

RESEARCH

Open Access



# A high-resolution genotype–phenotype map identifies the *TaSPL17* controlling grain number and size in wheat

Yangyang Liu<sup>1,2†</sup>, Jun Chen<sup>3†</sup>, Changbin Yin<sup>4†</sup>, Ziyang Wang<sup>1,2</sup>, He Wu<sup>1,2</sup>, Kuocheng Shen<sup>1,2</sup>, Zhiliang Zhang<sup>2,4</sup>, Lipeng Kang<sup>2,4</sup>, Song Xu<sup>2,4</sup>, Aoyue Bi<sup>2,4</sup>, Xuebo Zhao<sup>2,4</sup>, Daxing Xu<sup>2,4</sup>, Zhonghu He<sup>3,5</sup>, Xueyong Zhang<sup>3</sup>, Chenyang Hao<sup>3</sup>, Jianhui Wu<sup>6</sup>, Yan Gong<sup>3</sup>, Xuchang Yu<sup>1,2</sup>, Zhiwen Sun<sup>1,2</sup>, Botao Ye<sup>1,2</sup>, Danni Liu<sup>1,2</sup>, Lili Zhang<sup>1</sup>, Liping Shen<sup>1</sup>, Yuanfeng Hao<sup>3\*</sup>, Youzhi Ma<sup>3\*</sup>, Fei Lu<sup>2,4,7\*</sup> and Zifeng Guo<sup>1,2\*</sup>

<sup>†</sup>Yangyang Liu, Jun Chen, and Changbin Yin contributed equally to this work.

\*Correspondence: haoyuanfeng@caas.cn; mayouzhi@caas.cn; flu@genetics.ac.cn; guozifeng@ibcas.ac.cn

<sup>1</sup> Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

<sup>3</sup> Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100081, China

Full list of author information is available at the end of the article

## Abstract

**Background:** Large-scale genotype–phenotype association studies of crop germplasm are important for identifying alleles associated with favorable traits. The limited number of single-nucleotide polymorphisms (SNPs) in most wheat genome-wide association studies (GWASs) restricts their power to detect marker-trait associations. Additionally, only a few genes regulating grain number per spikelet have been reported due to sensitivity of this trait to variable environments.

**Results:** We perform a large-scale GWAS using approximately 40 million filtered SNPs for 27 spike morphology traits. We detect 132,086 significant marker-trait associations and the associated SNP markers are located within 590 associated peaks. We detect additional and stronger peaks by dividing spike morphology into sub-traits relative to GWAS results of spike morphology traits. We propose that the genetic dissection of spike morphology is a powerful strategy to detect signals for grain yield traits in wheat. The GWAS results reveal that *TaSPL17* positively controls grain size and number by regulating spikelet and floret meristem development, which in turn leads to enhanced grain yield per plant. The haplotypes at *TaSPL17* indicate geographical differentiation, domestication effects, and breeding selection.

**Conclusion:** Our study provides valuable resources for genetic improvement of spike morphology and a fast-forward genetic solution for candidate gene detection and cloning in wheat.

**Keyword:** Grain number, Grain size, GWAS, *TaSPL17*, Wheat



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Global cereal production should increase by around 50% during the first half of the twenty-first century to satisfy the expected rising demand driven by population growth [1]. Wheat (*Triticum aestivum* L.) is one of the most important crops worldwide. It supports more than 2.5 billion people [2] and accounts for around 20% of dietary calories and protein consumed by humankind [3], around 30% of global grain production, and around 45% of cereals used as food [4]. Since crop diseases, climate change, and abiotic stresses (e.g., drought and heat) cause significant grain yield loss, raising grain yield remains the primary goal in most wheat-producing areas [5–7].

Wheat spikes are the organs that determine grain yield. Wheat spikes consist of spikelets that produce reproductive structures called florets. An individual floret consists of lemma, palea, anthers, carpels, awns, and lodicules. Increased partitioning of assimilates to grains has a great influence on grain yield in wheat. Indeed, assimilate distribution between grains and spike chaff plays a critical role in determining grain weight in wheat [8–10]. However, the genetic connections between grains and other spike components (spike chaff) are largely unknown. A greater and global understanding of the genetic relationships between grains and spike chaff is critical for achieving desired gains in specific traits by harnessing different allelic combinations.

Genome-wide association study (GWAS) is a powerful strategy for dissecting the genetic basis of complex traits and identifying marker–trait associations [10–12]. Although numerous wheat GWAS have been conducted, the limited number of single-nucleotide polymorphisms (SNPs) (<1 million) used for these studies limit their ability to detect marker–trait associations. Due to the large genome size and high proportion of repetitive sequences in hexaploid wheat, progress in whole-genome sequencing is relatively slow, restricting wheat research. From 2012 to 2018, great strides have been achieved in genome sequencing of diploid, tetraploid, and hexaploid wheat [13–19]. The genome sequence of hexaploid wheat was released in 2018 (IWGSC, 2018), which greatly increased the number of SNPs available for investigation, revealing much about wheat evolution and domestication [20–25].

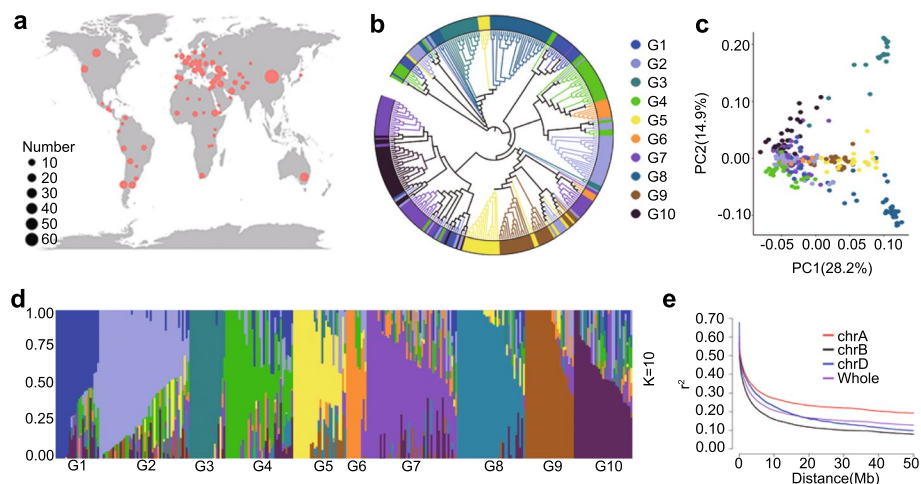
Although several studies identifying quantitative trait locus (QTLs) associated with spike morphology traits have been reported, few studies have focused on more than ten traits in a single study. Here, we conducted a large-scale GWAS based on approximately 40 million filtered SNPs for 27 spike morphology traits to dissect and understand the genetic basis of spike morphology. This approach may save a lot of time and work to get relatively precise locations of candidate genes leading to fast cloning of those genes for economically important traits. The genotype–phenotype map of these 27 traits revealed novel genomic regions associated with each putative trait and candidate genes. The 27 spike morphology traits enabled us to identify 392 additional genomic regions, indicating the value of dissecting spike morphology genetics for the detection of novel signals. Through GWAS, we identified and validated the gene *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE17* (*TaSPL17*), whose product positively regulates grain yield through grain number and size. Our results provide insights into the genetic basis of wheat spike morphology traits and serve as a resource for the genetic improvement of wheat grain yield.

## Results

### Genetic variation and population structure

In this study, we used genotype data from wheat accessions to create a whole-genome genetic variation map of wheat [20, 26]. The latest version of VMap (VMap 2.0) is based on 1062 wheat accessions with multiple ploidy levels [27]. From these, we selected 306 hexaploid wheat worldwide accessions [28]. High-coverage whole-genome sequencing ( $\sim 10\times$ ) enabled the high-confidence identification of 117,398,144 SNPs in the 306 wheat accessions; the false-positive error rate of variant calling (i.e., the proportion of segregating sites in the reference accession Chinese Spring) was only 0.00067%. We used 40,710,923 filtered SNPs (A genome, 17,509,041; B genome, 20,603,513; D genome, 2,598,369) with a minor allele frequency (MAF) of at least 0.05 for this study.

These 306 wheat accessions were mainly landraces collected from more than 70 countries (Fig. 1a, Additional file 1: Table S1). We constructed a phylogenetic tree (Fig. 1b), performed a principal component analysis (PCA) (Fig. 1c) and a population structure analysis (Fig. 1d) on these 306 accessions. The PCA results demonstrated that geography at the continental scale is the most important correlate of genetic structure. We classified the accessions into 10 genetic groups, G1–G10, comprising 23, 48, 19, 36, 28, 11, 48, 36, 26, and 31 samples, respectively (Additional file 1: Table S2). The first and second principal components (PC1 and PC2) of the PCA explained 28.2% and 14.9% of the standing variance, respectively, and reflected geographical factors, with PC1 separating European (G2) and Asian (G8) wheat accessions, and PC2 distinguishing between Middle Eastern (G5) and African (G9) wheat accessions (Fig. 1c, Additional file 1: Tables S1 and S2). These groups were consistent with the related phylogenetic tree. As  $K$  increased in the population structure analysis, the corresponding cross-validation error decreased and



**Fig. 1** Diversity represented by 306 hexaploid wheat accessions. **a** Geographical distribution of wheat accessions, as indicated by circles on the world map. Circle size is proportional to the number of accessions. **b** Phylogenetic tree of all accessions calculated from whole-genome SNPs. The colors indicate the 10 groups (G1–G10). **c** Principal component analysis (PCA) plot of the first two components (PC1 and PC2) of the 306 wheat accessions. **d** Population structure analysis with  $K=10$ . The y-axis quantifies cluster membership, and the x-axis represents the different accessions. The order and position of the accessions along the x-axis are consistent with those in the phylogenetic tree. **e** Decay of linkage disequilibrium (LD) in the wheat genome. LD decay was estimated based on SNP markers from all three genomes: A genome (chr A), B genome (chr B), and D genome (chr D)

reached a trough at  $K=10$  (Additional file 2: Figure S1). We also calculated the decay rate of linkage disequilibrium (LD) as the pairwise correlation coefficient ( $r^2$ ) from the maximum value to the half maximum, which reached 3.08 Mb, for all accessions with variation among different genomes (Fig. 1e, Additional file 1: Table S3). LD in the A, B, and D genomes was 6.04, 2.39, and 3.03 Mb, respectively.

### Overview of spike morphology traits

Previous work has focused on grain number and size to identify quantitative trait locus (QTLs) and other genomic regions related to grain yield [29]. Here, we looked for genomic signals in grain yield traits by dissecting the genetic connections between grains and other spike components (spike chaff), because assimilate partitioning is closely associated with spike morphology and grain yield traits in wheat. We thus dissected grain weight, number, and size using the traits associated with assimilate distribution between spike components. We analyzed a total of 27 spike morphology traits corrected on 306 worldwide wheat accessions. These 27 traits comprised 12 measured traits: spike length, spike dry weight, grain weight per spike, awn weight per spike, grain number per spike, spikelet number per spike, spike chaff weight per spike, rachis weight per spike, individual spikelet weight, grain number per spikelet, chaff weight per spikelet, and grain weight per spikelet (Additional file 1: Tables S4, S5 and S6). In addition, we used the 12 measured traits to determine 15 ratios/traits to estimate the distribution of assimilates among spike components (Additional file 1: Tables S4, S5 and S6). The 306 accessions displayed strong variation for all 27 traits (Additional file 1: Tables S4, S5 and S6, Additional file 2: Figures S2 and S3). We estimated the broad sense heritability ( $h^2$ ) for all 27 traits by calculating repeatability between raw phenotypes. Generally, direct traits exhibited high heritability, whereas the heritability of the indirect traits (the ratios between spike direct traits) was relatively low (Additional file 1: Table S4).

All traits followed a normal distribution with three exceptions: awn weight per spike, awn weight/grain weight, and awn weight/spike chaff weight (Additional file 2: Figure S3). The abnormal distribution of these traits may be attributed to the geographical differentiation of wheat awn. Interestingly, these three traits had relatively high heritability among the 27 traits (awn weight/spike weight,  $h^2=0.87$ ; awn weight/grain weight,  $h^2=0.73$ ; awn weight/spike chaff weight,  $h^2=0.72$ ) (Additional file 1: Table S4). The broad genotypic variation and relatively high heritability of most traits (Additional file 1: Tables S4, S5 and S6) suggest the large potential and a genetic basis for the increase in grain yield by manipulating spike morphology.

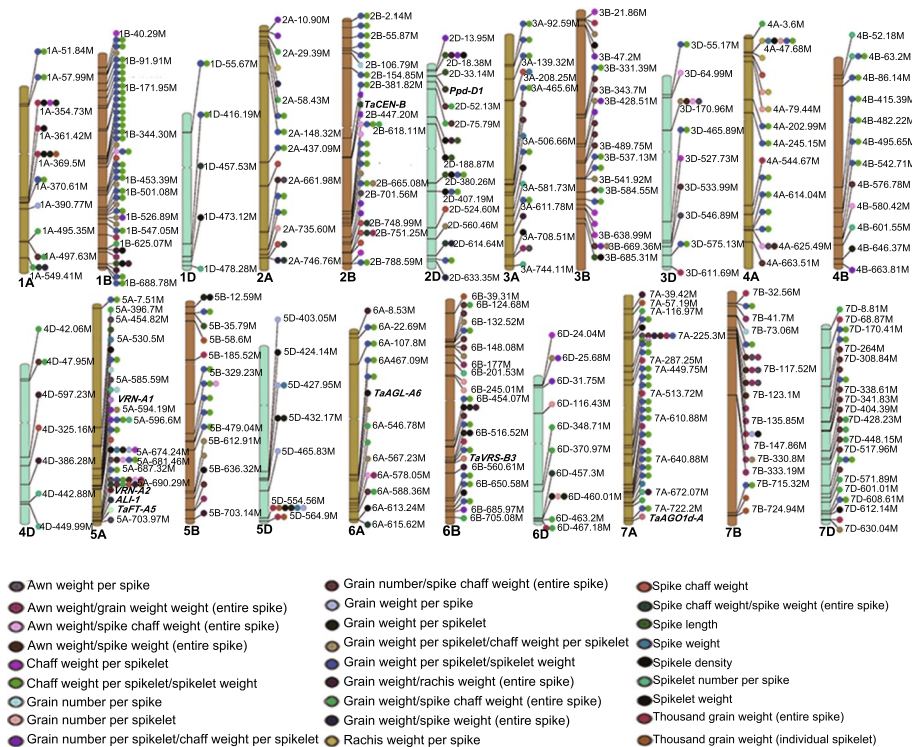
The strong genetic effects of spike length, spikelet number per spike, awn weight per spike, and spike density according to an analysis of variance (ANOVA) analysis were consistent with their relatively high heritability (Additional file 1: Tables S4, S5 and S6). The environmental effects and the influence of genotype  $\times$  environment were significant for grain number and thousand kernel weight (spikelet), which may explain their low heritability (Additional file 1: Tables S4, S5 and S6). The large residuals of spike chaff weight, chaff weight per spikelet, grain weight/chaff weight, grain weight per spikelet / chaff weight (spikelet), grain weight per spikelet / spikelet weight (spikelet), and chaff weight per spikelet / spikelet weight indicate that the dissection of these traits not only magnified the differences between genotypes, but also led to the strong variations

among replicates within individual genotype, which resulted in low heritability (Additional file 1: Tables S4, S5 and S6).

The high-resolution dissection of spike morphology above captured several important relationships (Additional file 2: Figure S4). For instance, grain weight per spikelet/spikelet weight and chaff weight/spikelet weight were strongly and negatively correlated. Consistently, grain weight per spike/spike weight exhibited close and negative connections with spike chaff weight/spike weight. Awn weight per spike was strongly and positively associated with spike chaff weight, suggesting that awn is an important part of spike chaff.

**GWAS for 27 spike morphology traits**

With the set of 40,710,923 SNPs defined above (MAF >0.05; missing rate <20%; missing genotype rate <0.1), we performed GWAS for 27 spike morphology traits. We identified 132,086 significant marker–trait associations ( $-\log_{10}[P \text{ value}] > 5.00$ ) across the three subgenomes, with 33,619 associations from the A genome, 90,158 from the B genome, and 8309 for the D genome (Additional file 1: Table S7). We also used LD and connections between markers to delineate 590 genomic regions associated with the 27 traits, comprised of 216 for the A genome, 254 for the B genome, and 120 for the D genome (Fig. 2, Additional file 1: Table S8). Of the 590 identified genomic regions, 422 were



**Fig. 2** Genotype–phenotype map for the 27 spike morphology traits. Genotype–phenotype map with the identified genomic regions for the corresponding traits in each chromosome were aligned to the reference sequence of bread wheat. Traits determined based on the entire spike are indicated where applicable. Individual spikelets were collected from the center of spike. The symbol “/” indicates the ratio between the two traits



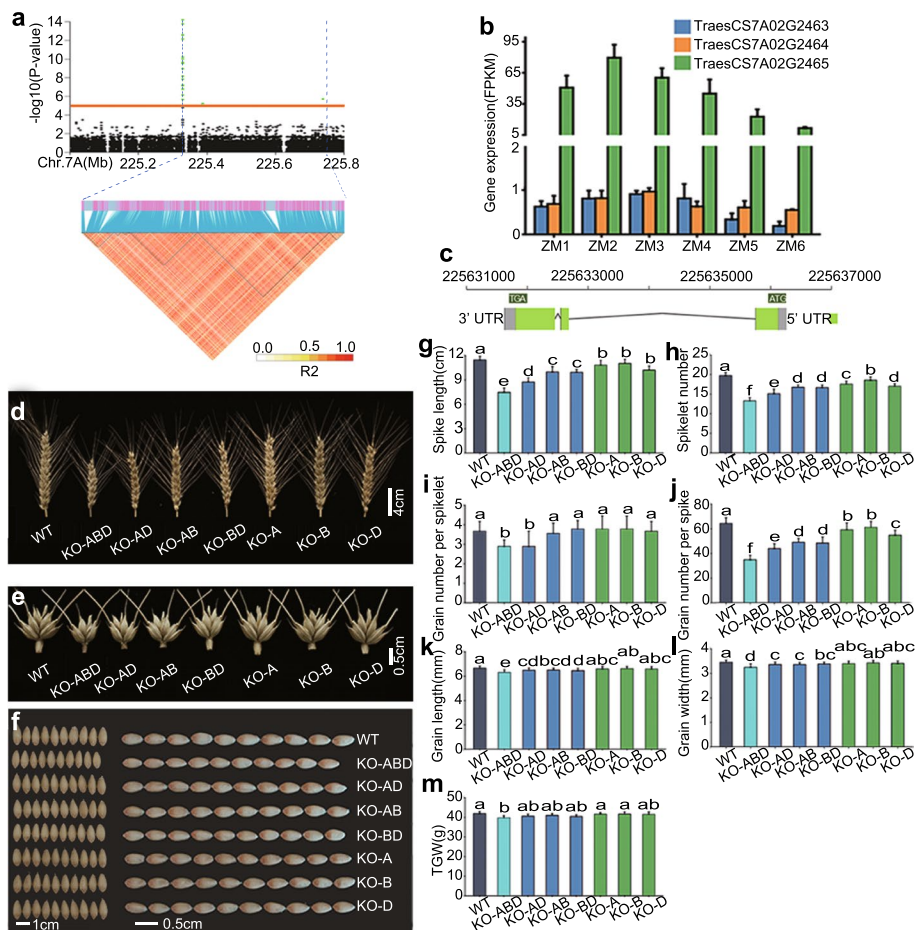
associated with more than one trait. Some associated genomic regions exhibited pleiotropic effects on the traits related to grain number/size and assimilate partitioning, while the genomic regions for grain number/size traits were not as strong as the regions for assimilate partitioning traits. For instance, we identified one region (1B: 628,535,896–629,920,890) when mapping for grain weight/rachis weight (entire spike) and spikelet density. The best SNP for the signal of grain weight/rachis weight (entire spike) (assimilate partitioning) were of much higher significance ( $-\log_{10} [P \text{ value}] = 9.37$ ) than the best SNP of spikelet density ( $-\log_{10} [P \text{ value}] = 6.07$ ). In addition, 392 peaks associated with assimilate distribution (ratios between spike component weight) were not observed for the remaining spike morphology/grain yield traits (Additional file 1: Tables S8 and S9). Haplotype analyses showed that 296 of 392 peaks have effects on the remaining spike morphology/grain yield traits (Additional file 1: Table S9). These results suggest that the ratios between the weights of spike components minimize the variance among the values of spike components, which in turn maximize the ability to detect differences for that spike component. This approach increases the power of GWAS to detect significant associations and highlights the necessity of conducting GWAS on assimilate partitioning traits.

We included awn weight as an internal control for GWAS. *Awn Length Inhibitor 1* (*ALI-1*, *TraesCS5A02G542800*) was previously confirmed as the causal gene for the *Tipped1* (*B1*) locus that controls awn initiation and development [30–32]. We detected 1325 marker-trait associations of awn weight, within the region (5A:690,292,073–703,437,181), which includes *ALI-1* (*TraesCS5A02G542800*) (Additional file 1: Table S8, Additional file 2: Figure S5). In addition, we identified *VERNALIZATION-A1* (*VRN-A1*), *VRN-A2*, *Photoperiod-D1* (*Ppd-D1*), *FLOWERING LOCUS T-A5* (*TaFT-A5*), *AGAMOUS-LIKE 6* (*TaAGL-A6*), *ARGONAUTE 1d* (*TaAGO1d-A*), *SIX-ROWED SPIKE-B3* (*TaVRS-B3*), and *CENTRORADIALIS-B* (*TaCEN-B*) as candidate genes for spike morphology traits (Fig. 2, Additional file 1: Table S8).

### GWAS identification of a candidate gene regulating grain number and size

To further validate the power of the assimilate partitioning traits for detecting significant associations in GWAS, we present GWAS data for eight traits related to assimilate partitioning between spike components. These eight traits magnify the existing variation among different wheat accessions, allowing us to observe strong signals associated with grain weight and number. We detected different peaks for the eight traits on different chromosomes (Additional file 2: Figure S5), including peaks for awn weight. *OsSPL14* (*SQUAMOSA PROMOTER BINDING PROTEIN-LIKE14*, also named as *IDEAL PLANT ARCHITECTURE1* [*IPA1*]) in rice (*Oryza sativa* L.) is involved in the regulation of spike architecture, spikelet number, and tillering [33–36]. *OsSPL14* is a new Green Revolution gene that plays an important role in regulating plant architecture in rice [37–39]. *HvSPL14* in barley (*Hordeum vulgare* L.) is associated with very similar spike phenotypes for its KO lines [40]. *TaSPL17* is orthologous to *OsSPL14* [41, 42]; however, its effects on wheat plant architecture and grain yield remain largely unknown. Using the threshold =  $-\log_{10} [P \text{ value}] > 5.00$  according to previous reports [12, 23, 43–45], one identified genomic region (225,328,984 bp to 225,737,586 bp) contained *TaSPL17-A* (*TraesCS7A02G246500*) and another two genes (*TraesCS7A02G246400*, *TraesCS7A02G246300*)

(Fig. 3a, Additional file 2: Figure S5). We extracted the relative transcript levels of the three genes at six developmental stages according to inflorescence morphological features from a published transcriptome dataset of wheat spike [46, 47]. Using the threshold  $= -\log_{10}[P \text{ value}] > 7.74$  and identifying genes in 500 kb upstream and downstream according to previous work [23], we found five genes in 500 kb upstream and



**Fig. 3** *TaSPL17-A* (*TraesCS7A02G246500*) is associated with spike morphology traits. **a** Local Manhattan plot and pairwise LD analysis showing the association between 20 markers and spike morphology trait in the region between 225,328,984 and 225,737,586 bp on chromosome 7A. Green dots indicate SNPs above the significance threshold ( $-\log_{10}[P \text{ value}] = 5.00$ ). Blue lines in the heatmap highlight strong LD with the significant variants. **b** Relative transcript levels for the three genes (*TraesCS7A02G246500*, *TraesCS7A02G246400*, *TraesCS7A02G246300*) within the associated region at six developmental stages: inflorescence meristem stage (ZM1), spikelet meristem stage (ZM2), glume primordium stage (ZM3), floral meristem stage (ZM4), stamen and pistil primordium stage (ZM5), and floral organ stage (ZM6). **c** Schematic diagram of *TraesCS7A02G246500*. Green rectangles and black lines indicate exons and introns, respectively. The gene is located on the reverse strand. **d, e** Spikes (**d**) and spikelets (**e**) of wild type (WT) and KO lines in the field. Scale bars = 4.00 cm (**d**) and 0.50 cm (**e**). **f** Grain width and length of WT and KO lines in the field. Scale bars = 1.00 cm and 5.00 mm for grain width and grain length, respectively. **g–m** Spike length (**g**), spikelet number per spike (**h**), grain number per spikelet (**i**), grain number per spike (**j**), grain width (**k**), grain length (**l**), and thousand grain weight (TGW) (**m**) in WT and KO lines in the field. Data are shown as means  $\pm$  SD ( $n = 20$ ). Triple KO line, KO-ABD; double KO line (*TraesCS7A02G246500* and *TraesCS7D02G245200* KO line), KO-AD; double KO line (*TraesCS7A02G246500* and *TraesCS7B02G144900* KO line), KO-AB; double KO line (*TraesCS7B02G144900* and *TraesCS7D02G245200* KO line), KO-BD; single KO line (*TraesCS7A02G246500* KO line), KO-A; single KO line (*TraesCS7B02G144900* KO line), KO-B; single KO line (*TraesCS7D02G245200* KO line), KO-D. Different lowercase letters indicate significant differences ( $P < 0.05$ ), as determined by ANOVA

downstream of the 7A: 225,328,984-225,329,125 (Additional file 1: Table S10). Based on the results of these two approaches, only one gene (*TraesCS7A02G246500*) was expressed at high levels during spikelet and floret initiation and development [46, 47], making it a high-confidence candidate, as the other two genes were almost not detectable during this phase (Fig. 3b). The distance between *TraesCS7A02G246500* and the lead SNP was less than 400 kb (Additional file 1: Table S11). The homoeologous genes of *TraesCS7A02G246500* (Fig. 3c) are *TraesCS7B02G144900* (*TaSPL17-B*, B genome) and *TraesCS7D02G245200* (*TaSPL17-B*, D genome).

### Phenotypic analyses of knockout lines

To explore the potential contribution of *TaSPL17-A*, *TaSPL17-B*, and *TaSPL17-D* to grain yield traits in wheat, we generated knockout (KO) lines in all homoeologs by designing one single guide RNA (sgRNA) targeting a conserved region within the third exon of all three genes (Additional file 1: Table S12, Additional file 2: Figure S6). We confirmed the transgenic status of these lines using PCR primers specific for each homoeolog to amplify fragments from both the wild-type (WT) and edited loci. Accordingly, we identified seven independent homozygous lines covering all mutant combinations (Additional file 2: Figure S6). A 1-bp deletion and a 1-bp insertion were most frequently identified at the target site. All mutations caused a shift in the reading frame and introduced premature translation termination codons (Additional file 1: Table S12, Additional file 2: Figure S6). The KO lines exhibited clear effects on plant architecture and the eight traits related to spike assimilate distribution (Additional file 1: Table S13, Additional file 2: Figures S7 and S8). Although these eight traits magnified the variation between the different wheat accessions, they may also enlarge the extent of variation between replicates for each accession, leading to the observed inconsistent effects on related traits (Additional file 1: Table S13). In addition, we also observed substantial differences in plant architecture and spike morphology traits between WT and KO lines (Additional file 1: Table S13). Plants from the triple KO line (KO-ABD) harboring a mutation in all three homoeologs were shorter and produced more tillers than WT, but the double and single KO lines did not show significant differences for these two traits relative to WT (Additional file 1: Table S13, Additional file 2: Figure S7).

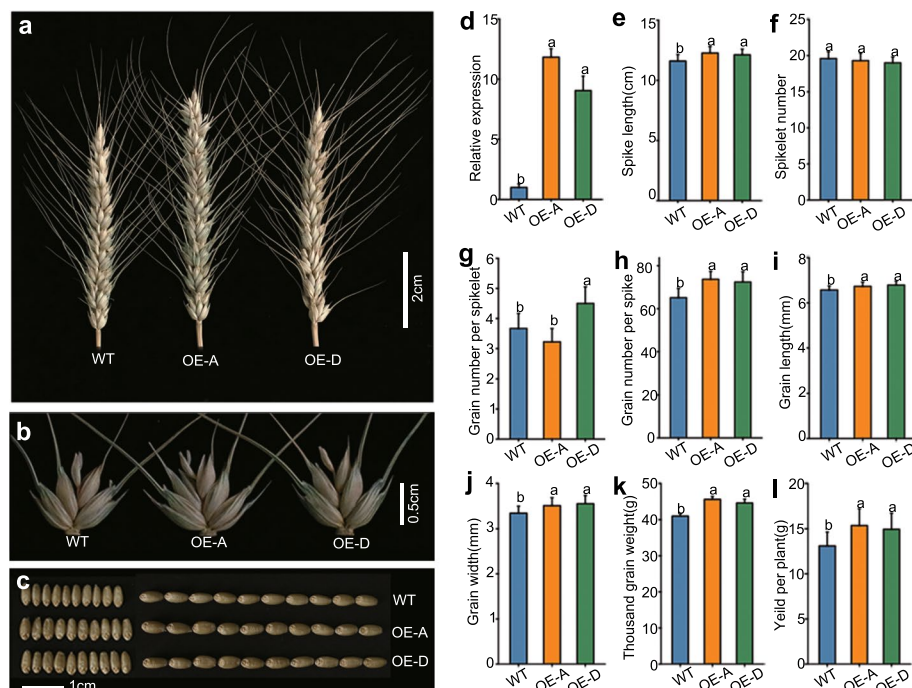
We observed consistent negative effects on spike length and spikelet number per spike in single, double, and triple KO lines ranging from 3.79 to 34.76% and from 5.84 to 32.74%, respectively (Fig. 3d, h, i, Additional file 1: Table S13). The triple KO line (KO-ABD) and one double KO line (KO-AD) exhibited a significant decrease in grain number per spikelet by 21.21% (0.78,  $P < 0.01$ ) and 21.21% (0.78,  $P < 0.05$ ), respectively, compared to WT (Fig. 3e, j, Additional file 1: Table S13). The other double and single KO lines did not obviously influence grain number per spikelet. The smaller spikelet number per spike, alongside the lower or unchanged grain number per spikelet, led to a 4.97-45.77% (3.20-29.45) decrease in grain number per spike in all the KO lines (Fig. 3k, Additional file 1: Table S13). The triple and double KO lines also showed a significant decrease in grain width and length ( $P < 0.05$ ), translating into slight reductions in TGW in these three KO lines (Fig. 3f, l-m, Additional file 1: Table S13). Notably, KO lines had decreased grain number and grain weight (Fig. 2j, m), but they showed higher grain number/spike chaff and grain weight/chaff (Additional file 2: Figure S8). KO lines



negatively affected both grains and chaff. We speculate that the higher grain number/spike chaff and grain weight/chaff maybe due to different sensitivities of grains and spike chaff to the knockout of this gene. This is consistent with the finding that some signals affecting grain number/size and assimilate partitioning traits were detected by GWAS of assimilate partitioning traits, but these signals were not observed in the GWAS results for grain number and size traits.

### Phenotypic analyses of overexpression lines

To determine the effects of higher *TaSPL17* transcript levels, we generated transgenic overexpression lines (OE lines) by transforming WT wheat (Fielder cultivar) with a construct containing the full-length coding sequence of *TaSPL17-A* or *TaSPL17-D* driven by the maize (*Zea mays* L.) *Ubiquitin* promoter. The respective OE lines affected traits associated with assimilate partitioning between spike components (Additional file 2: Figure S8). In addition, we observed differences in plant architecture and spike morphology between WT and OE lines (Fig. 4a–l, Additional file 1: Table S14). We obtained two independent transgenic lines (OE-A for *TaSPL17-A*; OE-D for *TaSPL17-D*) with about tenfold higher transcript levels relative to WT (Fig. 4d, Additional file 1: Table S14). Plant height for the two OE lines was significantly shorter than that of WT (87.78 cm, WT; 79.42 cm, OE-A; 78.20 cm, OE-D;  $P < 0.05$ ); however, the lines showed



**Fig. 4** Phenotypic characterization of *TaSPL17* OE lines in the field. **a, b** Spikes (**a**) and spikelets (**b**) of WT and OE plants. Scale bar, 2.00 or 0.50 cm for spike and spikelet, respectively. **c** Grain width and length of WT and OE lines. Scale bar = 1.00 cm. **d** Relative transcript levels of the corresponding genes in OE and WT. Values are means  $\pm$  SD ( $n = 3$ ). **e–l** Spike length (**e**), spikelet number per spike (**f**), grain number per spikelet (**g**), grain number per spike (**h**), grain length (**i**), grain width (**j**), TGW (**k**), and grain yield per spike (**l**) between WT and OE lines. Data are shown as means  $\pm$  SD ( $n = 20$ ). Different lowercase letters indicate significant differences ( $P < 0.05$ ), as determined by ANOVA

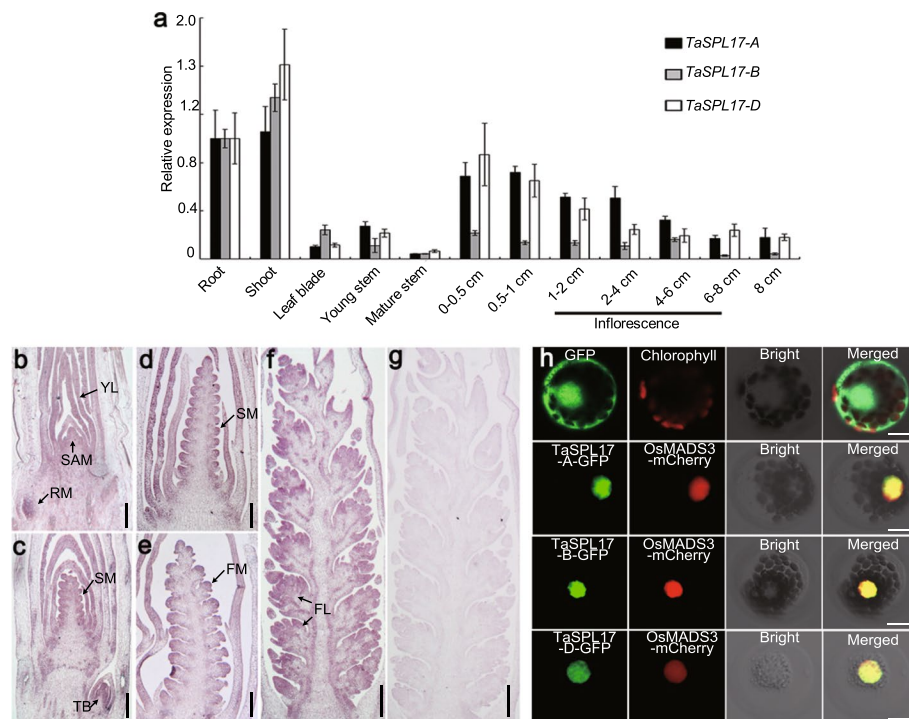
no obvious differences in tiller number (8.20, WT; 7.10, OE-A; 7.40, OE-D) (Additional file 1: Table S14). The two OE lines had increased spike length and grains per spike by 4.48-5.78% (0.52-0.67 cm) and 11.21-13.21% (7.30-8.60), respectively (Fig. 4e, h, Additional file 1: Table S14). Although we observed no significant difference between spikelet number per spike and grain number per spikelet (central spikelet) between OE-A and WT, apical and basal spikelets did have more grains in OE-A, which may explain the increased grain number per spike. The higher values measured for grain width and length resulted in an increase in TGW of 8.81-11.27% (3.61-4.62 g) (Fig. 4i-k, Additional file 1: Table S14) in the two OE lines relative to WT. The higher grain number per spike and higher TGW together with the unchanged tiller number led to a 14.05-17.23% (1.84-2.26 g) increase in grain weight per plant in the two OE lines (Fig. 4m, Additional file 1: Table S14). In summary, overexpressing *TaSPL17* increased grain yield per plant through its positive effects on grain number and size, highlighting the value of *TaSPL17* in wheat breeding.

#### ***TaSPL17* regulates the growth of the inflorescence meristem**

The triple KO line in *TaSPL17* (KO-ABD) produced smaller spikes compared to WT starting at the spike initiation developmental stage throughout spike development with spikes in the triple KO line being less than half the size of WT spikes at the late stage (55 days after germination) (Additional file 2: Figure S9). The rachis expanded and spikelet primordia formed gradually in WT plants, whereas this developmental process was blocked in the triple KO line (Additional file 1: Table S13, Additional file 2: Figure S9). The rachis of the triple KO line was shorter and formed a small number of spikelet meristems before they converted to floret meristems, contributing to the formation of short spikes. The KO/OE lines also showed clear effects on the process of spike development. Compared to WT, the terminal spikelet stage (TS), yellow anther stage (YA), heading stage (HD), and anthesis stage (AN) were promoted in OE lines, but delayed in KO lines (Additional file 2: Figure S10). We concluded that overexpression of *TaSPL17* accelerates spikelet/floret initiation and development.

#### ***TaSPL17* expression pattern and subcellular localization of its gene product**

We collected tissues from root, shoot, leaf blade, young stem, mature stem, and inflorescences at different developmental stages to measure expression levels of *TaSPL17-A*, *TaSPL17-B*, and *TaSPL17-D* (Fig. 5a). We detected high expression levels for all three homoeologs in roots and shoots, but low expression levels in leaf blades and mature stems. All three homoeologs were also highly expressed during early spike developmental stages, but their expression levels tended to decrease later with spike development. The expression level of *TaSPL17-B* was low relative to those of *TaSPL17-A* and *TaSPL17-D* (Fig. 5a). Considering its significant contribution to plant and spike morphogenesis, we examined the transcript pattern of *TaSPL17* in seedlings and young spikes at early differentiation stages using mRNA in situ hybridization. We determined that *TaSPL17* is strongly expressed in young actively dividing tissues such as root meristems, tiller buds, young leaves, and the shoot apical meristem. After the apex development entered the double ridge stage, *TaSPL17* was highly expressed in the spikelet meristem and floral



**Fig. 5** Expression pattern and subcellular localization of *TaSPL17*. **a** RT-PCR analysis showing the relative transcript levels of *TaSPL17-A*, *TaSPL17-B*, and *TaSPL17-D* in various organs (root, shoot, leaf blade, young and mature stems, and inflorescence at different developmental stages). The wheat *Actin* gene was used as an internal control. Values are shown as means  $\pm$  SD of three independent experiments and three biological replicates. **b–f** mRNA in situ hybridization reveals that *TaSPL17* is preferentially expressed in young leaves (YL), tiller buds (TB), root meristems (RM), the shoot apex meristem (SAM), spikelet meristems (SM), floret meristems (FM), and florets (FL). Arrows indicate the locations where *TaSPL17* is expressed. **g** *TaSPL17* sense probe was used as a negative control. **h** Subcellular localization of *TaSPL17* encoded by each subgenome. A construct encoding *TaSPL17-A-GFP*, *TaSPL17-B-GFP*, or *TaSPL17-D-GFP* was co-expressed with *OsMADS3-mCherry*, encoding a nuclear marker. From left to right: image of GFP (green), protoplast, and merged GFP and mCherry. Scale bars, 200  $\mu$ m in **b–e**, 500  $\mu$ m in **f, g**, or 20  $\mu$ m in **h**

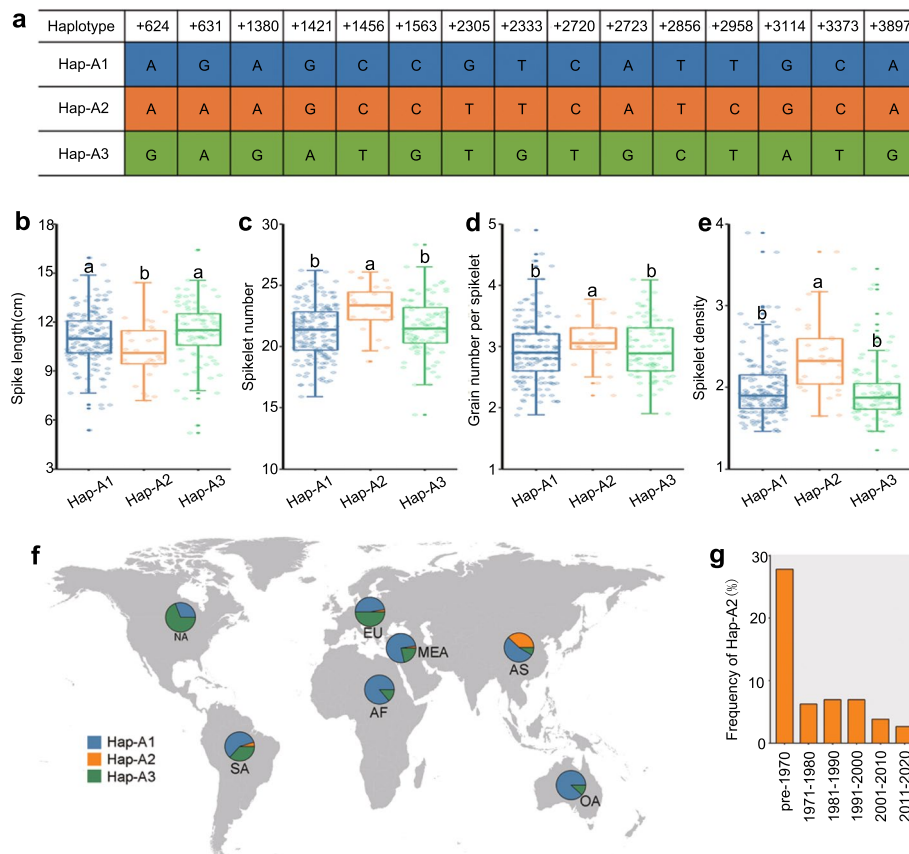
primordia and remained highly expressed in the glume, stamen, and lodicule primordia during the development of floral organs (Fig. 5b–g).

To examine the subcellular localization of *TaSPL17*, we fused *TaSPL17* to the green fluorescent protein (GFP) and co-transfected the resulting encoding constructs with a nuclear marker into wheat leaf protoplasts. *TaSPL17-A-GFP*, *TaSPL17-B-GFP*, and *TaSPL17-D-GFP* fusion proteins all localized exclusively to the nucleus of wheat protoplasts (Fig. 5h), indicating that *TaSPL17* is a nuclear protein and functions in the nucleus.

#### Geographic differentiation and breeding selection of *TaSPL17* haplotypes

Wheat breeding and domestication have exploited natural allelic variation that determines genomic regions and QTLs associated with agricultural traits. Understanding the genomic basis of this phenotypic variation in various germplasm is critical for making accurate selection decisions and for combining desired allelic combinations to improve wheat grain yield. We explored the genetic architecture of *TaSPL17-A* across the 306 worldwide wheat accessions used in this study, which revealed three major

*TaSPL17-A* haplotypes that account for 93.79% (294 accessions) of all haplotypes (Fig. 6a, Additional file 1: Table S15). Haplotype-A1 (Hap-A1), haplotype-A2 (Hap-A2), and haplotype-A3 (Hap-A3) were present in 173, 28, and 93 accessions, respectively (Additional file 1: Table S15). We determined the effects of the three haplotypes on the associated traits (Fig. 6b-e, Additional file 1: Table S15). Compared to Hap-A1, the wheat accessions harboring Hap-A2 had higher average value of spikelet number per spike by 8.85% (1.88) but lower average value of spike length by 7.30% (0.80 cm), leading to a higher spikelet density of 19.08%. Hap-A2 also had higher average value of grain number per spike by 13.78% (6.25) (Additional file 1: Table S16). Among the three haplotypes, we only detected one nonsynonymous SNP (at position + 3897, with the ATG defined as + 1–3 bp) that alters a proline (Hap-A1, Hap-A2) to a leucine residue (Hap-A3, with the changed amino acid at position + 200) (Fig. 6a). Considering the phenotypic differences among the three haplotypes, this nonsynonymous SNP



**Fig. 6** Haplotypes at *TaSPL17-A* affect spike morphology traits. **a** The three major haplotypes at *TaSPL17-A*. The first row indicates positions and the other three rows list the nucleotides for each haplotype. The positions are relative to the ATG of *TaSPL17-A*. + 624 and + 3897 indicate the first and last SNP in the three haplotypes of this gene. **b–e** Spike length (**b**), spikelet number per spike (**c**), grain number per spikelet (**d**), and spikelet density (**e**) between accessions carrying each of the three haplotypes ( $n = 173$  accessions [Hap-A1], 28 accessions [Hap-A2], 93 accessions [Hap-A3]) of *TaSPL17-A*. Different lowercase letters indicate significant differences ( $P < 0.05$ ). **f** Geographical distribution of the three haplotypes for each area (pie chart) AF, Africa; AS, Asia (except Middle East); EU, Europe; MEA, Middle East; NA, North America; OA, Oceania; SA, South America. **g** Frequency of Hap-A2 among 799 Chinese wheat accessions released from 1900 to 2020

was unlikely to be responsible for the observed differences for grain number and size traits between Hap-A2 and Hap-A1 or Hap-A3.

The three major haplotypes of *TaSPL17-A* displayed worldwide geographical differentiation. For each geographical area, we determined the distribution of the three haplotypes (Hap-A1, Hap-A2, Hap-A3) among the local accessions. Hap-A1 had the same sequence as the reference genome (Chinese Spring). We observed that more accessions harbored Hap-A1 (33.33–88.24%) or Hap-A3 (11.11–66.67%) than Hap-A2 (0–36.51%) (Fig. 6f, Additional file 1: Table S17, S18). Hap-A1 contributed to a large fraction of accessions in Oceania (88.24%), Africa (86.21%), and Middle East (76.19%) relative to other areas (Fig. 6f, Additional file 1; Tables S17 and S18). Hap-A2 was mainly present in landraces from China (82.14%) (Fig. 6f, Additional file 1: Tables S19 and S20). Hap-A3 was more widely detected among wheat accessions from North America (66.67%), Europe (50.67%), and South America (36.59%) relative to other areas (Fig. 6f, Additional file 1: Table S18).

We further explored the history of *TaSPL17-A* haplotypes in 799 Chinese wheat accessions released between 1900 and 2020 (Fig. 6g, Additional file 1: Tables S21 and S22). Hap-A1 is positively associated with TGW and increased in frequency from 51.48% of all accessions before 1970 to 68.75% (1971–1980), 75.00% (1981–1990), 68.70% (1991–2000), 67.69% (2001–2010), and 69.54% (2011–2020) (Additional file 1: Tables S21 and S22). Hap-A2 conferring high spikelet number and grain number per spikelet was mainly observed before 1970 and decreased sharply in frequency after 1971 among Chinese wheat accessions (Fig. 6g, Additional file 1: Tables S21 and S22), suggesting that Hap-A2 was selected against in the past 50 years. Tajima's *D* value of 4.16 for the 799 Chinese wheat accessions using VCFtools [48] indicated that these groups were subject to balanced selection. Due to its positive effects on grain number, Hap-A2 therefore has great potential for increasing grain yield in Chinese wheat breeding programs.

To explore the possible explanations for the effects of *TaSPL17-A* haplotypes on grain yield traits, we examined the differences between these three haplotypes. As mentioned above, the nonsynonymous SNP at position +3897 was unlikely to explain the phenotypic differences associated with variations among the three haplotypes of *TaSPL17-A*. *SPL* transcripts are often targeted by microRNAs (miR) from the miR156 family [49, 50]. Although *TaSPL17-A* carried a miR-binding site in its third exon (chr7A: +4070–4091), no SNP mapped to this region among the three haplotypes. Brassinosteroids (BRs) are plant hormones that play important roles in the regulation of plant growth and development. BRI1 EMS SUPPRESSOR 1 (BES1), a transcription factor, accumulates in the nucleus in response to BRs, where it binds to the promoters of and activates BR target genes to specifically regulate BR-mediated gene expression [51, 52]. One SNP (A/G, chr7A: 225,637,350) in the promoter region within a BES1-binding site may explain the phenotypic differences between Hap-A2 and the other two haplotypes. The alleles of the SNP were linked to the three haplotypes. Hap-A2 was linked to a G, while Hap-A1 and Hap-A3 were linked to an A at this position, suggesting a potential role for BRs in regulating spikelet/floret initiation and development.

We determined the LD value ( $r^2$ ) between the three haplotypes and the lead SNPs (Additional file 1: Table S23). The lead SNP: 7A-225,387,666 was in LD with one haplotype SNP: 7A-225,633,304 ( $r^2 = 0.77$ ). In addition, we identified seven



SNPs in *TaSPL17-A* based on Sanger sequencing, and three of which were the same as the SNPs in the haplotypes (Additional file 1: Table S23). We further compared these SNPs with the lead SNPs to investigate causal variations. The lead SNP: 7A-225,328,984 (position: Chr. 7A: 225,328,984, the position in *TaSPL17-A*: + 3897) was identified by Sanger sequencing and was included in the haplotypes (Additional file 1: Table S24). We speculated that SNP: 7A-225,328,984 plays a key role in determining phenotypic variations in this study.

For *TaSPL17-B* (*TraesCS7B02G144900*), we identified four haplotypes in the 306 worldwide wheat accessions that account for 96.08% (294 accessions) of all haplotypes. Haplotype-B1 (Hap-B1), haplotype-B2 (Hap-B2), haplotype-B3 (Hap-B3), and haplotype-B4 (Hap-B4) were present in 88, 59, 129, and 18 accessions, respectively (Additional file 1: Tables S25 and S26). Hap-B1 and Hap-B2 exhibited high similarity to the reference genome relative to Hap-B3 and Hap-B4. Of these four haplotypes, the wheat accessions harboring Hap-B4 had the highest average value of spikelet number per spike, and grain number per spikelet, resulting in the highest grain number per spike of all Hap-B haplotypes (Additional file 1: Tables S25 and S26). The four haplotypes of *TaSPL17-B* displayed worldwide geographical differentiation. Hap-B2 and Hap-B3 accounted for relatively high percentages (38.98 and 27.91%, respectively) in Europe (Additional file 1: Tables S25 and S26). Hap-B1 displayed its highest frequency in Asia (except Middle East) (43.18%) (Additional file 1: Tables S25 and S26). Hap-B4 was more widely distributed in South America (44.44%) relative to other areas (Additional file 1: Tables S25 and S26). We did not observe obvious differences for *TaSPL17-B* (*TraesCS7B02G144900*) haplotypes among the 799 wheat accessions released from 1900 to 2020. We detected only one SNP for *TaSPL17-D* (*TraesCS7D02G245200*), which did not show obvious effects on grain yield traits.

To investigate the effects of *TaSPL17-B* haplotypes on grain yield traits, we examined the differences between these four haplotypes. We did not detect any differences in the coding sequence or in the miR-binding site (chr7B + 3678-3701) between the haplotypes of *TaSPL17-B*. However, we identified three SNPs (chr7B: 187,775,793 bp, C/T; chr7B: 187,775,941 bp, C/G; chr7B: 187,776,637 bp, C/G) in the promoter region of *TaSPL17-B*, differentiating Hap-B4 (C.....C.....C) from Hap-B1, Hap-B2, and Hap-B3 (T.....G.....G). We concluded that the variation in the promoter region may contribute to the phenotypic differences of Hap-B4 with the other three haplotypes by regulating the expression of *TaSPL17-B*.

We also assessed the haplotype diversity for *TaSPL17* across the A and B genomes in the 306 wheat accessions. We detected seven *TaSPL17-A* and *TaSPL17-B* haplotype combinations. In most cases, the examined phenotypic traits displayed greater differences in combinations of haplotypes compared to the *TaSPL17-A* or *TaSPL17-B* haplotypes alone (Additional file 1: Table S15). We observed the most marked differences between the two combinations Hap-A3 + Hap-B3 and Hap-A2 + Hap-B1 (Additional file 1: Table S15). Notably, compared to Hap-A3 + Hap-B3, the combination Hap-A2 + Hap-B1 had increased grain number per spike, grain number per spikelet, and spikelet density by 13.35% (7.10), 7.83% (1.84), and 21.44% (0.51), respectively (Additional file 1: Table S16).

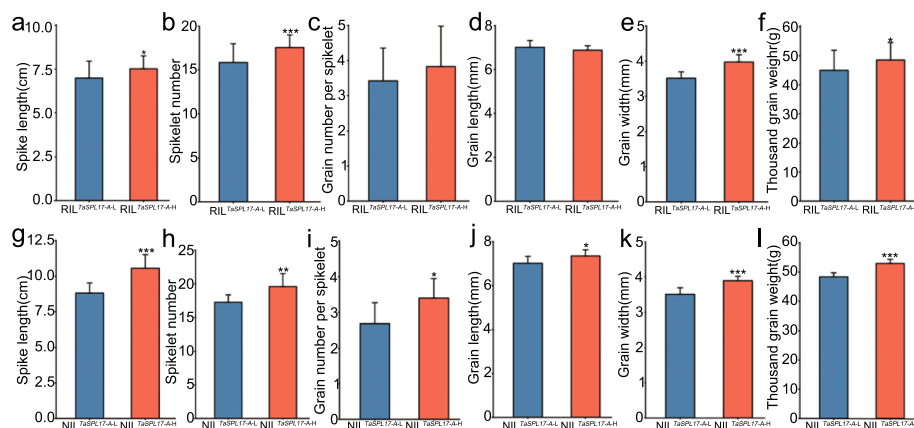
### Utilization of *TaSPL17* alleles for the improvement of grain yield traits

To relate the contributions of *TaSPL17* alleles to grain yield traits and evaluate the value of *TaSPL17* in wheat breeding, we introgressed the *TaSPL17* allele (A/G at +4245 bp in the gene, +1046 bp in the cDNA) conferring high spikelet number, grain number per spikelet, and grain size to improve the potential of grain yield.

Jingshuang16 is a wheat accession with good winter hardiness that was widely planted during the 1980s and 1990s in Northern China. However, its spikelet number, grain number per spikelet, and grain size are relatively low, making it a prime candidate to test the use of an allele that improves these traits. We sequenced the coding sequence of *TaSPL17-A* in multiple accessions and detected a nonsynonymous SNP (A/G at +4245 bp in the gene, +1046 bp in the cDNA) between Bainong64 (an elite wheat cultivar widely planted in China) and Jingshuang16. The nonsynonymous SNP altered one amino acid (+316 protein sequence, Glu/Asp) from aspartate (Bainong64) to glutamate (Jingshuang16).

To understand the contribution of the detected SNPs on grain yield traits, we examined recombinant inbred lines (RILs) generated from the parents Bainong64 (RIL<sup>*TaSPL17-A-H*</sup>, carrying the *TaSPL17-A* allele conferring high spikelet number, grain length, grain width, and TGW) and Jingshuang16 (RIL<sup>*TaSPL17-A-L*</sup>, carrying the *TaSPL17-A* allele conferring low spikelet number, grain length, grain width, and TGW). On average, RIL<sup>*TaSPL17-A-H*</sup> produced longer spikes, more spikelets per spike, wider and longer grains, and greater TGW than RIL<sup>*TaSPL17-A-L*</sup> (Fig. 7, Additional file 2: Figure S11).

We also constructed a near-isogenic line (NIL) that carries the *TaSPL17-A-H* allele from Bainong64 through repeated backcrossing into Jingshuang16. Compared to the parental Jingshuang16 line carrying the *TaSPL17-A-L* allele (NIL<sup>*TaSPL17-A-L*</sup>), the improved Jingshuang16 carrying the *TaSPL17-A-H* allele (NIL<sup>*TaSPL17-A-H*</sup>) had longer spikes (8.80 cm for NIL<sup>*TaSPL17-A-L*</sup>, 10.60 cm for NIL<sup>*TaSPL17-A-H*</sup>,  $P < 0.001$ ), more spikelets per spike (17.30 for NIL<sup>*TaSPL17-A-L*</sup>, 19.60 for NIL<sup>*TaSPL17-A-H*</sup>,  $P < 0.01$ ) and more



**Fig. 7** Introgression of *TaSPL17-A* favorable alleles in recombinant inbred lines (RILs) and near-isogenic lines (NILs) improves spike traits. Mean spike length, spikelet number, grain number per spikelet, grain width, grain length, and TGW are shown for original Jingshuang16 (RIL<sup>*TaSPL17-A-L*</sup> first row [a-f], NIL<sup>*TaSPL17-A-L*</sup> second row [g-l]) and improved Jingshuang16 (RIL<sup>*TaSPL17-A-H*</sup> first row [a-f], NIL<sup>*TaSPL17-A-H*</sup> second row [g-l]). Data are shown as means  $\pm$  SD. Significant differences were determined by Student's *t* test (two sided, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ )

grains per spikelet (2.70 for NIL<sup>TaSPL17-A-L</sup>, 3.40 for NIL<sup>TaSPL17-A-H</sup>,  $P < 0.05$ ), larger grains (grain width, 3.50 mm for NIL<sup>TaSPL17-A-L</sup>, 3.90 mm for NIL<sup>TaSPL17-A-H</sup>,  $P < 0.001$ ), longer grains (7.00 mm for NIL<sup>TaSPL17-A-L</sup>, 7.40 mm for NIL<sup>TaSPL17-A-H</sup>,  $P < 0.05$ ), and greater TGW (48.40 g for NIL<sup>TaSPL17-A-L</sup>, 53.00 g for NIL<sup>TaSPL17-A-H</sup>,  $P < 0.05$ ) (Fig. 7, Additional file 2: Figure S11). These results suggest that *TaSPL17* alleles have the potential to improve wheat grain yield.

## Discussion

This study achieves two significant objectives: first, we developed a large-scale phenotype-genotype map of 27 spike morphology traits for 306 worldwide wheat accessions; second, we clarified the mechanism of *TaSPL17* regulating grain number and size in wheat. In addition, we elaborated on the geographical differentiation and breeding selection of *TaSPL17* haplotypes and their value in wheat breeding. For the first objective, we carried out GWAS using 40 million SNPs. Although a substantial number of wheat GWAS has been previously conducted, many of these studies on complex traits were underpowered owing to the relatively small number of SNPs used (< 1 million).

Several studies reported that LD for wheat genome was above 1 Mb ( $r^2 = 0.50$ ) [23, 43, 44]. In this study, based on  $r^2 = 0.50$ , LD was 0.32, 0.007, 0.19 and 0.116 Mb for A, B, D, and whole genome, respectively. This observation is due to the strong genomic variation among the 306 worldwide wheat accessions. Traditionally, fine-mapping of a candidate gene in wheat would involve the development of several genetic populations and the identification of sufficient markers to obtain good genomic coverage, which may take several years. The large number of SNPs (~ 40 million SNPs, 2.8 SNPs per kb) and low LD used in this study allowed us to identify SNPs that are within or proximal to the candidate genes. This information may make it easier to obtain relatively precise locations of candidate genes.

Previous work identified very few major-effect genomic regions for grain yield, since grain yield is associated with grain number and size, which are influenced by different factors. We thus opted to dissect grain weight, number, and size using 27 traits. The GWAS for assimilate partitioning traits revealed strong signals that would not have been observed with GWAS conducted directly on grain yield, number, or size. The haplotypes at the identified peaks exhibited significant effects for the traits of grain number and size. The dissection of spike morphology using 27 traits allowed us to detect additional peaks, which is a powerful strategy to detect signals for grain yield traits in wheat.

Spikelet number per spike and grain number per spikelet are two determinants of grain number per spike in wheat. In addition, the “paired spikelet” is another promising trait for improving wheat spikelet number [53–56]. Spikelet number per spikelet is relatively stable in variable environments. Although a relatively large number of genes regulating spikelet number has been investigated [57–62], few studies have been reported in wheat on the molecular mechanism of floret fertility/grain number per spikelet [63], because grain number per spikelet exhibits high sensitivity in variable environments. In this study, we identified a gene regulating floret fertility/grain number per spikelet in wheat. The tradeoff between grain number and size negatively influences grain yield in wheat [64, 65]. We found that overexpressing *TaSPL17* increased grain number per

spikelet and TGW, overcoming the tradeoff between these two determinants of grain yield in wheat.

This study also illustrated the connection between known genes and spike morphology traits. Our work suggests that *VRN-A1*, *VRN-A2*, *Ppd-D1*, and *TaFT-A5* are not only associated with spike length, spikelet number, and density, but also connected to the assimilate partitioning among spike components (Fig. 2, Additional file 1: Table S8). In addition, *TaAGL-A6*, *TaAGO1d-A*, *TaVRS-B3*, and *TaCEN-B* consistently displayed close connections with two traits exhibiting the assimilate distribution between grains and chaff within individual spikelet: grain weight/spikelet weight (individual spikelet), and chaff weight/spikelet weight (individual spikelet) (Fig. 2, Additional file 1: Table S8).

*TaSPL17* encodes a plant-specific transcription factor from the SPL family, which regulates the expression of target genes by binding to their cognate GTAC cis-element [66]. SPL transcription factors are closely associated with grain yield in crops. In rice, *OsSPL13* and *OsSPL14* promote the development of spike branches to enhance grain number per spike and grain yield [35, 37, 67]. *OsSPL16* determines grain size and shape and improves rice yield [68, 69]. Haplotypes of *TaSPL21* are associated with wheat TGW [70]. Ectopic expression of *TaSPL16* in *Arabidopsis thaliana* increases organ size [71]. The indeterminate nature of wheat spikelets allows more than eight florets to form within one spikelet, in contrast to the fixed number in other cereals (rice, barley, and maize) [72–74]. Overexpressing *TaSPL13* increases florets and grains per spikelet in wheat [75]. By contrast, our results demonstrated that *TaSPL17* was highly expressed in spikelet and floral primordia and that the triple KO line lacking function of all three *TaSPL17* homoeologs suppressed the development of spikelets and florets, indicating that *TaSPL17* triggers spikelet and floret initiation and development.

Although some genes controlling grain number or size have been cloned, the effect of these genes in wheat breeding has been questioned. Our work explored the geographical distribution and breeding selection of *TaSPL17* haplotypes as well as their role in the improvement of grain yield in wheat. One haplotype of *TaSPL17* (Hap-A2), which improves grain number, was mainly present in landraces in China. We mainly identified Hap-A2 in Chinese wheat accessions released before 1970, and its frequency dropped from 1971 to 2020, suggesting that Hap-A2 was lost during modern wheat breeding in China. In China, wheat yield has achieved a great increase from the 1990s to today, mainly due to the genetic increase in TGW [76, 77]. Notably, the Yellow and Huai Winter Wheat Region is the most important wheat region in China. Wheat varieties in this region have relatively high spike number per unit area and medium grain number per spike. Although Hap-A2 exerts positive effects on grain number, it is not helpful in increasing grain size, which may partially explain why Hap-A2 was neglected during modern breeding in Chinese cultivars. When we introduced favorable *TaSPL17* alleles into modern wheat accessions, we observed improved grain number and size.

## Conclusion

In summary, our study provides a large amount of new genomic resources to further understand the genetic basis of spike morphology and to facilitate future breeding to improve wheat production.

## Methods

### Planting and phenotyping

For GWAS analyses, experiments were carried out in Zhaoxian, Hebei (37° 27' N, 113° 30' E, altitude 78 m) for two consecutive years. We planted six rows for each accession, and each row (1.50 m) included 15 plants, with 10.00 cm between rows. All field managements (e.g., irrigation, weed management, fertilization) were performed according to normal standards. Plants were irrigated when required. We measured 27 spike morphology traits at physiology maturity and calculated 15 ratios based on the 12 traits using 306 worldwide wheat accessions 2 years. We randomly selected the main shoots of five plants for each accession to measure the 12 traits. The 12 traits include spike length (cm), spike dry weight (g), grain weight per spike (g), awn weight per spike (g), spike chaff weight per spike (g), grain number per spike, total spikelet number per spike, rachis weight per spike (g), spikelet weight (g), grain number per spikelet, chaff weight per spikelet (g), grain weight per spikelet (g). Spike chaff includes glume, lemma, palea, and awn. Spike length does not include awns. Spikelet weight, grain number per spikelet, and chaff weight per spikelet were determined based on the spikelets in the center of spike.

Fruiting efficiency is often calculated at physiological maturity as the grain number per unit of chaff dry weight and has been proposed as a target trait to improve grain number and yield potential [78–80]. Therefore, fruiting efficiency was calculated for the entire spike and individual spikelets to use their results as traits for GWAS: grain number/spike chaff weight (entire spike) and grain number/spike chaff weight (individual spikelet). Wheat awns are long needle-like structures extending from the lemma in the florets and contribute to photosynthesis and grain production [81, 82]. Four traits were calculated to explore the contribution of awn to grains: awn weight/spike weight (entire spike), awn weight/grain weight (entire spike), awn weight/spike chaff weight (entire spike), and awn weight per spike. Wheat vascular bundles along the rachis constitute the transport system through which assimilates are transported to developing florets [83]. Grain weight/rachis weight (entire spike) was therefore used to assess the genetic connection between grains and rachis.

In total, the 12 measured traits were used to determine 15 traits to dissect spike components. The 15 traits comprised ten ratios for an entire spike and five ratios for individual spikelets. The ten ratios for the entire spike were thousand kernel weight ((grain weight/grain number) × 1000), the ratio between grain number per spike and spike chaff per spike (grain number/spike chaff weight), the ratio between grain weight per spike and spike chaff weight (grain weight/spike chaff weight), the ratio between grain weight per spike and spike weight (grain weight/spike weight), the ratio between grain weight per spike and rachis weight per spike (grain weight/rachis weight), spikelet density (the ratio between total spikelet number and spike length, total spikelet number/spike length), the ratio between spike chaff weight and spike weight (spike chaff/spike weight), the ratio between awn weight per spike and spike weight (awn weight/spike weight), the ratio between awn weight per spike and grain weight per spike (awn weight/grain weight), and the ratio between awn weight per spike and spike chaff weight per spike (awn weight/spike chaff). The five ratios for individual spikelet traits consisted of the ratio between grain weight per spikelet and spikelet chaff (grain weight/spikelet chaff weight), thousand kernel weight ((grain weight/grain number) × 1000) for the spikelet in



the center of spike, the ratio between grain number per spikelet and spike chaff weight (grain number/spike chaff weight), the ratio between grain weight per spikelet and spikelet weight (grain weight/spikelet weight), and the ratio between spikelet chaff weight and spikelet weight (spikelet chaff weight/spikelet weight).

For the phenotypic analyses of WT and transgenic lines, WT (cv. Fielder) and transgenic wheat were grown in an experimental field at the Chinese Academy of Agricultural Sciences in Beijing during the natural growing season. For the determination of grain width, length, spike length, and spikelet number per spike, the main shoots of 20 plants from each genotype were randomly selected. For the determination of grain number per spikelet (at the center of the spike) and TGW, nine and three spikes, respectively, were selected from the main shoots of different plants. For the RNA in situ hybridization and subcellular localization analysis, plants were grown in pots containing peat soil in controllable greenhouses with 70% relative humidity at 25 °C/20 °C, with a 16-h light/8-h dark photoperiod.

#### Genotype calling and SNP identification

SNP discovery and genotyping used a workflow by Zhou Y for the construction of the whole-genome genetic variation map of wheat [20]. The latest version of VMap (VMap 2.0) is based on 1062 wheat accessions with multiple ploidy levels. We selected 306 hexaploid wheat accessions from the current version of VMap (VMap 2.0) for this study. A total of 40,710,923 segregating SNPs with MAF > 0.05 were used in the GWAS.

Genotype imputation estimates missing genotypes from the reference panel. Genotype imputation is commonly performed in GWAS because it greatly increases the number of markers [84–86]. The SNPs for 799 wheat accessions were obtained by genotype imputation based on the SNPs of 306 worldwide wheat accessions. Genotype imputations were conducted based on the source data according to previous work [84–86].

#### Phylogenetic tree and population structure

To clarify the phylogenetic relationship from a genome-wide perspective, an individual-based neighbor-joining tree was constructed according to the *P* distance in Tassel (version 5.2.64) software and visualized using ITOL (<https://itol.embl.de/>). The population genetic structure was examined via an expectation maximization algorithm, as implemented in the program Admixture [87]. The number of assumed genetic clusters *K* ranged from 1 to 13, with 10,000 iterations for each run. Principle component analysis (PCA) was also performed to evaluate genetic structure in Plink (version 1.90b6.18) software. PCA was carried out to remove linked SNPs based on LD blocks and generated 2,756,291 SNPs for population structure analysis.

#### LD analysis

To conduct pairwise LD analysis of the associated genomic region for the *TaSPL17-A*, the software PopLDdecay [88] was used to calculate the LD coefficient ( $r^2$ ) between pairwise high-quality SNPs; the parameters were set as follows: `–allow-no-sex –maf 0.05 –geno 0.2 –r2 –ld-window 999999 –ld-window-r2 0`.

To determine genomic intervals, the associated SNPs of all traits above a threshold of five were combined and de duplication were extracted by VCFtools software. LD

between SNPs was calculated by Plink: `plink -vcf chr_ position. vcf. recode. vcf -allow-no-sex -maf 0.05 -geno 0.2 -r2 -ld-window 50000 -ld-window-r2 0 -out out_ file_ prefix2`. The results were used to combine SNPs to intervals. The significant markers were then delineated into intervals based on the LD between markers, where markers with  $r^2 > 0.1$  for the existence of LD were included in the same interval. If the distance between the peak SNPs of two adjacent loci was less than 5 Mb, these two loci were merged. The number of significant SNPs contained in the genomic region was also calculated. If the number of SNPs in corresponding genomic region was more than 10, these regions were defined as peaks or genomic regions. The most significant SNP means the peak SNP.

### Multiple comparisons and best linear unbiased estimates

The analysis was carried out using the agricolae package in R (agricolae: Statistical Procedures for Agricultural Research. R package version 1.3–1. <https://CRAN.R-project.org/package=agricolae>). The means of multiple groups were compared using the best significant difference (LSD) test, which required a one-way ANOVA.

The best linear unbiased estimates (BLUES) for 27 traits across 2 years were calculated using the lme4 package in R [89] with the linear mixed linear model:

$$Y_{ij} = u + \text{genoty}_i + \text{Year}_j + (\text{genoty} \times \text{Year})_{ij} + \text{error}_{ij}$$

Where  $Y_{ij}$  is the trait of the  $i$ th genotype in  $j$ th year;  $u$  is the mean;  $\text{genoty}_i$  is the genotype effect of the  $i$ th genotype;  $\text{Year}_j$  is the effect of the  $j$ th year;  $(\text{genoty} \times \text{Year})_{ij}$  is the genotype-year interaction;  $\text{error}_{ij}$  is the error of the  $j$ th year and items were set to random.

### Broad sense heritability

Phenotypic data from all environments were analyzed by analysis of variance (ANOVA) in R. Broad sense heritability ( $H^2$ ) of 27 traits was calculated across environments from variance components according to the formula:

$$H^2 = \left( \sigma_G^2 \right) / \left( \sigma_G^2 + \left( \sigma_{GE}^2 \right) / 2 + \left( \sigma_e^2 \right) / 2r \right)$$

Where  $\sigma_G^2$  is the genotypic variance,  $\sigma_{GE}^2$  is the genotype by the year effect,  $\sigma_e^2$  is the residual error, 2 means two years and  $r$  is the average number of replications ( $r=5$ ).

### GWAS

A large-scale GWAS on 27 spike morphology traits was carried out, using 40,710,923 SNPs ( $MAF > 0.05$ ). Genome-wide association analysis performed in GEMMA (version 0.98.4) software using the mixed linear model (MLM) analysis:

$$y = X\alpha + S\beta + K\mu + e$$

where  $y$  represents phenotype;  $\alpha$  and  $\beta$  are fixed effects representing marker effects and non-marker effects; and  $\mu$  represents unknown random effects.  $X$ ,  $S$ , and  $K$  are

the incidence matrices for  $\alpha$ ,  $\beta$ , and  $\mu$ , respectively, and  $e$  is a vector of random residual effects. The top three PCs were used to build up the  $S$  matrix in Plink (version 1.90b6.18). The matrix of simple matching coefficients was used to build up the kinship ( $K$ ) matrix. The  $S$  matrix and kinship ( $K$ ) matrix were performed for population structure correction. To determine the threshold, we performed permutation test for each trait by disrupting each 100 times and performed 100 GWAS analyses according to the MLM model. We ranked these  $P$ -values from smallest to largest for each trait and used a significance level of  $\alpha=0.5$  and the 0.05 quantile of ranked  $P$ -values as the corrected significance level. The detailed results of the 100 runs of the traits are shown in Additional file 1: Table S27, with the significance levels of 7.67–18.93 for 27 traits. We also performed Bonferroni correction of  $0.05/n$ , where  $n$  is the number of independent markers determined by Plink (version 1.90b6.18) (window size 50,000 bp, step size 5,  $r^2=0.2$ ,  $n=2,756,291$ ). This threshold of  $P$  value was determined by  $0.05/2,756,291$  ( $P=1.81e-07$ ),  $-\log_{10} [P \text{ value}]=7.74$ . To balance false positives and false negatives, so we conservatively chose  $-\log_{10} [P \text{ value}]=5.00$  as a moderate threshold for calling significant association according to previous GWAS work in wheat [12, 23, 43–45]. The GWAS results were visualized by Manhattan plots generated with the “CMplot” R package (<https://github.com/YinLiLin/R-CMplot>). The candidate genes in associated regions were identified and confirmed as significantly associated with the corresponding traits based on the two thresholds ( $-\log_{10} [P \text{ value}]=5.0$  and 7.74), which displayed consistent results. The trait-associated markers were aligned to the wheat reference genome RefSeq v.1.0 and visualized using Phenogram (<http://visualization.ritchielab.org/phenograms/plot>).

#### RNA extraction and reverse transcription-quantitative PCR (RT-qPCR) analyses

Total RNA was extracted from various tissues using an RNeasy plant mini kit (Qiagen, Valencia, CA) in accordance with the manufacturer’s instructions and then treated with DNase I (TaKaRa, Tokyo, Japan) to remove genomic DNA contamination. First-strand cDNA was synthesized using a PrimeScript first-strand cDNA synthesis kit (TaKaRa, Tokyo, Japan). qPCR was performed using gene-specific primers and SYBR Premix Ex Taq reagents (Takara, Tokyo, Japan) on an ABI 7500 Fast real-time PCR system (Applied Biosystems, Foster City, CA). The wheat *Actin* gene was used as an internal control. qPCR was performed on three independent biological replicates. For each biological replicate, every sample was collected from at least six individual plants.

#### Histological analysis

For microscopy observations, shoot apices of WT and the triple KO line (KO-ABD) plants were collected every 5 days from the vegetative stage, shortly before phase transition, to the end of floral differentiation. The shoot apices were carefully dissected and fixed in FAA solution (50% [v/v] ethanol, 5% [v/v] glacial acetic acid and 5% [v/v] formaldehyde). The samples were embedded in Paraplast (Leica Microsystems, Germany) after dehydration through a graded ethanol series (70% [v/v], 80%, 90%, and 100%) and hyalinized using a xylene:ethanol series with v/v ratios of 1:2, 1:1, 2:1, and 1:0. Then, the samples were sectioned into 8- $\mu$ m sections using a Leica RM2245 rotary microtome. After removing Paraplast using a series of xylene and ethanol solutions, the sections

were stained with toluidine blue before images were captured using a Leica ICC50 HD microscope.

#### **RNA in situ hybridization**

RNA in situ hybridization was performed as described previously (Bradley et al., 1993) [90]. A specific 347-bp fragment of the *TaSPL17*-specific coding region was subcloned into the pGEM-T Easy vector (Promega, Madison, Wisconsin, USA) and used as a template to generate both antisense and sense RNA probes. Shoot apices of WT wheat seedlings at the third-leaf stage and young inflorescences were fixed using an RNase-free FAA fixative solution and then embedded in Paraplast (Leica Microsystems, Germany). Digoxigenin-labeled RNA probes were prepared using a DIG Northern starter kit (Roche) according to the manufacturer's instructions. Slides were observed under brightfield using a Leica DMR microscope.

#### **Subcellular localization analysis**

The coding sequences of *TaSPL17-A*, *TaSPL17-B*, and *TaSPL17-D* were individually cloned in-frame and downstream of the sequence of *GFP* and placed under the control of the cauliflower mosaic virus (CaMV) 35S promoter in the vector pAN580. Protoplasts were prepared from wheat mesophyll cells, and the *TaSPL17-A/B/D-GFP* fusion constructs and the nuclear marker *OsMADS3-mCherry* were co-transfected into wheat leaf protoplasts for transient expression using the polyethylene glycol (PEG4000)-mediated method. A Leica TCS-SP4 confocal microscope was used for fluorescence detection.

#### **Construction of gene editing and OE vectors**

To generate mutants in *TaSPL17-A*, *TaSPL17-B*, and *TaSPL17-D* (*TraesCS7A02G246500*, *TraesCS7B02G144900*, and *TraesCS7D02G245200*, respectively), two sgRNA target sequences were designed according to the conserved exon sequences of *TaSPL17-A*, *TaSPL17-B*, and *TaSPL17-D* using the online tool CRISPR-P 2.0 (<http://crispr.hzau.edu.cn/CRISPR2/>). A segment containing the wheat *U3* promoter was amplified from the pCBC-MT1T2 vector using a pair of primers containing the two designed sgRNAs and then cloned into the *Bsa*I site of the clustered regularly interspaced short palindromic repeats CRISPR/Cas9 vector pBUE411, using T4 DNA ligase (TransGen Biotech, China).

To simultaneously disrupt *TaSPL17* in all subgenomes (A, B, and D), two sgRNAs (sg1 and sg2) were designed targeting the conserved regions within the first and third exons of *TaSPL17-A*, *TaSPL17-B*, and *TaSPL17-D*, and the corresponding sgRNA/Cas9 vector was transformed into the wheat cultivar Fielder via *Agrobacterium* (EHA105) (*Agrobacterium tumefaciens*)-mediated transformation (Figure S3). Twenty bialaphos-resistant transgenic events harboring the T-DNA of the sgRNA/Cas9 vector were identified. *TaSPL17-A*, *TaSPL17-B*, and *TaSPL17-D*-specific PCR primers were designed to amplify fragments from both the WT *TaSPL17* and edited *TaSPL17* loci. DNA sequencing revealed that 50% (10 of 20)  $T_0$  events harbored compound heterozygous mutations at the gRNA2 site, whereas no mutations were detected at the gRNA1 site. Seeds of 10 mutant  $T_0$  events were collected and planted in the greenhouse, and 10  $T_1$  plants for each  $T_0$  event were selected to detect mutation types

at the gRNA2 site. A 1-bp deletion and a 1-bp insertion were most frequently identified at the target site. All mutations caused a frameshift leading to the introduction of premature translation termination codons.

To avoid the influence of potential off-target sites on phenotypic and functional analyses, the four most likely off-target sites with two or three mismatches were selected for amplification and analysis using specific primers. The results of sequencing did not reveal any mutations in the examined potential off-target sites of all 25 homozygous *TaSPL17* mutants.

To generate overexpressing plants, the coding sequences of *TaSPL17-A*, *TaSPL17-B*, and *TaSPL17-D* were individually amplified using DNA polymerase KOD FX (Toyobo, Osaka, Japan) from WT cDNA and cloned into the pMWB110 vector [91] using an In-Fusion Advantage PCR cloning kit (Clontech Laboratories, Takara, Tokyo, Japan). The resulting plasmids were transformed into the spring wheat cultivar Fielder via *Agrobacterium*-mediated transformation [92]. All primer sequences used in this study are listed in Supplemental Table S28.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-023-03044-2>.

**Additional file 1: Table S1.** Summary of the 306 worldwide wheat accessions. **Table S2.** Diversity and genetic structure analyses of the 306 wheat accessions. **Table S3.** The determination of linkage disequilibrium ( $r^2$ ) in this study. **Table S4.** The heritability of 27 traits in this study. **Table S5.** The effects of genotype, environment, genotype  $\times$  environment interaction and heritability for the 27 traits in this study. **Table S6.** Overview of the 27 spike morphology traits in this study. **Table S7.** Detailed information of all significant associated SNPs for the investigated traits ( $-\log_{10}[P \text{ value}] > 5.00$ ). **Table S8.** Detailed information of all genomic intervals for the investigated traits. **Table S9.** The effects of 392 genomic intervals associated with assimilate partitioning for the grain yield traits. **Table S10.** The expression of candidate genes (FPKM). **Table S11.** The distance between lead SNPs and *TraesCS7A02G246500*. **Table S12.** Targeted mutagenesis of *TaSPL17* in different mutant lines. **Table S13.** The phenotypic data of knock out lines in this study. **Table S14.** The phenotypic data of overexpression lines in this study. **Table S15.** Overview of spike morphology traits among the haplotypes of *TaSPL17*. **Table S16.** The differences of spike morphology traits among the haplotypes of *TaSPL17*. **Table S17.** Worldwide distribution of the three major haplotypes of *TaSPL17-A*. **Table S18.** The percentage of worldwide distribution of the three major haplotypes of *TaSPL17-A*. **Table S19.** The Hap-A2 information of *TaSPL17-A* in 306 wheat accession. **Table S20.** The percentage of the three major haplotypes of *TaSPL17-A* in different regions. **Table S21.** The distribution of the twelve major haplotypes of *TaSPL17-A* in 799 Chinese wheat accessions. **Table S22.** The percentage of the three major haplotypes of *TaSPL17-A* in 799 Chinese wheat accessions. **Table S23.** The linkage disequilibrium ( $r^2$ ) between lead SNPs and haplotype SNPs. **Table S24.** The identified SNPs of *TaSPL17-A* by Sanger sequencing. **Table S25.** The distribution of the four major haplotypes of *TaSPL17-B* in different areas. **Table S26.** The distribution of the four major haplotypes of *TaSPL17-B* in different areas. **Table S27.** The results of permutation test. **Table S28.** Primers used in this study.

**Additional file 2: Figure S1.** Cross-validation (CV) errors of ADMIXTURE runs. **Figure S2.** Overview of phenotypic data according to the geographical origin of accessions. **Figure S3.** The distribution of phenotypic values of the 27 spike morphology traits for the 306 wheat accessions. **Figure S4.** The associations among the 27 spike morphology traits. **Figure S5.** Manhattan plot displaying the GWAS result of the eight traits. **Figure S6.** Homozygous targeted mutagenesis of *TaSPL17*. **Figure S7.** Phenotypic characterization of the *TaSPL17* mutant lines in field. **Figure S8.** Traits associated with spike assimilate partitioning in wild type (WT), overexpression (OE) and knock out (KO) lines. **Figure S9.** Early stages of spike development in wild type (WT) and the triple knockout (KO) line. **Figure S10.** Duration of the terminal spikelet stage (TS), yellow anther (YA), heading stage (HD) and anthesis stage (AN) in wild type (WT), the overexpression (OE) and the triple knockout (KO) lines. **Figure S11.** The spikes for NILs and RILs of *TaSPL17* in field.

**Additional file 3.** Review history.

## Acknowledgements

We thank Prof. Kang Chong, Prof. Tai Wang, Prof. Yunyuan Xu (Institute of Botany, Chinese Academy of Sciences) and Prof. Zefu Lu (Institute of Crop Science, Chinese Academy of Agricultural Sciences) for their valuable suggestions and comments. We thank Dr. Dongdong Li and Dr. Shuqin Jiang (College of Agronomy and Biotechnology, China Agricultural University, Beijing, China) for the suggestions of data analyses.

## Review history

The review history is available as Additional file 3.



**Peer review information**

Wenjing She was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Authors' contributions**

Z.G., F.L., Y.M., Y.H., and Z.H. designed and supervised the experiment. Y.L., Z.Z., L.K., X.S., A.B., X.Z., D.X., X.Z., and C.H. performed sequencing, genome variant calls, and GWAS analyses. J.C., Y.L., and Y.G. carried out gene expression analysis and functional validation. Y.L., C.Y., Z.W., H.W., K.S., J.W., X.Y., Z.S., B.Y., D.L., L.Z., and L.S. conducted field experiments and phenotyping. Y.H. and Z.H. developed the genetic populations. Y.L. and J.C. plotted manuscript figures. All authors were involved in the writing and editing of the manuscript.

**Funding**

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDA24010104-2), the Agricultural Science and Technology Innovation Program (CAAS-ZDRW202109) and the National Key Research and Development Program (2020YFE0202300, 2022YFF1002904), the National Natural Science Foundation of China (32225038, 31921005, and 32270269), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA24020201), and the Hainan Yazhou Bay Seed Lab (B21HJ0001 and B21HJ0111).

**Availability of data and materials**

The genotypes of 306 wheat accession used in this study have been deposited in the Genome Variation Map (<https://bigd.big.ac.cn/gvm>) under accession number GVM000463 [20, 26–28]. The genes in Additional file 1: Table S8 [24, 30, 93–134] in the study were reported previously.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. <sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China. <sup>3</sup>Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100081, China. <sup>4</sup>State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 10011, China. <sup>5</sup>International Maize and Wheat Improvement Center (CIMMYT) China Office, c/o CAAS, Beijing 100081, China. <sup>6</sup>State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University, Yangling 712100, Shaanxi, China. <sup>7</sup>CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100093, China.

Received: 27 April 2023 Accepted: 21 August 2023

Published online: 28 August 2023

**References**

- Rosegrant MW, Cline SA. Global food security: challenges and policies. *Science*. 2003;302:1917–9.
- Shewry PR, Hey SJ. The contribution of wheat to human diet and health. *Food Energy Secur*. 2015;4:178–202.
- Shiferaw B, Smale M, Braun HJ, Duveiller E, Reynolds M, Muricho G. Crops that feed the world 10 Past successes and future challenges to the role played by wheat in global food security. *Food Secur*. 2013;5:291–317.
- Chand R. Challenges to ensuring food security through wheat. *CAB reviews: perspectives in agriculture, veterinary science, nutrition and natural resources*. 2009.
- Singh RP, Singh PK, Rutkoski J, Hodson DP, He XY, Jorgensen LN, Hovmoller MS, Huerta-Espino J. Disease impact on wheat yield potential and prospects of genetic control. *Annu Rev Phytopathol*. 2016;54(54):303–22.
- Wheeler T, von Braun J. Climate change impacts on global food security. *Science*. 2013;341:508–13.
- Zampieri M, Ceglar A, Dentener F, Toreti A. Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. *Environ Res Lett*. 2017;12:064008.
- Gonzalez FG, Terrile II, Falcon MO. Spike fertility and duration of stem elongation as promising traits to improve potential grain number (and yield): variation in modern argentinean wheats. *Crop Sci*. 2011;51:1693–702.
- Pretini N, Vanzetti LS, Terrile II, Donaire G, Gonzalez FG. Mapping QTL for spike fertility and related traits in two doubled haploid wheat (*Triticum aestivum* L.) populations. *BMC Plant Biol*. 2021;21:353–71.
- Guo ZF, Chen DJ, Alqudah AM, Roder MS, Ganai MW, Schnurbusch T. Genome-wide association analyses of 54 traits identified multiple loci for the determination of floret fertility in wheat. *New Phytol*. 2017;214:257–70.
- Guo ZF, Chen DJ, Roder MS, Ganai MW, Schnurbusch T. Genetic dissection of pre-anthesis sub-phase durations during the reproductive spike development of wheat. *Plant J*. 2018;95:909–18.
- Juliana P, Poland J, Huerta-Espino J, Shrestha S, Crossa J, Crespo-Herrera L, Toledo FH, Govindan V, Mondal S, Kumar U, et al. Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat Genet*. 2019;51:1530–42.
- Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*. 2012;491:705–10.

14. Jia JZ, Zhao SC, Kong XY, Li YR, Zhao GY, He WM, Appels R, Pfeifer M, Tao Y, Zhang XY, et al. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*. 2013;496:91–5.
15. International Wheat Genome Sequencing C. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*. 2014;345:286–99.
16. Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo NX, Zhu TT, Wang L, Wang Y, et al. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature*. 2017;551:498–519.
17. Zhao GY, Zou C, Li K, Wang K, Li TB, Gao LF, Zhang XX, Wang HJ, Yang ZJ, Liu X, et al. The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nature Plants*. 2017;3:946–55.
18. Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak C, Stein N, Choulet F, Distelfeld A, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*. 2018;361:661–73.
19. Ling HQ, Ma B, Shi XL, Liu H, Dong LL, Sun H, Cao YH, Gao Q, Zheng SS, Li Y, et al. Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature*. 2018;557:424–33.
20. Zhou Y, Zhao XB, Li YW, Xu J, Bi AY, Kang LP, Xu DX, Chen HF, Wang Y, Wang YG, et al. *Triticum* population sequencing provides insights into wheat adaptation. *Nat Genet*. 2020;52:1412–22.
21. Scott MF, Fradgley N, Bentley AR, Brabbs T, Corke F, Gardner KA, Horsnell R, Howell P, Ladejobi O, Mackay IJ, et al. Limited haplotype diversity underlies polygenic trait architecture across 70 years of wheat breeding. *Genome Biol*. 2021;22:137–67.
22. He F, Pasam R, Shi F, Kant S, Keeble-Gagnere G, Kay P, Forrest K, Fritz A, Hucl P, Wiebe K, et al. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat Genet*. 2019;51:896–906.
23. Hao CY, Jiao CZ, Hou J, Li T, Liu HX, Wang YQ, Zheng J, Liu H, Bi ZH, Xu FF, et al. Resequencing of 145 landmark cultivars reveals asymmetric sub-genome selection and strong founder genotype effects on wheat breeding in China. *Mol Plant*. 2020;13:1733–51.
24. Guo WL, Xin MM, Wang ZH, Yao YY, Hu ZR, Song WJ, Yu KH, Chen YM, Wang XB, Guan PF, et al. Origin and adaptation to high altitude of Tibetan semi-wild wheat. *Nat Commun*. 2020;11:5085–96.
25. Cheng H, Liu J, Wen J, Nie XJ, Xu LH, Chen NB, Li ZX, Wang QL, Zheng ZQ, Li M, et al. Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. *Genome Biol*. 2019;20:136–42.
26. Zhao X, Guo Y, Kang L, Yin C, Bi A, Xu D, Zhang Z, Zhang J, Yang X, Xu J, et al. Population genomics unravels the Holocene history of bread wheat and its relatives. *Nat Plants*. 2023;9:403–19.
27. Bi A, XD, Kang L, Guo Y, Song X, Zhao X, Zhang J, Zhang Z, Li Y, Yin C, et al. An integrated map of genetic variation from 1,062 wheat genomes. *bioRxiv*. Preprint. <https://doi.org/10.1101/2023.03.31.535022>.
28. Liu Y, Shen K, Yin C, Xu X, Yu X, Ye B, Sun Z, Dong J, Bi A, Zhao X, et al. Genetic basis of geographical differentiation and breeding selection for wheat plant architecture traits. *Genome Biol*. 2023;24:114–39.
29. Cao SH, Xu DG, Hanif M, Xia XC, He ZH. Genetic architecture underpinning yield component traits in wheat. *Theor Appl Genet*. 2020;133:1811–23.
30. Wang DZ, Yu K, Jin D, Sun LH, Chu JF, Wu WY, Xin PY, Gregova E, Li X, Sun JZ, et al. Natural variations in the promoter of *Awn Length Inhibitor 1 (ALI-1)* are associated with awn elongation and grain length in common wheat. *Plant J*. 2020;101:1075–90.
31. DeWitt N, Guedira M, Lauer E, Sarinelli M, Tyagi P, Fu DL, Hao QQ, Murphy JP, Marshall D, Akhunova A, et al. Sequence-based mapping identifies a candidate transcription repressor underlying awn suppression at the *B1* locus in wheat. *New Phytol*. 2020;225:326–39.
32. Huang DQ, Zheng Q, Melchikart T, Bekkaoui Y, Konkin DJF, Kagale S, Martucci M, You FM, Clarke M, Adamski NM, et al. Dominant inhibition of awn development by a putative zinc-finger transcriptional repressor expressed at the *B1* locus in wheat. *New Phytol*. 2020;225:340–55.
33. Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, et al. Regulation of *OsSPL14* by *OsmiR156* defines ideal plant architecture in rice. *Nat Genet*. 2010;42:541–4.
34. Duan EC, Wang YH, Li XH, Lin QB, Zhang T, Wang YP, Zhou CL, Zhang H, Jiang L, Wang JL, et al. *OsSH11* regulates plant architecture through modulating the transcriptional activity of *IPA1* in rice. *Plant Cell*. 2019;31:1026–42.
35. Miura K, Ikeda M, Matsubara A, Song XJ, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M. *OsSPL14* promotes panicle branching and higher grain productivity in rice. *Nat Genet*. 2010;42:545–50.
36. Song XG, Meng XB, Guo HY, Cheng Q, Jing YH, Chen MJ, Liu GF, Bing Wang B, Wang YH, Li JY, Yu H. Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size. *Nat Biotechnol*. 2022;40:1403–13.
37. Wang J, Zhou L, Shi H, Chern M, Yu H, Yi H, He M, Yin JJ, Zhu XB, Li Y, et al. A single transcription factor promotes both yield and immunity in rice. *Science*. 2018;361:1026–8.
38. Wang BB, Wang HY. *IPA1*: A new "green revolution" gene? *Mol Plant*. 2017;10:779–81.
39. Liu MM, Shi ZY, Zhang XH, Wang MX, Zhang L, Zheng KZ, Liu JY, Hu XM, Di CR, Qian Q, et al. Inducible overexpression of ideal plant architecture1 improves both yield and disease resistance in rice. *Nature Plants*. 2019;5:389–400.
40. Thiel J, Koppolu R, Trautewig C, Hertig C, Kale SM, Erbe S, Mascher M, Himmelbach A, Rutten T, Esteban E, et al. Transcriptional landscapes of floral meristems in barley. *Sci Adv*. 2021;7:eabf0832.
41. Zhang B, Liu X, Zhao GY, Mao XG, Li A, Jing RL. Molecular characterization and expression analysis of *Triticum aestivum* squamosa-promoter binding protein-box genes involved in ear development. *J Integr Plant Biol*. 2014;56:571–81.
42. Zhu T, Liu Y, Ma LT, Wang XY, Zhang DZ, Han YC, Ding Q, Ma LJ. Genome-wide identification, phylogeny and expression analysis of the *SPL* gene family in wheat. *BMC Plant Biol*. 2020;20:420–34.
43. Pang YL, Liu CX, Wang DF, St Amand P, Bernardo A, Li WH, He F, Li LZ, Wang LM, Yuan XF, et al. High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. *Mol Plant*. 2020;13:1311–27.
44. Yu SZ, Wu JH, Wang M, Shi WM, Xia GM, Jia JZ, Kang ZS, Han DJ. Haplotype variations in QTL for salt tolerance in Chinese wheat accessions identified by marker-based and pedigree-based kinship analyses. *Crop Journal*. 2020;8:1011–24.

45. Franco MF, Polacco AN, Campos PE, Pontaroli AC, Vanzetti LS. Genome-wide association study for resistance in bread wheat (*Triticum aestivum* L.) to stripe rust (*Puccinia striiformis* f. sp. *tritici*) races in Argentina. *BMC Plant Biol.* 2022;22:543–60.
46. Bonnett OT. The development of the wheat spike. *J Agric Res.* 1936;53:445–51.
47. Cao P, Fan WJ, Li PJ, Hu YX. Genome-wide profiling of long noncoding RNAs involved in wheat spike development. *BMC Genomics.* 2021;22:543–60.
48. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8.
49. Wang JW, Czech B, Weigel D. miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell.* 2009;138:738–49.
50. Wang H, Wang HY. The miR156/SPL module, a regulatory hub and versatile toolbox, gears up crops for enhanced agronomic traits. *Mol Plant.* 2015;8:677–88.
51. Bai QW, Li CX, Wu L, Liu H, Ren HY, Li GS, Wang QL, Wu G, Zheng BW. Engineering chimeras by fusing plant receptor-like kinase EMS1 and BRI1 reveals the two receptors' structural specificity and molecular mechanisms. *Int J Mol Sci.* 2022;23:2155–73.
52. Yin YH, Vafeadas D, Tao Y, Yoshida S, Asami T, Chory J. A new class of transcription factors mediates brassinosteroid-regulated gene expression in *Arabidopsis*. *Cell.* 2005;120:249–59.
53. Dixon LE, Greenwood JR, Bencivenga S, Zhang P, Cockram J, Mellers G, Ramm K, Cavanagh C, Swain SM, Boden SA. *TEOSINTE BRANCHED1* regulates inflorescence architecture and development in bread wheat (*Triticum aestivum*). *Plant Cell.* 2018;30:563–81.
54. Dixon LE, Pasquariello M, Badgami R, Levin KA, Poschet G, Ng PQ, Orford S, Chayut N, Adamski NM, Brinton J, et al. MicroRNA-resistant alleles of *HOMEBOX DOMAIN-2* modify inflorescence branching and increase grain protein content of wheat. *Sci Adv.* 2022;8:eabn5907.
55. Boden SA, Cavanagh C, Cullis BR, Ramm K, Greenwood J, Finnegan EJ, Trevaskis B, Swain SM. *Ppd-1* is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nat Plants.* 2015;1:1–6.
56. Wang YG, Du F, Wang J, Wang K, Tian C, Qi XQ, Lu F, Liu XG, Ye XG, Jiao YL. Improving bread wheat yield through modulating an unselected AP2/ERF gene. *Nat Plants.* 2022;8:930–9.
57. Li Y, Li L, Zhao M, Guo L, Guo X, Zhao D, Batool A, Dong B, Xu H, Cui S, et al. Wheat *FRIZZY PANICLE* activates *VERNALIZATION1-A* and *HOMEBOX4-A* to regulate spike development in wheat. *Plant Biotechnol J.* 2021;19:1141–54.
58. Shaw LM, Lyu B, Turner R, Li CX, Chen FJ, Han XL, Fu DL, Dubcovsky J. *FLOWERING LOCUS T2* regulates spike development and fertility in temperate cereals. *J Exp Bot.* 2019;70:193–204.
59. Finnegan EJ, Ford B, Wallace X, Pettolino F, Griffin PT, Schmitz RJ, Zhang P, Barrero JM, Hayden MJ, Boden SA, et al. Zebularine treatment is associated with deletion of *FT-B1* leading to an increase in spikelet number in bread wheat. *Plant, Cell Environ.* 2018;41:1346–60.
60. Chen ZY, Ke WS, He F, Chai LL, Cheng XJ, Xu HW, Wang XB, Du DJ, Zhao YD, Chen XY, et al. A single nucleotide deletion in the third exon of *FT-D1* increases the spikelet number and delays heading date in wheat (*Triticum aestivum* L.). *Plant Biotechnol J.* 2022;20:920–33.
61. Kuzay S, Lin H, Chen S, Woods DP, Zhang J, Lan T, von Korff M, Dubcovsky J. *WAPO-A1* is the causal gene of the 7AL QTL for spikelet number per spike in wheat. *Plos Genet.* 2022;18:1–25.
62. Gauley A, Boden SA. Stepwise increases in *FT1* expression regulate seasonal progression of flowering in wheat (*Triticum aestivum*). *New Phytol.* 2021;229:1163–76.
63. Sakuma S, Golan G, Guo ZF, Ogawa T, Tagiri A, Sugimoto K, Bernhardt N, Brassac J, Mascher M, Hensel G, et al. Unleashing floret fertility in wheat through the mutation of a homeobox gene. *Proc Natl Acad Sci USA.* 2019;116:5182–7.
64. Molerio G, Joynson R, Pinera-Chavez FJ, Gardiner LJ, Rivera-Amado C, Hall A, Reynolds MP. Elucidating the genetic basis of biomass accumulation and radiation use efficiency in spring wheat and its role in yield potential. *Plant Biotechnol J.* 2019;17:1276–88.
65. Xie Q, Sparkes DL. Dissecting the trade-off of grain number and size in wheat. *Planta.* 2021;254:3–19.
66. Birkenbihl RP, Jach G, Saedler H, Huijser P. Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. *J Mol Biol.* 2005;352:585–96.
67. Si LZ, Chen JY, Huang XH, Gong H, Luo JH, Hou QQ, Zhou TY, Lu TT, Zhu JJ, Shangguan YY, et al. *OsSPL13* controls grain size in cultivated rice. *Nat Genet.* 2016;48:447–56.
68. Wang SK, Wu K, Yuan QB, Liu XY, Liu ZB, Lin XY, Zeng RZ, Zhu HT, Dong GJ, Qian Q, et al. Control of grain size, shape and quality by *OsSPL16* in rice. *Nat Genet.* 2012;44:950–4.
69. Wang SK, Li S, Liu Q, Wu K, Zhang JQ, Wang SS, Wang Y, Chen XB, Zhang Y, Gao CX, et al. The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat Genet.* 2015;47:949–54.
70. Zhang B, Xu WN, Liu X, Mao XG, Li A, Wang JY, Chang XP, Zhang XY, Jing RL. Functional conservation and divergence among homoeologs of *TaSPL20* and *TaSPL21*, two SBP-Box genes governing yield-related traits in hexaploid wheat. *Plant Physiol.* 2017;174:1177–91.
71. Cao RF, Guo LJ, Ma M, Zhang WJ, Liu XL, Zhao HX. Identification and functional characterization of squamosa promoter binding protein-like gene *TaSPL16* in wheat (*Triticum aestivum* L.). *Front Plant Sci.* 2019;10:1–4.
72. Yadav SR, Prasad K, Vijayraghavan U. Divergent regulatory *OsMADS2* functions control size, shape and differentiation of the highly derived rice floret second-whorl organ. *Genetics.* 2007;176:283–94.
73. Waddington SR, Cartwright PM, Wall PC. A quantitative scale of spike initial and pistil development in barley and wheat. *Ann Bot.* 1983;51:119–30.
74. Skibbe DS, Wang XJ, Borsuk LA, Ashlock DA, Nettleton D, Schnable PS. Floret-specific differences in gene expression and support for the hypothesis that tapetal degeneration of *Zea mays* L. occurs via programmed cell death. *J Genet Genomics.* 2008;35:603–16.
75. Li L, Shi F, Wang YQ, Yu XF, Zhi JJ, Guan YB, Zhao HY, Chang JL, Chen MJ, Yang GX, et al. *TaSPL13* regulates inflorescence architecture and development in transgenic wheat (*Triticum aestivum* L.). *Plant Sci.* 2020;296:115016.

76. Sajjad M, Ma XL, Khan SH, Shoaib M, Song YH, Yang WL, Zhang AM, Liu DC. *TaFlo2-A1*, an ortholog of rice *Flo2*, is associated with thousand grain weight in bread wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 2017;17:164–75.
77. Qin L, Hao CY, Hou J, Wang YQ, Li T, Wang LF, Ma ZQ, Zhang XY. Homologous haplotypes, expression, genetic effects and geographic distribution of the wheat yield gene *TaGW2*. *BMC Plant Biol.* 2014;14:107–26.
78. Abbate PE, Andrade FH, Lazaro L, Bariffi JH, Berardocco HG, Inza VH, Marturano F. Grain yield increase in recent argentine wheat cultivars. *Crop Sci.* 1998;38:1203–9.
79. Slafer GA, Elia M, Savin R, Garcia GA, Terrile II, Ferrante A, Miralles DJ, Gonzalez FG. Fruiting efficiency: an alternative trait to further rise wheat yield. *Food and Energy Security.* 2015;4:92–109.
80. Pretini N, Alonso MP, Vanzetti LS, Pontaroli AC, Gonzalez FG. The physiology and genetics behind fruiting efficiency: a promising spike trait to improve wheat yield potential. *J Exp Bot.* 2021;72:3987–4004.
81. Motzo R, Giunta F. Awnedness affects grain yield and kernel weight in near-isogenic lines of *durum* wheat. *Aust J Agric Res.* 2002;53:1285–93.
82. Rebetzke GJ, Bonnett DG, Reynolds MP. Awns reduce grain number to increase grain size and harvestable yield in irrigated and rainfed spring wheat. *J Exp Bot.* 2016;67:2573–86.
83. Whingwiri EE, Kuo J, Stern WR. The vascular system in the rachis of a wheat ear. *Ann Bot.* 1981;48:189–201.
84. Shi S, Yuan N, Yang M, Du ZL, Wang JY, Sheng X, Wu JY, Xiao JF. Comprehensive assessment of genotype imputation performance. *Hum Hered.* 2017;83:107–16.
85. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet.* 2018;103:338–48.
86. Browning BL, Tian XW, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet.* 2021;108:1880–90.
87. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
88. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics.* 2019;35:1786–8.
89. Bates D, Machler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015;67:1–48.
90. Bradley D, Carpenter R, Sommer H, Hartley N, Coen E. Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of *Antirrhinum*. *Cell.* 1993;72:85–95.
91. Liu H, Wang K, Jia Z, Gong Q, Lin Z, Du L, Pei X, Ye X. Efficient induction of haploid plants in wheat by editing of *TaMTL* using an optimized *Agrobacterium*-mediated CRISPR system. *J Exp Bot.* 2020;71:1337–49.
92. Wang K, Liu H, Du L, Ye X. Generation of marker-free transgenic hexaploid wheat via an *Agrobacterium*-mediated co-transformation strategy in commercial Chinese wheat varieties. *Plant Biotechnol J.* 2017;15:614–23.
93. Assanga SO, Fuentealba M, Zhang G, Tan C, Dhakal S, Rudd JC, Ibrahim AMH, Xue Q, Haley S, Chen J, et al. Mapping of quantitative trait loci for grain yield and its components in a US popular winter wheat TAM 111 using 90K SNPs. *PLoS One.* 2017;12:e0189669.
94. Beales J, Turner A, GriYths S, Snape JW, Laurie DA. A pseudo-response regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). *Theor Appl Genet.* 2007;115:721–33.
95. Bi XJ, van Esse W, Mulki MA, Kirschner G, Zhong JS, Simon R, von Korff M. *CENTRORADIALIS* interacts with *FLOWERING LOCUS T*-like genes to control floret development and grain number. *Plant Physiol.* 2019;180:1013–30.
96. Chai L, Xin M, Dong C, Chen Z, Zhai H, Zhuang J, Cheng X, Wang N, Geng J, Wang X, et al. A natural variation in Ribonuclease H-like gene underlies *Rht8* to confer “Green Revolution” trait in wheat. *Mol Plant.* 2022;15:377–80.
97. Chen Y, Yan Y, Wu TT, Zhang GL, Yin H, Chen W, Wang S, Chang F, Gou JY. Cloning of wheat keto-acyl thiolase 2B reveals a role of jasmonic acid in grain weight determination. *Nat Commun.* 2020;11:6266–77.
98. Cuthbert JL, Somers DJ, Brule-Babel AL, Brown PD, Crow GH. Molecular mapping of quantitative trait loci for yield and yield components in spring wheat (*Triticum aestivum* L.). *Theor Appl Genet.* 2008;117:595–608.
99. van Esse GW, Walla A, Finke A, Koornneef M, Pecinka A, von Korff M. Six-Rowed Spike3 (*VRS3*) is a histone demethylase that controls lateral spikelet development in barley. *Plant Physiol.* 2017;174:2397–408.
100. Guan P, Lu L, Jia L, Kabir MR, Zhang J, Lan T, Zhao Y, Xin M, Hu Z, Yao Y, et al. Global QTL analysis identifies genomic regions on chromosomes 4A and 4B harboring stable QTLs for yield-related traits across different environments in wheat (*Triticum aestivum* L.). *Front Plant Sci.* 2018;9:529–47.
101. Hai L, Guo HJ, Wagner C, Xiao SH, Friedt W. Genomic regions for yield and yield parameters in Chinese winter wheat (*Triticum aestivum* L.) genotypes tested under varying environments correspond to QTL in widely different wheat materials. *Plant Sci.* 2008;175:226–32.
102. Halliwell J, Borrill P, Gordon A, Kowalczyk R, Pagano ML, Saccomanno B, Bentley AR, Uauy C, Cockram J. Systematic investigation of *FLOWERING LOCUS T*-like poaceae gene families identifies the short-day expressed flowering pathway gene, *TaFT3* in wheat (*Triticum aestivum* L.). *Front Plant Sci.* 2016;7:1–15.
103. Huang XQ, Cloutier S, Lycar L, Radovanovic N, Humphreys DG, Noll JS, Somers DJ, Brown PD. Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). *Theor Appl Genet.* 2006;113:753–66.
104. Jia HY, Wan HS, Yang SH, Zhang ZZ, Kong ZX, Xue SL, Zhang LX, Ma ZQ. Genetic dissection of yield-related traits in a recombinant inbred line population created using a key breeding parent in China’s wheat breeding. *Theor Appl Genet.* 2013;126:2123–39.
105. Jia ML, Li YA, Wang ZY, Tao S, Sun GL, Kong XC, Wang K, Ye XG, Liu SS, Geng SF, et al. *TaAA21* represses *TaARF25*-mediated expression of *TaERFs* required for grain size and weight development in wheat. *Plant J.* 2021;108:1754–67.
106. Kong XC, Wang F, Geng SF, Guan JT, Tao S, Jia ML, Sun GL, Wang ZY, Wang K, Ye XG, et al. The wheat *AGL6*-like MADS-box gene is a master regulator for floral organ identity and a target for spikelet meristem development manipulation. *Plant Biotechnol J.* 2022;20:75–88.
107. Kumar A, Mantovani EE, Seetan R, Soltani A, Echeverry-Solarte M, Jain S, Simsek S, Doehlert D, Alamri MS, Elias EM, et al. Dissection of genetic factors underlying wheat kernel shape and size in an elite x nonadapted cross using a high density SNP linkage map. *Plant Genome.* 2016;9:1–22.

108. Kuzay S, Xu YF, Zhang JL, Katz A, Pearce S, Su ZQ, Fraser M, Anderson JA, Brown-Guedira G, DeWitt N, et al. Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. *Theor Appl Genet.* 2019;132:2689–705.
109. Li JH, Xie LN, Tian XL, Liu SY, Xu DA, Jin H, Song J, Dong Y, Zhao DH, Li GY, et al. *TaNAC100* acts as an integrator of seed protein and starch synthesis exerting pleiotropic effects on agronomic traits in wheat. *Plant J.* 2021;108:829–40.
110. Li FJ, Wen WE, He ZH, Liu JD, Jin H, Cao SH, Geng HW, Yan J, Zhang PZ, Wan YX, Xia XC. Genome-wide linkage mapping of yield-related traits in three Chinese bread wheat populations using high-density SNP markers. *Theor Appl Genet.* 2018;131:1903–24.
111. Liu G, Jia LJ, Lu LH, Qin DD, Zhang JP, Guan PF, Ni ZF, Yao YY, Sun QX, Peng HR. Mapping QTLs of yield-related traits using RIL population derived from common wheat and Tibetan semi-wild wheat. *Theor Appl Genet.* 2014;127:2415–32.
112. Lozada DN, Mason RE, Sukumaran S, Dreisigacker S. Validation of grain yield QTLs from soft winter wheat using a CIMMYT spring wheat panel. *Crop Sci.* 2018;58:1964–71.
113. Ma FF, Xu YF, Ma ZQ, Li LH, An DG. Genome-wide association and validation of key loci for yield-related traits in wheat founder parent Xiaoyan 6. *Mol Breeding.* 2018;38:91–106.
114. Mizuno N, Nitta M, Sato K, Nasuda S. A wheat homologue of *PHYTOCLOCK 1* is a candidate gene conferring the early heading phenotype to einkorn wheat. *Genes Genet Syst.* 2012;87:357–67.
115. Mohler V, Albrecht T, Castell A, Diethelm M, Schweizer G, Hartl L. Considering causal genes in the genetic dissection of kernel traits in common wheat. *J Appl Genet.* 2016;57:467–76.
116. Moore JW, Herrera-Foessel S, Lan C, Schnippenkoetter W, Ayliffe M, Huerta-Espino J, Lillemo M, Viccars L, Milne R, Periyannan S, et al. A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. *Nat Genet.* 2015;47:1494–8.
117. Feng N, Song GY, Guan JT, Chen K, Jia ML, Huang DH, Wu JJ, Zhang LC, Kong XY, Geng SF, et al. Transcriptome profiling of wheat inflorescence development from spikelet initiation to floral patterning identified stage-specific regulatory genes. *Plant Physiol.* 2017;174:1779–94.
118. Pallotta M, Schnurbusch T, Hayes J, Hay A, Baumann U, Paull J, Langridge P, Sutton T. Molecular basis of adaptation to high soil boron in wheat landraces and elite cultivars. *Nature.* 2014;514:88–101.
119. Schilling S, Kennedy A, Pan S, Jermin LS, Melzer R. Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol.* 2020;225:511–29.
120. Sukumaran S, Lopes M, Dreisigacker S, Reynolds M. Genetic analysis of multi-environmental spring wheat trials identifies genomic regions for locus-specific trade-offs for grain weight and grain number. *Theor Appl Genet.* 2018;131:985–99.
121. Sun XY, Wu K, Zhao Y, Kong FM, Han GZ, Jiang HM, Huang XJ, Li RJ, Wang HG, Li SS. QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. *Euphytica.* 2009;165:615–24.
122. Wang JS, Liu WH, Wang H, Li LH, Wu J, Yang XM, Li XQ, Gao AN. QTL mapping of yield-related traits in the wheat germplasm 3228. *Euphytica.* 2011;177:277–92.
123. Wang JY, Li L, Li CN, Yang X, Xue YH, Zhu Z, Mao XG, Jing RL. A transposon in the vacuolar sorting receptor gene *TaVSR1-B* promoter region is associated with wheat root depth at booting stage. *Plant Biotechnol J.* 2021;19:1456–67.
124. Wang LF, Ge HM, Hao CY, Dong YS, Zhang XY. Identifying loci influencing 1,000-Kernel weight in wheat by microsatellite screening for evidence of selection during breeding. *PLoS One.* 2012;7:e29432.
125. Wang SX, Zhu YL, Zhang DX, Shao H, Liu P, Hu JB, Zhang H, Zhang HP, Chang C, Lu J, et al. Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using SNP markers. *PLoS One.* 2017;12:e0188662.
126. Wu XY, Cheng RR, Xue SL, Kong ZX, Wan HS, Li GQ, Huang YL, Jia HY, Jia JZ, Zhang LX, Ma ZQ. Precise mapping of a quantitative trait locus interval for spike length and grain weight in bread wheat (*Triticum aestivum* L.). *Mol Breed.* 2014;33:129–38.
127. Wurschum T, Leiser WL, Langer SM, Tucker MR, Longin CFH. Phenotypic and genetic analysis of spike and kernel characteristics in wheat reveals long-term genetic trends of grain yield components. *Theor Appl Genet.* 2018;131:2071–84.
128. Xiong H, Zhou C, Fu M, Guo H, Xie Y, Zhao L, Gu J, Zhao S, Ding Y, Li Y, et al. Cloning and functional characterization of *Rht8*, a “green revolution” replacement gene in wheat. *Mol Plant.* 2022;15:373–6.
129. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene *VRN1*. *Proc Natl Acad Sci USA.* 2003;100:6263–8.
130. Yan LL, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science.* 2004;303:1640–4.
131. Yang X, Zhong SB, Zhang QJ, Ren Y, Sun CW, Chen F. A loss-of-function of the dirigent gene *TaDIR-B1* improves resistance to Fusarium crown rot in wheat. *Plant Biotechnol J.* 2021;19:866–8.
132. Zhang JL, Gizaw SA, Bossolini E, Hegarty J, Howell T, Carter AH, Akhunov E, Dubcovsky J. Identification and validation of QTL for grain yield and plant water status under contrasting water treatments in fall-sown spring wheats. *Theor Appl Genet.* 2018;131:1741–59.
133. Zhang JH, Zhang HT, Li SY, Li JY, Yan L, Xia LQ. Increasing yield potential through manipulating of an *ARE1* ortholog related to nitrogen use efficiency in wheat by CRISPR/Cas9. *J Integr Plant Biol.* 2021;63:1649–63.
134. Zikhali M, Wingen LU, Leverington-Waite M, Specl S, Griffiths S. The identification of new candidate genes *Triticum aestivum* *FLOWERING LOCUS T3-B1* (*TaFT3-B1*) and *TARGET OF EAT1* (*TaTOE1-B1*) controlling the short-day photoperiod response in bread wheat. *Plant Cell Environ.* 2017;40:2678–90.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.