V.7.1.0 software package. The genomic origin of each sequence was firstly confirmed by comparing with reference genomic sequence obtained from URGI (https://urgi.versailles.inra.fr/blast/blast.php). The gene structure of *TaWUS-like-5A*, *TaWUS-like-5B* and *TaWUS-like-5D* was determined using MegAlign (DNASTAR Lasergene 7.1.0) through aligning coding and genomic sequences.

Phylogenetic tree analysis

In order to investigate the evolutionary relationship of *TaWUS-like* gene, a BLAST search was performed in the ENSEMBL database (http://asia.ensembl.org/index.html) based on *TaWUS-like* gene ID. The amino acid sequences of *TaWUS-like* proteins of the targeted species were downloaded, followed by the construction of phylogenetic tree of *TaWUS-like* proteins from a complete alignment of 36 *TaWUS-like* protein sequences. Neighbor-joining method with 1000 bootstrap replicates and p-distance substitution model using MEGA 7.0 (https://www.megasoftware.net/) was used to construct the phylogenetic tree.

Functional marker development

Population-VIII was used to detect polymorphisms in *TaWUS-like-5A*, *TaWUS-like-5B* and *TaWUS-like-5D*. For KASP marker development and high throughput genotyping, five gel free KASP markers were developed on selected SNP sites by following standard KASP guidelines (http://www.lgcgenomics.com). *KASP-5A-1126* at -1126 bp and *KASP-5A-764* at -764 bp were developed on the SNP sites in *TaWUS-like-5A* upstream region. *KASP-5B-883* at 883 bp and *KASP-5B-1672* at 1672 bp were developed on the selected SNP sites in *TaWUS-like-5B* coding region. *KASP-5D-226* at 226 bp was developed for solitary SNP in *TaWUS-like-5D* downstream region. Detailed information about KASP markers is given in Supplementary Table S4.

Moreover, the developed KASP markers were applied across the all seven wheat populations. In general, KASP mixture consisted of 30 µL common primers (100 µM), 12 µL of each tailed primer (100 µM) and 46 µL ddH_2O. KASP assays were tested in 5 µL PCR reaction mixture. Each reaction contains 2.4 µL DNA (25 ng/µL), 2.4 µL of 2X KASP master mixture, 0.04 µL $MgCl_2$/DMSO, 0.06 µL primer mixture and 0.1 µL ddH_2O. Fluorescence levels were detected by QuantStudio™ Flex (Applied Biosystems by Life Technologies) and data were visualized by using QuantStudio™ Real-time PCR software v.1.3 (Applied Biosystems by Life Technologies).

Statistical analysis

DNAMAN (9.0, Lynnon, Quebec, Canada) was used to compare amino acid sequences. Descriptive statistics, one-way ANOVA and Tukey test was performed for variance and significance analysis using SPSS (21.0, Chicago, USA). Association analysis was used to check marker trait association in the recorded data set. Associations were considered statically significant at $p<0.05$. The effects of allelic variations on each trait were analysed by Student's *t*-test at $p<0.05$ (even 0.01). Allelic frequencies were calculated in all seven wheat populations. Polymorphic Information Content (PIC) and gene diversity (H_e) were calculated using https://www.gene-calc.pl/pic website.

RESULTS

Chromosomal location of *TaWUS-likes*

To investigate the function of *TaWUS-like* in wheat, three copies of genomic sequences of *TaWUS-like* obtained from CS. All three homoeologous gene of *TaWUS-likes* were composed of two exons and one intron with the full length base pairs of 1903 bp, 1903 bp, 1912 bp respectively, and encodes putative protein products of 318, 321, and 322 amino acids. These proteins had a high identity of 94.75% and all contained a homeobox domain. The amino acid

sequence similarity of *TaWUS-like-5A* and *TaWUS-like-5B* is 89.10% and that between *TaWUS-like-5A* and *TaWUS-like-5D* is 92.90% and that between *TaWUS-like-5B* and *TaWUS-like-5D* is 91.36%. But the homology with rice, maize and *Arabidopsis thaliana* are low, about 29% (Figure 1).

A phylogenetic tree was constructed with the putative amino acid sequences of *TaWUS-like* members form wheat and other species. The results were in consistence with previous studies and as anticipated the three copies of *TaWUS-like* gene belong to the WUS branch of the WUSCHEL-related TF superfamily [19]. The deduced amino acid sequences showed a high homology with *WUSCHEL-related* TF family members of the monocotyledonous plants such as LOC0s01g62310, Zm0001d042920 and relatively lower homology with dicotyledonous species for example AT5G59340 (Figure 2A). Scanning site analysis indicated that TaWUS-like-5A/5B/5D have conserved homeobox domain (Figure 1).

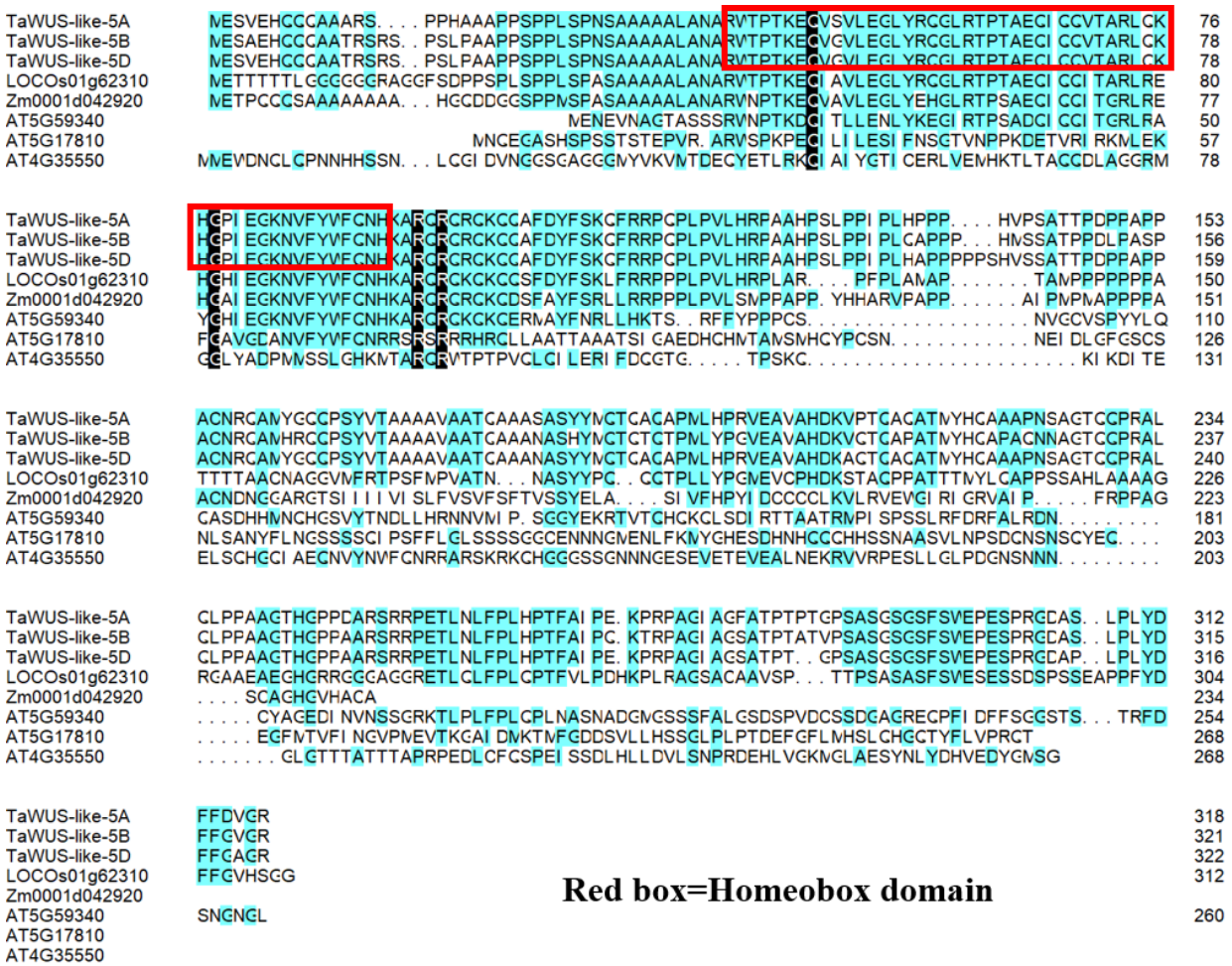

Figure 1. Protein amino acid sequence alignment of *TaWUS-likes* with the other homeobox domain containing TFs in rice (LOC0s01g62310), maize (Zm0001d042920) and *Arabidopsis* (AT5G59340, AT5G17810, AT4G35550). The homeobox domain is indicated by the red rectangle.

BLAST, nulli-tetrasomic and ditelomsoic lines showed that *TaWUS-likes* were located on short arm of chromosome *5A*, *5B* and *5D*, hence were named as *TaWUS-like-5A*, *TaWUS-like-5B* and *TaWUS-like-5D* accordingly (Figure 2B). The amplified fragment sizes of *TaWUS-like-5A/5B/5D* were 3507 bp, 2983 bp and 3076 bp, respectively.

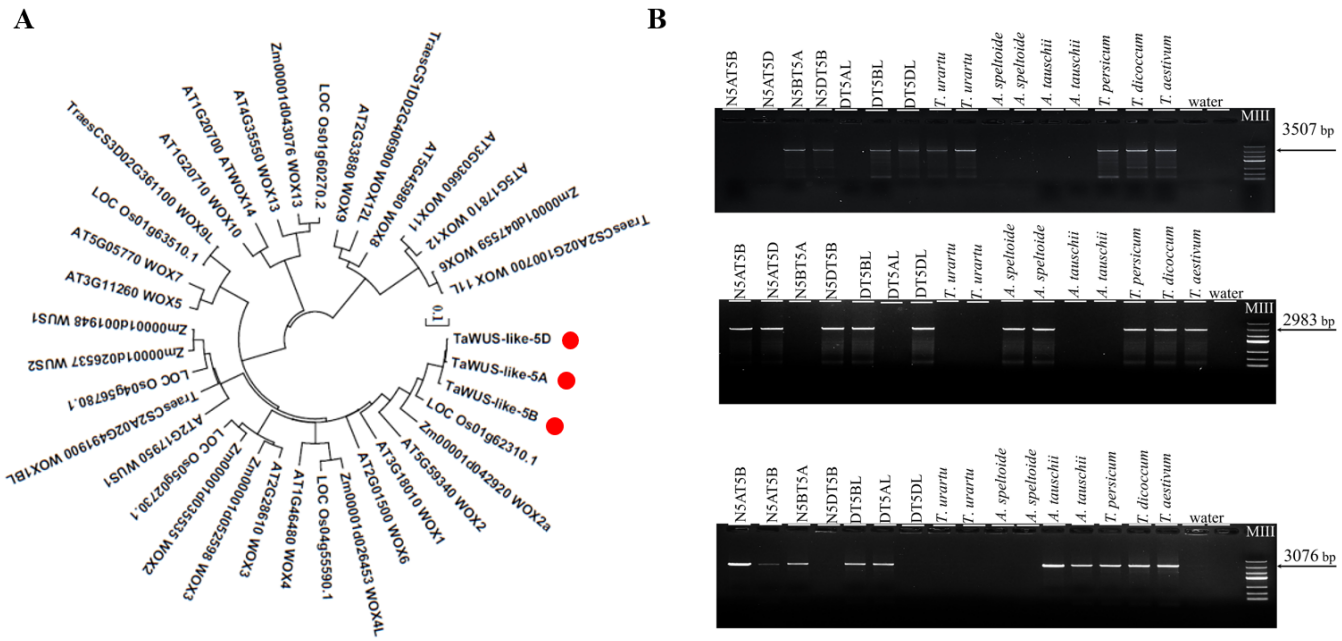

Figure 2. Phylogenetic tree and chromosomal location of *TaWUS-likes*. (A) Phylogenetic tree constructed from the *WUSCHEL* members protein sequences of 36 different species using the neighbor-joining method with a bootstrap value of 1000 by MEGA7.0. Three homoeologous proteins of *TaWUS-likes* are marked in red dots. (B) *TaWUS-likes* were located on chromosome *5A/5B/5D* using Nulli-Tetrasomic (NT) and Ditelosomic (DT) lines of Chinese Spring. *T. Urartu* (AA), *A. speltoide* (BB), *A. tauschii* (DD), *T. persicum* (AABB), *T. dicoccum* (AABB), *T. aestivum* (AABBDD), MIII–DNA marker III.

TaWUS-like-5A-HapI was significantly correlated with GN

Four allelic variations were identified in the upstream region of *TaWUS-like-5A*. Based on the two selected SNPs (-1126 bp and -764 bp) (Figure 3A), KASP markers named *KASP-5A-1126* and *KASP-5A-764* (Figure 3B) were developed to distinguish *TaWUS-like-5A-HapI* and *TaWUS-like-5A-HapII*.

Population-I was screened by the newly developed FMs. Association analysis revealed that, *5A-HapI* haplotype was significantly associated with higher GN, which is 3% more GN than *5A-HapII* ($p<0.05$ or $p<0.01$) (Figure 3C). Wheat accession containing *5A-HapI* also showed lower PH as compared to *5A-HapII*, which is 5.7% lower than *5A-HapII* ($p<0.01$) (Figure 3D). In addition, *5A-HapII* was the most frequent haplotype available in 80% accessions of Chinese modern cultivars.

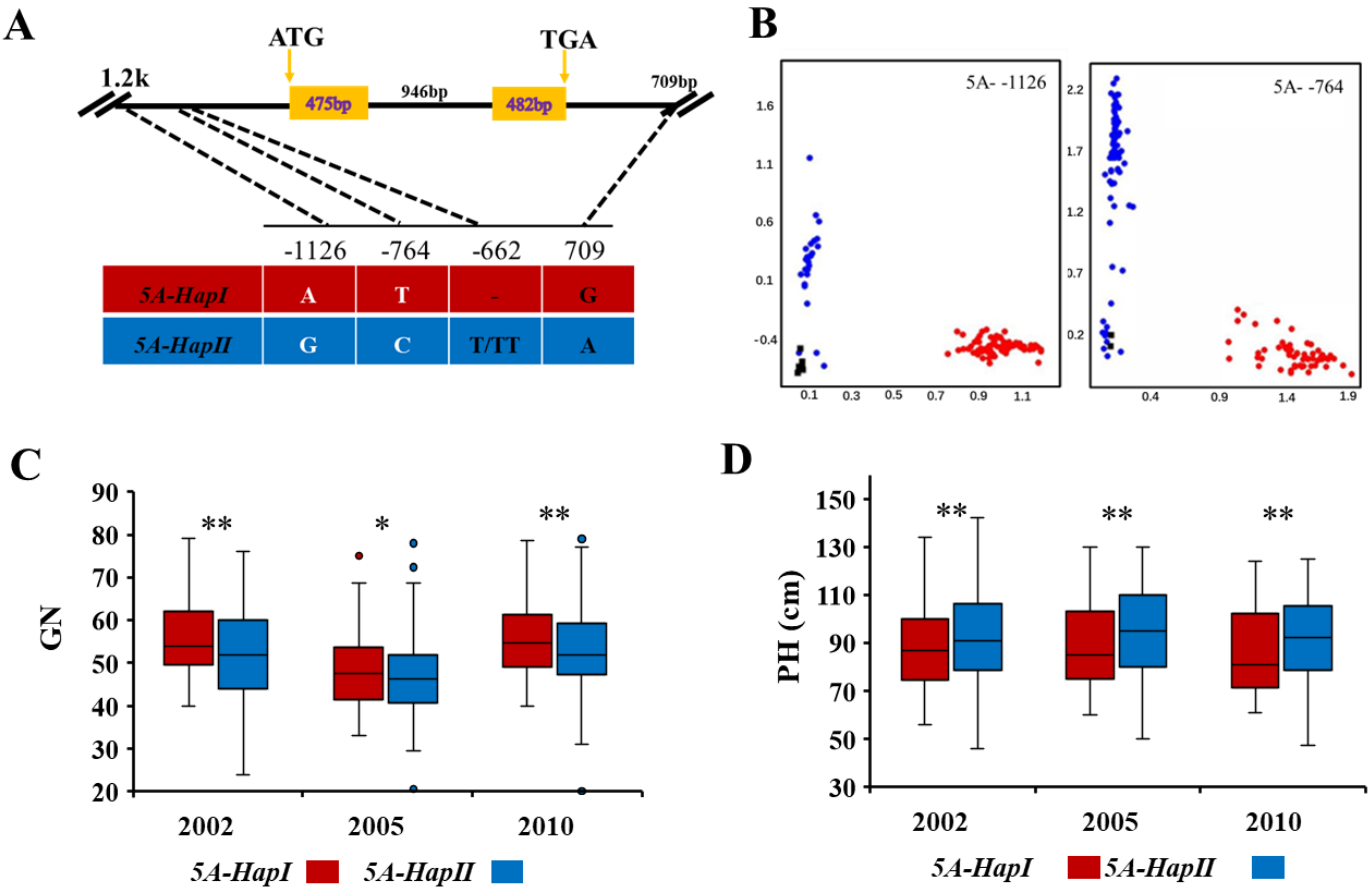

Figure 3. Gene structure and allelic discrimination plot and phenotype comparison for *TaWUS-like-5A*. (A) The top panel shows a schematic diagram of the gene and 1.2 kb promoter structure of *TaWUS-like-5A*. The ATG start codon was designated as position 1 bp. The bottom panel shows the polymorphic sites of *TaWUS-like-5A*. (B) Scatter plot for KASP assays showing clustering of accessions on X-(HEX) and Y-(FAM) axes. Small black squares indicate non-templet control. (C-D) Association analysis of *TaWUS-like-5A* haplotypes: Association of *TaWUS-like-5A* haplotypes with GN (C) and PH (D) in the Chinese modern cultivars in three different environments; the x-axis represents different environments. The asterisks indicate significant differences between haplotypes (Student's *t*-test, **p*<0.05, ***p*<0.01). **Note:** *5A-HapI*; *5A-HapII*.

TaWUS-like-5B-HapI associated with TGW and GN

Nineteen polymorphic sites were identified in the genomic region of *TaWUS-like-5B* (Figure 4A). Among the identified polymorphic sites, 14 are located in the promoter region, while four were identified in coding region and one in downstream region, forming two haplotypes. On the basis of the selected polymorphic site (at 883 bp and 1672 bp), KASP markers were developed to distinguish *TaWUS-like-5B-HapI* and *TaWUS-like-5B-HapII* (Figure 4B).

The newly developed KASP markers were then applied to Population-I. In Population-I, *5B-HapI* was most frequently occurring (75%) haplotype, followed by *5B-HapII* (25%). The association analysis showed that *5B-HapI* was significantly associated with higher TGW ($p<0.05$) (Figure 4C) and GN ($p<0.05$ or $p<0.01$) (Figure 4D). Wheat accession possessing *5B-HapI* had 3% more TGW than *5B-HapII*, respectively.

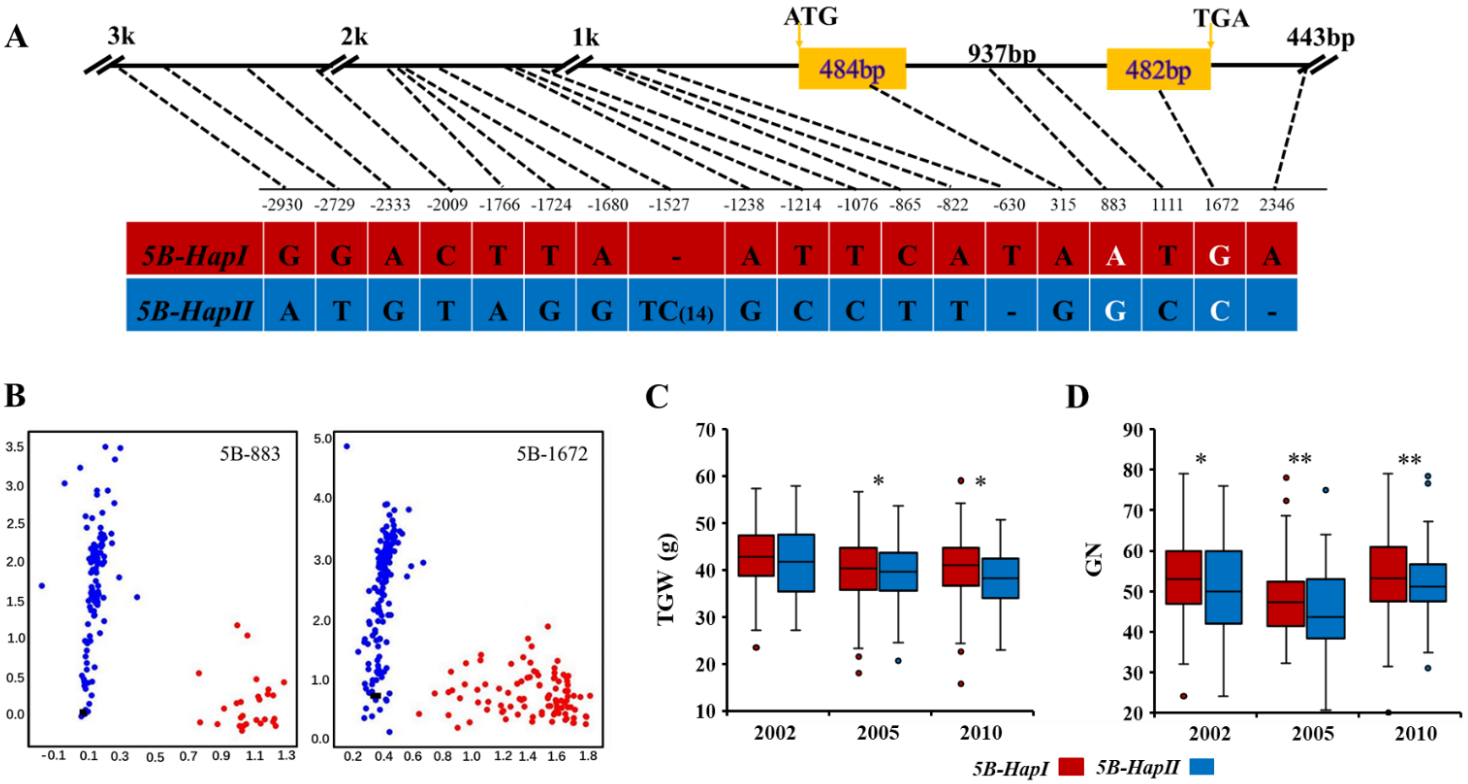

Figure 4. Gene structure and allelic discrimination plot and phenotype comparison for *TaWUS-like-5B*. (A) The top panel shows a schematic diagram of the gene and 3 kb promoter structure of *TaWUS-like-5B*. The ATG start codon was designated as position 1 bp. The bottom panel shows the polymorphic sites of *TaWUS-like-5B*. (B) Scatter plot for KASP assays showing clustering of accessions on X-(HEX) and Y-(FAM) axes. Small black squares indicate non-templet control. (C-D) Association analysis of *TaWUS-like-5B haplotypes*: Association of *TaWUS-like-5B* haplotypes with TGW (C) and GN (D) in the Chinese modern cultivars in three different environments; the x-axis represents different environments. The asterisks indicate significant differences between haplotypes (Student's *t*-test, **p*<0.05, ***p*<0.01). **Note:** *5B-HapI*; *5B-HapII*.

TaWUS-like-5D-HapI associated with TGW

Two SSR loci in the upstream and a solitary SNP (226 bp) (Figure 5A) was identified in the downstream region of TGA (stop codon) of *TaWUS-like-5D* forming two haplotypes. KASP marker (5D-226) was developed on the SNP site (Figure 5B).

Population-I was genotyped by using the newly developed KASP marker. Significant differences were recorded between the two haplotypes in three environments. Association analyses showed that genotypes possessing *5D-HapI* have higher TGW (3.4% higher) and lower PH (9.3% lower) as compared to genotypes possessing *5D-HapII* (Figures 5C-5D). In addition, *5D-HapI* haplotypes showed early maturity compared with *5D-HapII* (*p*<0.05) in two environments (Supplementary Figure 1).

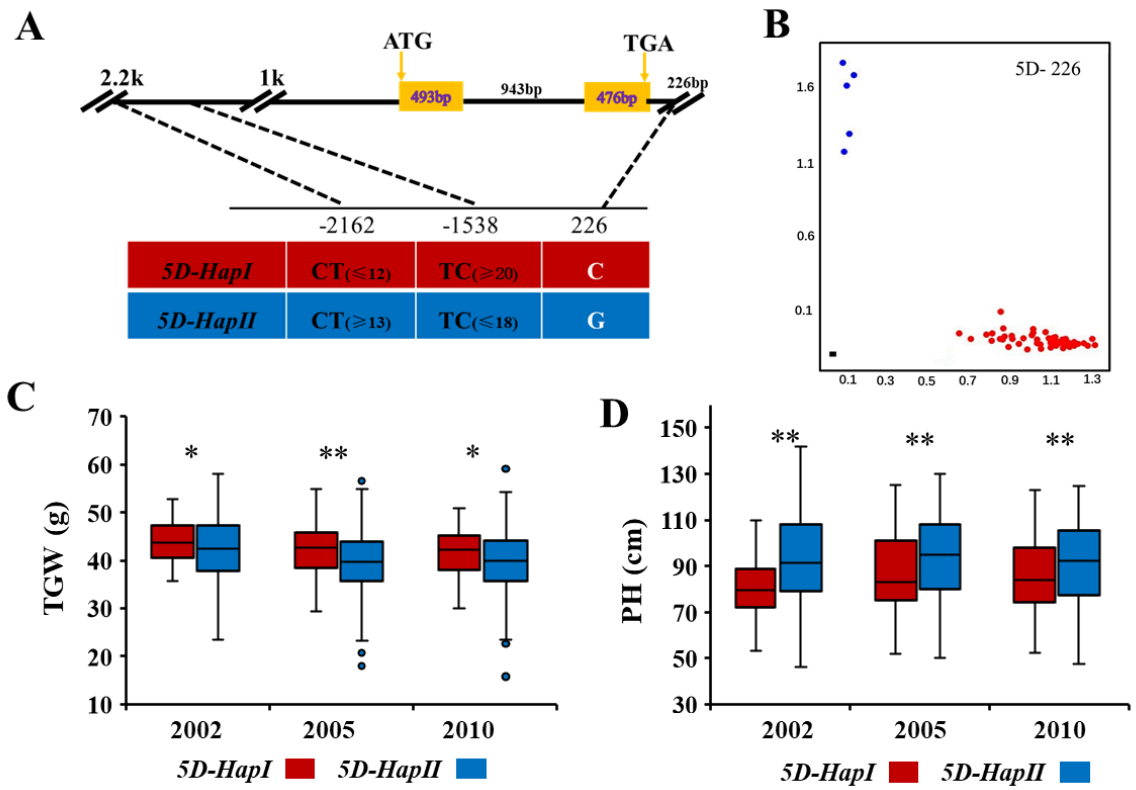

Figure 5. Gene structure and allelic discrimination plot and phenotype comparison for *TaWUS-like-5D*. (A) The top panel shows a schematic diagram of the gene and 2.2 kb promoter structure of *TaWUS-like-5D*. The ATG start codon was designated as position 1 bp. The bottom panel shows the polymorphic sites of *TaWUS-like-5D*. (B) Scatter plot for KASP assays showing clustering of accessions on X-(HEX) and Y-(FAM) axes. Small black squares indicate non-templet control. (C-D) Association analysis of *TaWUS-like-5D* haplotypes: Association of *TaWUS-like-5D* haplotypes with TGW (C) and PH (D) in the Chinese modern cultivars in three different environments; the x-axis represents different environments. The asterisks indicate significant differences between haplotypes (Student's *t*-test, **p*<0.05, ***p*<0.01). **Note**: *5D-HapI*; *5D-HapII*.

Geographic distribution of *TaWUS-like-5A/5B/5D* haplotypes

To determine, whether the favored allelic variations of *TaWUS-like-5A/5B/5D* were selected in wheat breeding, we investigated the geographic distribution of favorable allelic variations in China, Pakistan, Europe, Russia, North America and CIMMYT wheat germplasm.

On the basis of wheat, China is divided into ten agro-ecological zones [20,21]. In Chinese wheat landraces (Population-II), frequency of favored *TaWUS-like-5A-HapI* was very low (3.2%) and *TaWUS-like-5A-HapII* was the pre-dominant haplotype (96.8%) in all major wheat growing regions (Figure 6A). In Chinese modern wheat cultivars (Population-I), a significant increment (21%) in *TaWUS-like-5A-HapI* frequency took place (Figure 6B). Although positive selection of favored haplotype took place, the selection pressure was not as strong as we anticipated (Figures 6A- 6B). For *TaWUS-like-5B*, the frequencies of favored *HapI* were less (23.6%) in Population-II (Figure 6C). Remarkable increment (up to 75%) was observed in the Population-I (Figure 6D), suggesting strong breeding selection of these allelic variations. For *TaWUS-like-5D*, the proportions of favored *HapI* were still lower (11.5%) in Chinese landraces (Figure 6E), however, different frequency changes occurred in Population-I (modern cultivars). As Figure 6F indicated, the frequency of the favored *5D-HapI* showed an obvious increment in Chinese wheat zone-II, wheat zone-IV, wheat zone-

VI, wheat zone-X regions, but a dramatic decline in zone-IX, a slight decrease in zone-VII and zone-III, and the others almost kept the same frequency. In China, wheat zone-I, wheat zone-II, wheat zone-III, wheat zone-IV and wheat zone-VIII are the major wheat production area, holding about 85% proportion of wheat production, especially for the wheat zone-II, almost holds 45%-50% proportion of the wheat production. In addition, the main breeding target of the modern cultivars in wheat zone-IV focus on grain size and high TGW. Thus, in these TGW targeted major wheat regions, breeders still prefer the favored haplotypes of *TaWUS-like-5D-HapI*, suggesting *TaWUS-like-5D-HapI* positively contribute to yield and thus has underwent breeding selections in China. However, as a whole, the favored haplotypes still remain very low in Chinese modern cultivars (10.6%) (Figure 6).

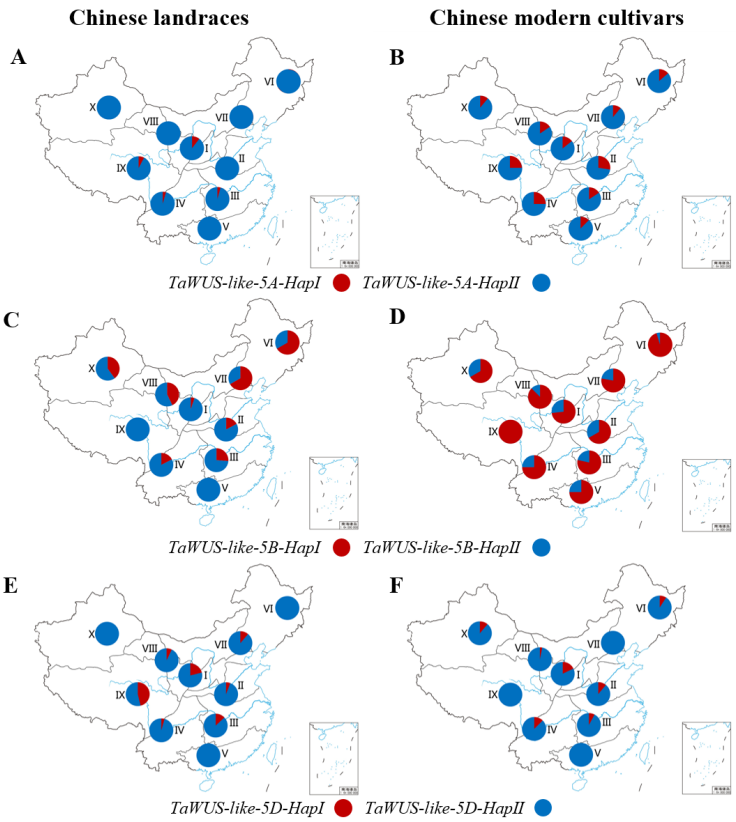

Figure 6A-6F. Geographic distribution of *TaWUS-like*s haplotypes in Chinese landraces and modern cultivars. I, The Northern Winter Wheat region; II, The Yellow and Huai River Valleys Wheat region; III, The Middle and Low Yangtze Valleys winter Wheat region; IV, The Southwestern winter Wheat region; V, The Southern winter Wheat region; VI, The Northeastern Spring Wheat region; VII, The Northern Spring Wheat region; VIII, The Northwestern Spring Wheat region; IX, The Qinghai-Tibetan Plateau Spring-Winter Wheat region; X, The Xinjiang Winter-Spring Wheat region. Size of pie diagram is directly proportional to the number of accessions. **Note**: *TaWUS-like-5A-HapI*, *TaWUS-like-5B-HapI*, *TaWUS-like-5D-HapI*; *TaWUS-like-5A-HapII*, *TaWUS-like-5B-HapII*, *TaWUS-like-5D-HapII*.

Among non-Chinese regions i.e. Pakistan, Europe, Russia, North America and CIMMYT wheat germplasm, favored haplotypes of *TaWUS-like-5A/TaWUS-like-5D* remain in lower frequencies except for *5B-HapI* which were apparently selected (Figures 7A-7C). However, the favored haplotypes of *TaWUS-like*s all showed a positive breeding pressure though the selection pressure on *TaWUS-like-5A/TaWUS-like-5D* haplotypes among Chinese and non-Chinese wheat

germplasm differed greatly in degree. Overall, lower favorable allelic frequencies indicate underutilization of these favored haplotypes both in Chinese and non-Chinese wheat breeding programs.

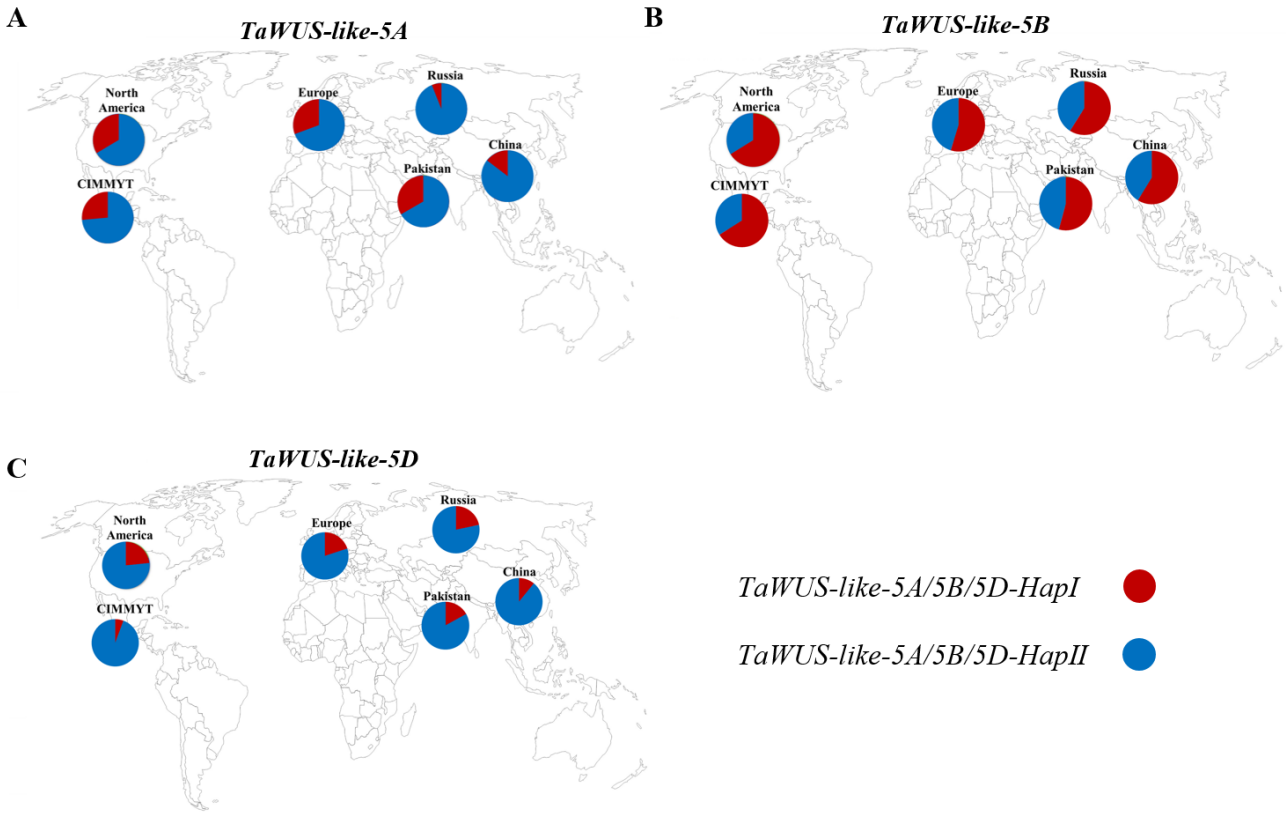

Figure 7. Geographic distribution of *TaWUS-like*s haplotypes in six different global regions. (A) *TaWUS-like-5A*. (B) *TaWUS-like-5B*. (C) *TaWUS-like-5D*. **Note**: *TaWUS-like-5A/5B/5D-HapI*; *TaWUS-like-5A/5B/5D-HapII*.

Domestication and breeding evolution of *TaWUS-like*s in global wheat germplasm

To survey the evolutionary history of *TaWUS-like*s, we analyzed the *TaWUS-like*s polymorphism degree in wheat progenitor accessions, hexaploid landraces (CLR) and hexaploid modern cultivars (CMC). Results showed that the PIC of *TaWUS-like-5B* (0.620) in CLR was much higher than those in *TaWUS-like-5D* (0.183) and *TaWUS-like-5A* (0.06) (Table1), indicating a higher SNP polymorphism or gene diversity in *TaWUS-like-5B* genome than the ones in *TaWUS-like-5D* and *TaWUS-like-5A* genome. Similar results obtained in H_e analysis. Interestingly, during polyploidization events, gene diversity increased in *TaWUS-like-5A* (PIC from 0.06 in CLR to 0.277 in CMC), while decreased in *TaWUS-like-5B* (PIC from 0.620 in CLR to 0.577 in CMC) and *TaWUS-like-5D* (PIC from 0.183 in CLR to 0.172 in CMC) (Table 1), showing a slight breeding pressure of *TaWUS-like-5D* and *TaWUS-like-5A* in Chinese breeding process. However, during domestication process, gene diversity of *TaWUS-like-5A* reduced dramatically (PIC from 0.269 in AABB to 0.06 in AABBDD), while the *TaWUS-like-5B* increased obviously (PIC from 0.164 in AABB to 0.620 in AABBDD), similar results also obtained in H_e analysis (Supplementary Table 5). These observations suggest that *TaWUS-like-5A/TaWUS-like-5B* were all subject to intense historic selection but at different evolution stages.

Table 1. Diversity at each locus in Chinese and non-Chinese wheat germplasm.

Polymorphic Information Content (PIC)							
Locus	**CLR**	**CMC**	**Europe**	**Pakistan**	**Russia**	**North America**	**CIMMYT**
TaWUS-like-5A	0.060*	0.277*	0.341	0.358	0.106	0.348	0.313*
TaWUS-like-5B	0.620**	0.577**	0.555	0.495	0.449	0.577	0.583**
TaWUS-like-5D	0.183**	0.172**	0.273	0.242	0.282	0.294	0.106**
Genetic diversity (H_e)							
TaWUS-like-5A	0.062*	0.332*	0.435	0.466	0.113	0.449	0.389*
TaWUS-like-5B	0.672**	0.626**	0.624	0.578	0.533	0.633	0.662**
TaWUS-like-5D	0.204**	0.190**	0.326	0.282	0.34	0.359	0.113**

Note: CLR=Chinese wheat Landraces; CMC=Chinese Modern wheat Cultivars; Higher PIC and H_e indicating higher polymorphism; worldwide wheat breeding has not reduced the overall genetic diversity of *TaWUS-like-5A* (indicated by *), while genetic diversity of *TaWUS-like-5B* and *TaWUS-like-5D* decreased in China and CIMMYT (indicated by **).

Genotype frequency of *TaWUS-likes* in Chinese wheat breeding

To evaluate the allelic variation frequencies of *TaWUS-likes*, we partitioned the studied wheat diversity panel into seven sub-groups by 10 years, namely 1930s, 1940s, 1950s, 1960s, 1970s, 1980s and 1990s. Results showed that the frequencies of *TaWUS-like-5A/TaWUS-like-5B/TaWUS-like-5D* favorable allelic variations were slow but increasing trend from 1930s to 1990s (Figures 8A-8C), however, the frequency proportion of favorable allelic variations was still very low for *TaWUS-like-5A/TaWUS-like-5D* in Chinese modern cultivars (Population-I) compared that with the unfavorable haplotypes. According to Population-I, the average TGW and GN exhibited continuous increasing trend from pre-1960s to 1961-2000s, while PH exhibited decreasing trend from pre-1960s to 1961-2000s (Figure 8D). These outcomes suggest that favorable allelic variations of *TaWUS-likes* are valuable and could be selected to improve wheat grain yield.

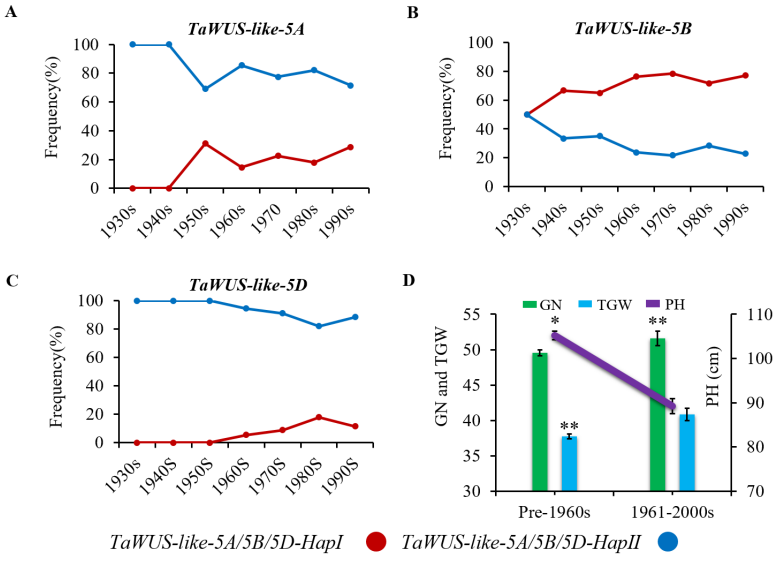

Figure 8. Changes of haplotypes frequencies and traits of *TaWUS-likes* in China. (A-C) *TaWUS-like* haplotypes frequencies in China from 1930s to 1990s. (D) Trend changes in TGW, GN and Plant Height in China. **Note:** *TaWUS-like-5A/5B/5D-HapI*; *TaWUS-like-5A/5B/5D-HapII*; GN; TGW; PH.

DISCUSSION

Resequencing small-scale core wheat accessions combined by large-scale KASP genotyping provides effective strategy for identifying superior alleles and marker-assisted breeding in wheat

Numerous studies have proved that MAS is vital for speeding up genetic gain to get high prediction rate of economically important traits and develop new germplasm in crops, therefore, marker based breeding is of utmost important for global food security in upcoming times [22]. *TaWUS-like-5D*, an important TGW-related gene cloned by our lab, regulate sucrose metabolism and utilization in grain endosperm therefore affecting wheat yield greatly (unpublished data). Moreover, *TaWOX-like-5D* plays a vital role in wheat architecture due to its residual meristem functions [23]. However, the genome allelic variations and evolutionary history of *TaWUS-like-5D* and its homoeologs of *TaWUS-like-5A* and *TaWUS-like-5B*, were still not characterized yet, thus limited the wide application of *TaWUS-like*s gene in yield improvement and wheat breeding. In this study, we used resequencing combined by KASP strategy to systematically characterize the allelic variations of *TaWUS-like* homoeologs.

Genotyping by sequencing (GBS) has emerged as an effective and accurate breeding strategy to identify genomic sequence polymorphism in different crop plants such as rice, maize, sorghum, and wheat [24-28]. However, in wheat, conducting GBS in a larger wheat population (>50 accessions) still is cost and time-consuming and technologically challenging due to the wheat higher ploidy level (3 sub-genome) and high proportion of repeat sequences (>85%) [11]. Thus, exploit or combination of the other high-throughput cost-effective molecular markers is also vital for identifying or MAS of superior alleles in wheat [29,30]. KASP is such a uniplex and flexible genotyping platform that utilizes a unique form of competitive allele-specific PCR combined with a novel, homogeneous, fluorescence-based reporting system for identification and measurement of genetic variation occurring at the nucleotide level to detect SNPs or Inserts and Deletions (InDels), moreover, KASP enables high throughput with a time- and cost-effective way [31,32]. KASP assay has been proved to be successively utilized in the genetic improvement of tomato and field crops [33-35]. Therefore, in this study, we first used a diverse set of 34 wheat accessions from China representing 70% diversity in Chinese wheat germplasm and accurately identified twenty SNPs and six InDels across the three *TaWUS-like* homoeologous genes by GBS, then grouped these variations into haplotypes and conduct the MTAs, finally five pairs of KASP markers were developed to distinguish *TaWUS-like* allelic variations and haplotypes in a larger wheat population. We found these KASP markers can easily genotype thousands of wheat germplasm (1347 wheat germplasm) within very short periods (96-well plates, 3-5 days). Moreover, the developed KASP markers can also be used to fingerprint that articulates *TaWUS-like* in any large breeding populations. Therefore, the application of KASP analysis in molecular breeding could significantly improve wheat breeding efficiency. We suggested that KASP markers can also be utilized in combination with other functional markers such as Cleaved Amplified Polymorphic Sequences (CAPS) and Simple Sequence Repeats (SSR) etc; to genotype yield-related traits. Thus, GBS (a small representative wheat set) combined by KASP genotyping assay could be a valuable method to be used in any large wheat breeding population.

Haplotype based MTAs are more informative than bi-allelic SNPs in wheat

In this study, we used haplotype not SNP to investigate MTAs. That because haplotype-based MTAs are more informative than bi-allelic SNPs [36]. Haplotype data can not only capture the associations that elude identification by solitary SNPs but also the epistatic interactions between SNPs. Hence, haplotype-based MTAs could increase prediction accuracies. We found the favored haplotypes of *TaWUS-like-5A-HapI/TaWUS-like5B-HapI/TaWUS-like-5D-HapI* were significantly associated with TGW (*5B-HapI and 5D-HapI*) and GN (*5A-HapI* and *5B-HapI*) in wheat population, accurately predict 61%-78% genotype among Chinese modern cultivars not considering the genetic

effects of the other yield related genes. Spring wheat collection from Pakistan (153 accessions) also showed positive selection trend of favored haplotypes of *TaWUS-likes*, suggesting haplotype based association are more effective than SNP thus these haplotypes can be instrumental to improve wheat grain yield in breeding.

More sequence polymorphisms presented in *TaWUS-like-5B* genome than those of *TaWUS-like-5A* and *TaWUS-like-5D* genome

Higher PIC and H_e observed were in the *TaWUS-like-5B* genome both in hexaploid CLR and in hexaploid CMC, suggesting the presence of more frequent sequence polymorphisms in *TaWUS-like-5B* gene than those of *TaWUS-like-5A* and *TaWUS-like-5D* genes. Similar results were obtained by Hao, *et al*. through resequencing 145 wheat accessions and identifying the polymorphisms in a sub-genome scale [37]. It is well known that wheat B sub-genome has relatively broader genetic base compared to A sub-genome and D sub-genomes, therefore more sequence polymorphisms in *TaWUS-like-5B* is probably due to the diverse *Aegilops speltoides* (the cross-pollinator ancestors for *Triticum aestivum* L.) natures. In addition, *Triticum aestivum* A and *Triticum aestivum* D sub-genome donors were self-pollinators that might be another reason for less sequence polymorphism for *TaWUS-like-5A/TaWUS-like-5D* compared to *TaWUS-like-5B* [38].

Favored haplotypes of *TaWUS-likes* experienced positive but less intensive selection point towards underutilization of *TaWUS-likes* in wheat breeding

Before 1960s, wheat disease resistance was the breeders main choice trait in China. At 1960s-1980s, the breeding objective was changed to breed for yield traits such as PH, TGW, GN and SN *etc*. Post 1990s, the objective turned to high yield and quality [37]. The changes of Chinese breeding result in the selection of favorable allelic variations in different genes. In this study, from Chinese wheat landraces to modern cultivars, the proportion of favored haplotypes of *TaWUS-like* homoeologs showed slight increment (Figures 8A-8C). A rapid increase of *TaWUS-like-5A-HapI* occurred from 1940s-1950s, whereas *TaWUS-like-5D-HapI/TaWUS-like-5B-HapI* showed a gradually increasing trend since 1950s/1930s, indicating favored haplotypes of *TaWUS-likes* experienced positive breeding selection but at different breeding stages. We assumed that, based on the haplotype trait association analysis, the selected traits of *TaWUS-like-5A*, *TaWUS-like-5B* and *TaWUS-like-5D* genes might be lopsidedly correlated with PH, GN and TGW respectively. This concept can be supported by the bias selection of *TaWUS-likes* at different evolution stage and by different selection degree according to the PIC and H_e analysis. In addition, we have transferred the *TaWUS-like-5D* into wheat background, and found that OE of *TaWUS-like-5D* resulted in less TGW and more serious disease symptom than the wild-type wheat receptor cultivar KN199 under field conditions (unpublished data), thus we do not exclude the possibility of *TaWUS-like-5A* was also contributed to adaptation traits such as disease resistance or flowering (meristem function like *TaWUS-like-5D*) since their amino acid sequence was very conserved therefore be selected strongly at wheat domestication periods.

Interestingly, during polyploidization events (CLR to CMC), gene diversity increased dramatically in *TaWUS-like-5A*, while *TaWUS-like-5B* and *TaWUS-like-5D* decreased slightly, showing a less breeding pressure for *TaWUS-like-5B* and *TaWUS-like-5D* in Chinese breeding process except *TaWUS-like-5A*. Higher H_e for *TaWUS-likes-5A* in Chinese and global modern wheat cultivars may be illustrated that conventional breeding by artificial hybridization has increased H_e of *TaWUS-like-5A*. Since the adaptation trait of *TaWUS-like-5A* preferred in worldwide breeding, therefore the usage of wide geographic sources of germplasm by present-day breeders has not reduced the overall genetic diversity of *TaWUS-like-5A*. Zhao, et al. have reported that the diversity of functional gene increased gradually during successive decades of wheat breeding [13]. This concept, supported by our different association roles of *TaWUS-like* homoeologs

in Chinese modern cultivars, highlights the possibility to avert the narrowing of genetic diversity by introgression of diverse germplasm in wheat breeding program.

Accessions from Pakistan, Europe, North America, Russia and CIMMYT were also used to evaluate *TaWUS-likes* haplotypes in five different geographical regions. In these regions, favored haplotypes experienced positive selection but the selection intensity was not as strong as we anticipated except for *5B-HapI* (Figures 7A-7C), indicating *5B-HapI* has undergone more strong selections than *5A-HapI* and *5D-HapI* both in China and in global. Given that the favored haplotypes of *TaWUS-likes*, experienced less positive selection intensity in Chinese and non-Chinese studied wheat breeding history clearly indicates the underutilization in breeding. Although, the studied wheat germplasm has different population structures, still the unconscious selection of the favored haplotypes is likely due to the high linkage disequilibrium for major yield related genes selected during selection breeding. For example, we observed that both haplotypes of all *TaWUS-like* genes contributed equally for TGW trait in Pakistani wheat program. However, in China, the increase of favored haplotypes in *5A* and *5B* was more apparent than the one in *5D* in major Chinese wheat production areas (Zone-I, Zone-II and Zone-III). Therefore, these observations suggest that *TaWUS-like-5A/TaWUS-like-5B/TaWUS-like-5D* were all subject to positive breeding selection but at different grain yield traits and by different degree among Chinese and global germplasm.

CONCLUSION

We identified allelic variations in *TaWUS-like* homoeologs among seven geographically different wheat germplasm. The favored haplotype of *TaWUS-like-5B* has been fully selected and utilized in global, whereas *TaWUS-like-5A* and *TaWUS-like-5D* could be further utilized in yield improvement breeding both in China and global. Association analysis showed that the favored haplotypes of *TaWUS-like* homoeologs were significantly associated with grain yield related traits. This evidence indicated that *TaWUS-likes* maybe involved in grain or spike development. These functional molecular markers can be used in future studies to assess their usefulness as selection criteria for improving grain yield related traits.

AUTHORS CONTRIBUTIONS

H.L. conceived the project, performed paper drafting (introduction and discussion section) and revised the whole manuscript; X.S., S.R. and Y.W. performed experiments, data interpretation and paper drafting (results section); J.H., T.L., C.H. and X.Z. provided valuable suggestions and comments. All authors have read and agreed to the published version of the manuscript.

FUNDING

The research was supported by the National Natural Science Foundation of China (No. 31471492 and 91935304), the National Key Research and Development Program of China (2016YFD0100402), the Innovation Project of the Institute of Chinese Academy of Agricultural Sciences, and National Crop Genomics and Speed Breeding Centre for Agriculture Sustainability-ADP-2021-22 of Pakistan (Lo21002838).

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

Supporting information is available from the Wiley Online Library or from the author.

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

We declare that these experiments comply with the current laws and ethical standards of China and Pakistan.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

REFERENCES

1. Ju L, et al. JAZ proteins modulate seed germination through interaction with ABI5 in bread wheat and *Arabidopsis*. New Phytol. 2019;223:246-260.
2. Zhang L, et al. Identification of a novel ERF gene, *TaERF8*, associated with plant height and yield in wheat. BMC Plant Biol. 2020;20:1-12.
3. Liu H, et al. Identification and validation of quantitative trait loci for kernel traits in common wheat (*Triticum aestivum* L.). BMC Plant Biol. 2020;20:529.
4. Jamil M, et al. Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. BMC Plant Biol. 2019;19:1-18.
5. Zhang D, et al. Identifying loci influencing grain number by microsatellite screening in bread wheat (*Triticum aestivum* L.). Planta. 2012;236:1507-1517.
6. Rasheed A, et al. From markers to genome-based breeding in wheat. Theor Appl Genet. 2019;132:767-784.
7. Ramstein GP, et al. Breaking the curse of dimensionality to identify causal variants in breeding 4. Theor Appl Genet. 2019;132:559-567.
8. Tong J, et al. High resolution genome wide association studies reveal rich genetic architectures of grain zinc and iron in common wheat (*Triticum aestivum* L.). Front Plant Sci. 2022;13:840614.
9. Rehman SU, et al. Development and exploitation of KASP assays for genes underpinning drought tolerance among wheat cultivars from Pakistan. Front Genet. 2021;12:684702.
10. Zhuang M, et al. The wheat short root length 1 gene *TaSRL1* controls root length in an auxin-dependent pathway. J Exp Bot. 2021;72:6977-6989.
11. Li A, et al. Wheat breeding history reveals synergistic selection of pleiotropic genomic sites for plant architecture and grain yield. Mol Plant. 2022;15:504-519.
12. Yang J, et al. Cloning, characterization of *TaGS3* and identification of allelic variation associated with kernel traits in wheat (*Triticum aestivum* L.). BMC Genet. 2019;20:98.
13. Zhao J, et al. Global status of 47 major wheat loci controlling yield, quality, adaptation and stress resistance selected over the last century. BMC Plant Biol. 2019;19:5.

14. Liu J, et al. The conserved and unique genetic architecture of kernel size and weight in maize and rice. Plant Physiol. 2017;175:774-785.
15. Leibfried A, et al. WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators. Nature. 2005;438:1172-1175.
16. Jha P, et al. WUSCHEL: a master regulator in plant growth signaling. Plant Cell Rep. 2020;39:431-444.
17. Zhang N, et al. A core regulatory pathway controlling rice tiller angle mediated by the *lazy1*-dependent asymmetric distribution of auxin. Plant Cell. 2018;30:1461-1475.
18. Wang W, et al. *Dwarf tiller1*, a WUSCHEL-related homeobox transcription factor, is required for tiller growth in rice. PLoS Genet. 2014;10:e1004154.
19. Li M, et al. Genome-wide identification and analysis of the *WUSCHEL-related* homeobox (*WOX*) gene family in allotetraploid *Brassica napus* reveals changes in *WOX* genes during polyploidization. BMC Geno. 2019;20:317.
20. Hou J, et al. Global selection on sucrose synthase haplotypes during a century of wheat breeding. Plant Physiol. 2014;164:1918-1929.
21. He ZH, et al. A history of wheat breeding in China. J Comp Neurol. 2001;523:805-813.
22. Zhang X, et al. *TaCol-B5* modifies spike architecture and enhances grain yield in wheat. Science. 2022;376:180-183.
23. Si X, et al. A sheathed spike gene, *TaWUS-like* inhibits stem elongation in common wheat by regulating hormone levels. Int J Mol Sci. 2021;22:11210.
24. Spindel J, et al. Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet. 2015;11:1004982.
25. Wang N, et al. Applications of Genotyping-By-Sequencing (GBS) in maize genetics and breeding. Sci Rep. 2020;10:16308.
26. Morris GP, et al. Population genomic and genome-wide association studies of agroclimatic traits in s*orghum*. Proc Natl Acad Sci. 2013;110:453-458.
27. Alipour H, et al. Genotyping-By-Sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. Front Plant Sci. 2017;8:1293.
28. Singh N, et al. In-silico detection of aneuploidy and chromosomal deletions in wheat using genotyping-by-sequencing. Plant Meth. 2020;16:1-6.
29. Jang YR, et al. High-throughput analysis of high-molecular weight glutenin subunits in 665 wheat genotypes using an optimized MALDI-TOF-MS method. 3 Biotech. 2021;11:1-8.
30. Lee SB, et al. A rapid, reliable RP-UPLC method for large-scale analysis of wheat HMW-GS alleles. Molecul. 2021;26:6174.
31. Jiang P, et al. Linkage and association mapping and kompetitive allele-specific PCR marker development for improving grain protein content in wheat. Theor Appl Genet. 2021;134:3563-3575.
32. Makhoul M, et al. Overcoming polyploidy pitfalls: a user guide for effective SNP conversion into KASP markers in wheat. Theor Appl Genet. 2020;133:2413-2430.
33. Devran Z, et al. Development of molecular markers for the *Mi-1* gene in tomato using the KASP genotyping assay. Hortic Environ Biotechnol. 2016;57:156-160.

34. Semagn K, et al. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed. 2014;33:1-14.
35. Rasheed A, et al. Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. Theor Appl Genet. 2016;129:1843-1860.
36. Stephens JC, et al. Haplotype variation and linkage disequilibrium in 313 human genes. Science. 2001;293:489-493.
37. Hao C, et al. Resequencing of 145 landmark cultivars reveals asymmetric sub-genome selection and strong founder genotype effects on wheat breeding in China. Mol Plan. 2020;13:1733-1751.
38. Marcussen T, et al. Ancient hybridizations among the ancestral genomes of bread wheat. Science. 2014;345:1250092.

Citation: Liu H and Zhang X, et al. Plant Specific Homeobox *TaWUS-like* Gene Associated with Thousand Grain Weight and Grain Number in Common Wheat. RRJ Microbiol Biotechnol. 2023;12:001.