Check for updates

Identification of molecular markers and candidate regions associated with grain number per spike in Pubing3228 using SLAF-BSA

Jiansheng Wang[1,2]*, Erwei Wang[3], Shiping Cheng[1,2]
and Aichu Ma[3]

[1]College of Chemistry and Environment Engineering, Pingdingshan University, Pingdingshan,
Henan, China, [2]Henan Key Laboratory of Germplasm Innovation and Utilization of Eco-economic
Woody Plant, Pingdingshan, Henan, China, [3]Pingdingshan Academy of Agricultural Science,
Pingdingshan, Henan, China

Grain number per spike, a pivotal agronomic trait dictating wheat yield, lacks a comprehensive understanding of its underlying mechanism in Pubing3228, despite the identification of certain pertinent genes. Thus, our investigation sought to ascertain molecular markers and candidate regions associated with grain number per spike through a high-density genetic mapping approach that amalgamates site-specific amplified fragment sequencing (SLAF-seq) and bulked segregation analysis (BSA). To facilitate this, we conducted a comparative analysis of two wheat germplasms, Pubing3228 and Jing4839, known to exhibit marked discrepancies in spike shape. By leveraging this methodology, we successfully procured 2,810,474 SLAF tags, subsequently resulting in the identification of 187,489 single nucleotide polymorphisms (SNPs) between the parental strains. We subsequently employed the SNP-index association algorithm alongside the extended distribution (ED) association algorithm to detect regions associated with the trait. The former algorithm identified 24 trait-associated regions, whereas the latter yielded 70. Remarkably, the intersection of these two algorithms led to the identification of 25 trait-associated regions. Amongst these regions, we identified 399 annotated genes, including three genes harboring non-synonymous mutant SNP loci. Notably, the APETALA2 (AP2) transcription factor families, which exhibited a strong correlation with spike type, were also annotated. Given these findings, it is plausible to hypothesize that these genes play a critical role in determining spike shape. In summation, our study contributes significant insights into the genetic foundation of grain number per spike. The molecular markers and candidate regions we have identified can be readily employed for marker-assisted breeding endeavors, ultimately leading to the development of novel wheat cultivars possessing enhanced yield potential. Furthermore, conducting further functional analyses on the identified genes will undoubtedly facilitate a comprehensive elucidation of the underlying mechanisms governing spike development in wheat.

1 Introduction

Wheat (*Triticum aestivum* L.), a globally cultivated staple crop, serves as a primary food source for approximately 35% of the world's population. Given its status as the most abundant and traded food commodity, its economic significance cannot be understated (Guo et al., 2018). The spike, an essential reproductive structure of wheat, assumes a critical role in grain production and maintenance (Faris et al., 2014). Previous investigations have extensively explored the association between spike morphology and seed yield, specifically focusing on spike densities (SD), spike length (SL), and spikelet number per spike (SNS) (Chai et al., 2019; Li T. et al., 2021; You et al., 2021; Yu et al., 2022). The shape of the wheat spike exhibits a direct correlation with yield and is thus considered a pivotal agronomic trait during the process of domestication and selective breeding of wheat. Notably, the characteristics of the spike, including grain number per spike, exert a direct influence on yield levels. Therefore, gaining a comprehensive understanding of the genetic regulation underlying grain number per spike in wheat holds immense theoretical significance and practical value (Cui et al., 2012).

A comprehensive understanding of the physiological, genetic, and developmental mechanisms governing spikelet morphology is of paramount importance, as it not only contributes to the augmentation of spikelet numbers but also facilitates the improvement of spikelet fruiting or setting rates (Slafer et al., 2015; Gámez et al., 2020). However, conventional breeding techniques employed to enhance wheat yields encounter significant challenges, given the considerable costs and labor-intensive nature associated with incorporating these traits into extensive germplasm collections (Cui et al., 2012). Consequently, there arises a pressing need to explore innovative methodologies that can effectively elevate wheat yield potential.

The utilization of molecular markers plays a pivotal role in molecularly assisted selection (MAS) for plant breeding, as it establishes a link between phenotypes and molecular markers. Over time, various types of molecular markers have been developed, encompassing the first-generation markers such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), and amplified fragment length polymorphism (AFLP), as well as the second-generation markers like simple sequence repeats (SSR) and inter-simple sequence repeats (ISSR). Nonetheless, these markers exhibit limitations in terms of throughput, accuracy, time consumption, labor intensiveness, and cost (Zheng et al., 2018; Clifton-brown et al., 2019; Chen et al., 2022). To surmount these challenges, a third generation of molecular markers known as SNPs has emerged. SNPs are DNA sequence variants that arise from a single nucleotide

Abbreviations: AFLP, Amplified fragment length polymorphism; BSA, Bulked Segregation Analysis; ED, Euclidean Distance; ISSR, Inter-simple sequence repeats; MAS, Molecularly assisted selection; RAPD, Random amplified polymorphic DNA; RIL, Recombinant inbred line; RFLP, Restriction fragment length polymorphism; SD, Spike densities; SL, Spike length; SLAF-seq, Site-specific amplified fragment sequencing; SNPs, Single nucleotide polymorphisms; SNS, Spikelet number per spike; SSR, Simple sequence repeats.

alteration in the genome sequence. They are typically distributed extensively across the genome and represent the most abundant form of genomic polymorphism, thereby making SNP markers the densest available marker type. The utilization of high-throughput DNA sequencing technologies enables the rapid and efficient identification of a vast number of SNPs in species, facilitating genotyping at high marker densities (Graner et al., 2013). Moreover, SNPs are often employed for population structure determination and linkage disequilibrium (LD) analysis (Jain et al., 2014). Concurrently, insertion/deletion (InDel) markers, which are also employed for fine mapping and marker-assisted selection in diverse crops (Liang et al., 2011; Wang et al., 2012; Das et al., 2015), can exert an influence on gene expression and function, given their location within coding and regulatory regions (Jain et al., 2014). Variants present within these regions can induce alterations in protein function and modulate gene expression, as exemplified by SNPs located on exons of the *GA2ox* gene, impacting spikelet sterility in wheat (Alqudah et al., 2020). Consequently, the discovery of SNPs assumes paramount significance for investigating genomic variation in crop species (Shen et al., 2017). In a genome-wide association study conducted on wheat, ten stable SNPs were identified, demonstrating associations with seeding emergence rate (SER) and tiller number at various fertility stages. Ultimately, a subset of these alleles exhibited efficacy in enhancing the seeding emergence rate and tiller number at diverse fertility stages (Chen et al., 2017).

SNPs have emerged as a highly effective tool for plant breeding through the implementation of MAS. However, the application of SNP-based array technology presents certain limitations in the detection of novel loci. To address this, the development of SLAF-seq markers has provided a promising solution, offering advantages such as deep sequencing, reduced sequencing costs, optimized marker efficiency, and suitability for large populations (Sun et al., 2013; Wang et al., 2022). In conjunction with BSA techniques, the SLAF-BSA method has been successfully employed for the identification of major QTLs associated with specific traits in various plant species. Unlike individual trait analysis, the BSA technique circumvents this requirement by focusing on individuals exhibiting contrasting extreme phenotypes. This integrated approach offers a cost-effective and accurate means to elucidate the genetic architecture underlying target traits in plant species lacking a reference genome (Michenlmore and Kesseli, 1991; Du et al., 2019). The efficacy of the SLAF-BSA method has been demonstrated across diverse plant species, including peas, oilseed rape, and pepper, where it has successfully identified QTL profiles and candidate genes associated with specific traits such as leaf shape, seed weight, first flowering node, and resistance to blast root rot. Thus, the SLAF-BSA method emerges as a potent tool for the discovery of genetic markers linked to desirable traits, ultimately facilitating plant breeding endeavors (Geng et al., 2016; Xu et al., 2016; Zhang et al., 2018; Zheng et al., 2018).

Pubing3228, a novel wheat germplasm derived from the hybridization of common wheat and *Agropyron*, exhibits numerous superior agronomic traits, particularly in relation to spike characteristics. However, the molecular basis governing the spike traits of Pubing3228 remains largely unexplored. Therefore,

the primary objective of this investigation was to establish a high-resolution genetic map utilizing the SLAF-BSA approach in two wheat germplasms, namely Pubing3228 and Jing4839, in addition to F_2 population generated from two exceptional wheat cultivars. By analyzing the extreme spike types observed in the two wheat germplasms, we aimed to identify stable molecular markers and candidate regions with the potential to contribute to future endeavors in fine mapping and gene cloning, specifically targeting key genes associated with enhanced wheat yield. This study represents a significant advancement in our comprehension of the genetic mechanisms underlying wheat yield traits, thereby paving the way for the development of more efficient strategies for wheat breeding and cultivation.

2 Materials and methods

2.1 Plant material and phenotypic data collection

Pubing3228, a genetically stable wheat derivative resulting from the cross between wheat and *A. cristatum*, exhibits desirable spike traits such as a long spike, high number of spikelets per spike, and high grain number per spike. This particular wheat germplasm was developed by the research team led by Lihui Li at the Chinese Academy of Agricultural Science. In order to explore the phenotypic differences in spike traits, two wheat germplasms, Pubing3228 and Jing4839, were carefully selected for a cross. The resulting second-generation offspring served as the sequencing material for the present study. In a field environment, the materials were planted in randomized complete blocks with three replications. Each block consisted of three rows, with each row measuring 2 meters in length and maintaining a 30 cm distance between rows. Standard local practices were followed for all field management activities. At the stage of physiological maturity, 10 representative plants per genotype from each replication were harvested and manually threshed. The grain number per spike (GNS) was subsequently recorded. The distinctive characteristics of the two wheat spike types are visually presented in Figure 1.

2.2 SLAF library construction and high-throughput sequencing

2.2.1 Enzyme digestion protocol design

To select the most suitable enzyme sectioning scheme, the researchers utilized enzyme section prediction software to predict the enzyme sections of the reference genome. The selection principles were based on four criteria. Firstly, the proportion of enzyme sections located in repetitive sequences was kept as low as possible. Secondly, the enzyme sections were distributed as evenly as possible on the genome. Thirdly, the length of the enzyme sections matched the specific experimental system (Davey et al., 2013). Finally, the number of enzyme sections (SLAF tags) obtained was expected to meet the target number of tags.

2.2.2 Experimental procedure

The genomic DNA of each qualified sample was digested using the optimal digestion protocol that was selected through the enzyme section prediction software. The resulting enzymatic fragments, or SLAF tags, were subjected to several steps including 3′-end addition A treatment, ligation of Dual-index sequencing connectors, PCR amplification, purification, mixing, gum cutting to select the target fragments, and library quality control. Finally, the libraries were sequenced using Illumina HiSeqTM with PE125bp sequencing.

2.2.3 SLAF-BSA

For the BSA analysis, the spikes of both parental lines and selected spikes from the F_2 population were segregated into two distinct pools, namely pools aa and ab. The identification of candidate genes was carried out utilizing two algorithms: the SNP-index association algorithm and the ED association algorithm. The SNP-index method, which is a recently published approach for marker association analysis, relies on the detection of genotype frequency disparities between the pools (Fekih et al., 2013; Takagi et al., 2013). The primary aim of this method is to discern significant deviations in genotype frequencies within mixed pools by employing the ΔSNP-index statistic. The calculation procedure can be summarized as follows:

$$SNPindex(aa) = Maa/(Maa + Paa)$$

Maa indicates that the aa pool is derived from the depth of the female parent; Paa indicates that the aa pool is derived from the depth of the male parent.

$$SNPindex(ab) = Mab/(Mab + Pab)$$

Mab indicates that the ab pool is derived from the depth of the female parent; Pab indicates that the ab pool is derived from the depth of the male parent.

$$\Delta SNPindex = SNPindex(aa) - SNPindex(ab)$$

In this investigation, the BSA technique was employed to ascertain candidate genes associated with wheat yield. To facilitate this analysis, wheat spikes obtained from both parental lines and selected spikes from the F_2 population were carefully segregated into two distinct pools denoted as pools aa and ab. Subsequently, two algorithms, namely SNP-index and ED, were utilized for candidate gene selection. The SNP-index algorithm, a recent development in marker association analysis, exploits the disparities in genotype frequencies observed between the mixed pools, quantified by the ΔSNP-index statistic. Conversely, the ED algorithm employs sequencing data to identify markers that exhibit significant differences between the pools, thereby facilitating the assessment of the genomic region's association with the trait of interest. It is important to note that the ED value for non-target loci is expected to be 0, given that the BSA project constructs two hybrid pools that are primarily identical except for the variations present in the target trait-related loci. A higher ED value indicates a more pronounced distinction between the two hybrid pools, thereby
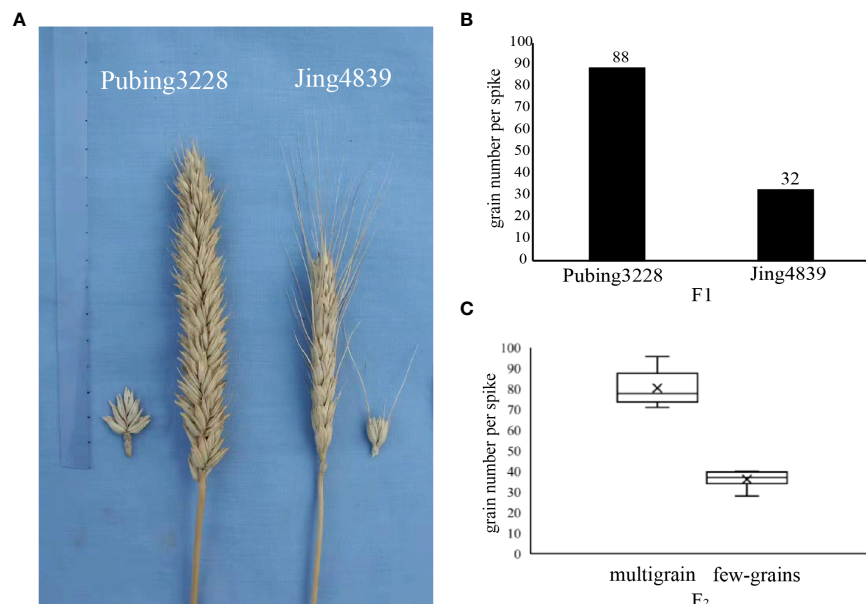
FIGURE 1
Wheat parental and hybrid second-generation spike counts. **(A)** Morphological characteristics of two varieties of paternal and maternal parent. **(B)** Average grain number of two paternal and maternal parent wheat varieties. **(C)** The number of two extreme characteristics after F_2 generation trait separation.

signifying a stronger association with the trait under investigation.

$$ED = \sqrt{(A_{mut} - A_{wt})^2 + (C_{mut} - C_{wt})^2 + (G_{mut} - G_{wt})^2 + (T_{mut} - T_{wt})^2}$$

Amut represents the frequency of A bases observed in the mutant mixing pool, while Awt signifies the frequency of A bases in the wild-type mixing pool. Similarly, Cmut denotes the frequency of C bases in the mutant mixing pool, whereas Cwt corresponds to the frequency of C bases in the wild-type mixing pool. Furthermore, Gmut indicates the frequency of G bases in the mutant mixing pool, while Gwt represents the frequency of G bases in the wild-type mixing pool. Lastly, Tmut denotes the frequency of T bases observed in the mutant mixing pool, and Twt signifies the frequency of T bases in the wild-type mixing pool.

3 Result

Significant distinctions were observed between Pubing3228 and Jing4839 in terms of spike length, morphology per spikelet, awning length, and number of grains per spike, as depicted in Figures 1A, B. Pubing3228 exhibited longer spikes, larger morphology per spikelet, and an absence of awning throughout the spike, whereas Jing4839 displayed smaller spikes, smaller morphology per spikelet, and longer awning throughout the spike. The average grain count per spike was determined to be 88 for Pubing3228 and 32 for Jing4839.

In order to identify potential molecular markers and candidate regions associated with the target trait, a mixed pool sample was created by crossing the two wheat varieties, and the resulting F_2 generation was classified into two distinct types based on grain number. The multi-spike grain type exhibited an average of 80

grains per spike, whereas the few-spike grain type had an average of 36 grains per spike, as illustrated in Figure 1C. The statistical information regarding the seed grain situation of the F_2 generation is presented in Supplementary Table 1. Subsequently, the second generation progeny was subjected to sequencing and subsequent analysis to construct the mixed pool sample.

3.1 Characterization of SLAF sequencing data and SNPs

In this investigation, the RsaI enzyme was employed to enzymatically fragment the reference genome, targeting specific sections with a length range of 464-484 base pairs. This enzymatic cleavage procedure yielded a total of 311,781 SLAF tags. Subsequently, the parental and mixed pool samples underwent the SLAF-seq sequencing protocol, resulting in the generation of 239.26 million reads after applying a quality filtering criterion of 80% Q30. Among these reads, 2,810,474 SLAF tags were successfully identified, exhibiting a sequencing depth of 11.67X and 8.32X for each parental sample, and 20.10X and 24.15X for the mixed pool samples, respectively. Notably, these SLAF tags were evenly distributed across the 21 chromosome pairs, as illustrated in Figure 2A. Furthermore, a total of 187,489 SNPs were detected between the parental samples, while 164,018 SNPs were identified within the mixed pool samples, as depicted in Figure 2B. A comprehensive summary of these findings is presented in Table 1.

The table displays the following information: ID (the unique identifier of the sample), SLAF number (the number of SLAF tags in the corresponding sample), Total depth (the total sequencing depth of the SLAF tags in the corresponding sample, i.e., the total number
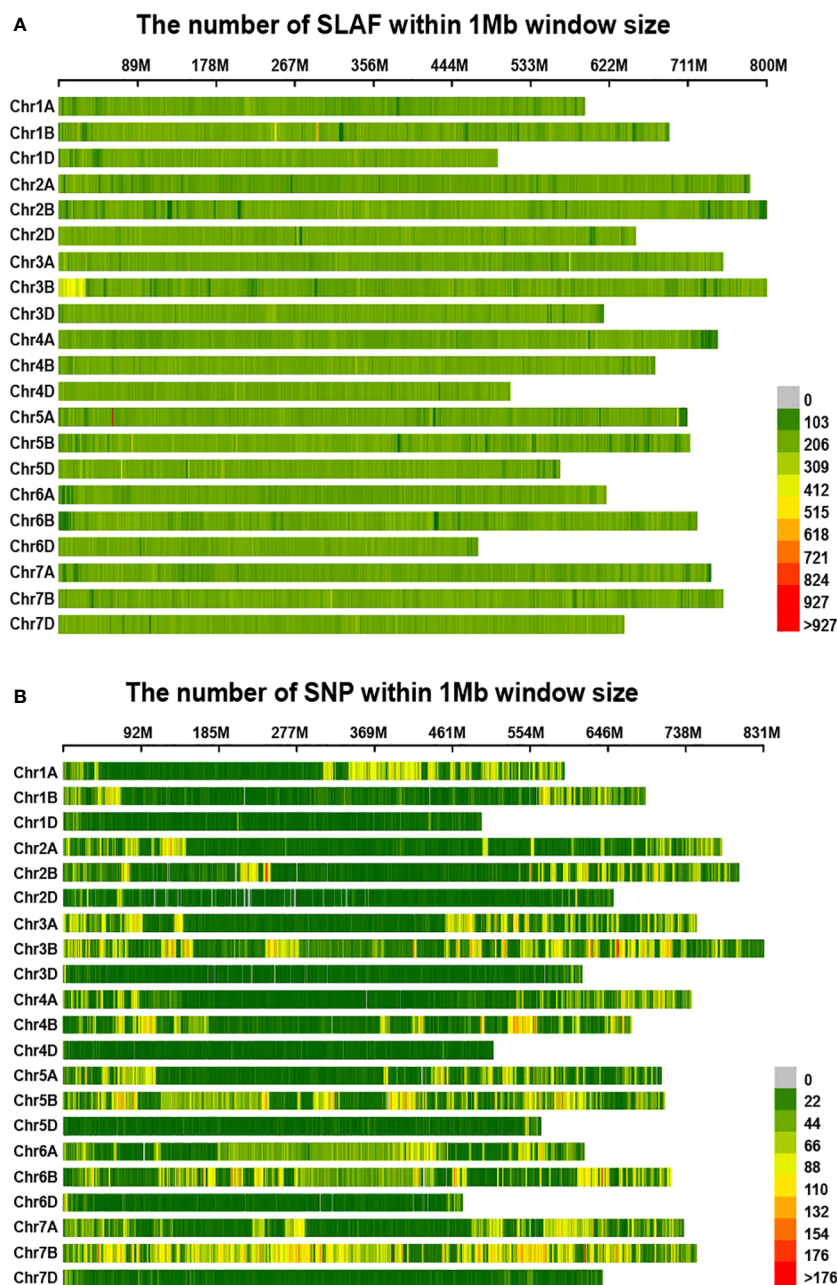
FIGURE 2
Distribution of SLAF-tags **(A)** and SNPs **(B)** on each reference chromosome in wheat. **(A, B)** The horizontal axis of the figure illustrates the chromosomal lengths, where each yellow band represents an individual chromosome. The genome is partitioned into 1 Mb segments. The shading intensity of each window corresponds to the number of labels contained within it, with darker windows indicating a higher label count, while lighter windows indicate a lower label count.

of reads developed from the SLAF tags), and Average depth (the average number of sequencing reads of the corresponding sample on each SLAF).

3.2 High-quality SNP screening

Prior to conducting the association analysis, a comprehensive screening of SNP loci was carried out. A total of 88,211 high-quality SNP loci were identified by applying filters to exclude loci with

multiple genotypes, loci with less than 4 read support, loci with consistent genotypes among mixed pools, and loci with recessive mixed pool genes not inherited from the recessive parent (Table 2).

3.3 Correlation analysis

The ED algorithm, a widely utilized approach, is employed to detect markers that exhibit significant differences between pools and to evaluate regions associated with specific traits (Hill et al., 2013). In

TABLE 1 SLAF tag statistics.

ID	SLAF number	Total depth	Average depth
M	1,463,703	17,074,359	11.665,2
P	1,519,806	12,647,895	8.322,0
aa	1,851,584	37,225,010	20.104,4
ab	1,936,123	46,749,748	24.146,1
Total	2,810,474	113,697,012	40.454,7

the context of the BSA project, two mixed pools are generated to identify disparities in loci related to the target trait, while minimizing variations in non-target loci. Consequently, the ED value for non-target loci is expected to converge towards 0. To mitigate background noise, the fourth power of the initial ED was utilized as the associated value, and the SNPNUM method was employed to fit the ED values. The association threshold for analysis was determined as 0.07, calculated as the median plus three standard deviations (SD) of the fitted values for all loci (Hill et al., 2013).

Seventy regions encompassing a total of 1,850 genes were identified, with these regions spanning a cumulative length of 241.70 Mb (Figure 3A). Among these genes, 10 harbored non-synonymous mutant SNP loci. Chromosome-wise distribution revealed that 647 genes were located on chromosome 3A, 913 genes on chromosome 5B, and 290 genes on chromosome 7B. For further analysis, detailed information on the screened genes is provided in Supplementary Table 2 (Figures 3B–D).

3.4 SNP-index method association results

3.4.1 SNP index analysis

The SNP index method, a recent development in marker association analysis based on genotype frequency differences between mixed pools, was utilized in this study (Fekih et al., 2013; Takagi et al., 2013). To mitigate false positive findings, the ΔSNP-index statistic was employed, and fitting of ΔSNP-index values for markers located on the same chromosome was conducted using their genomic positions. The SNPNUM method was employed to fit the ΔSNP-index values, and regions surpassing the association threshold were considered as trait-associated regions. Figure 4 depicts the distribution of SNP-index and ΔSNP-index for each of the two mixed pools. By employing the 99th percentile of the fitted ΔSNP-index (0.40), a total of 24 regions spanning 44.33 Mb and encompassing 400 genes were identified (Supplementary Table 3).

TABLE 2 SNP filtering statistics.

Total_SNP	MAL	LRS	GCLMP	UPFL	HSNP
361,030	643	167,992	12,596	91,588	88,211

MAL, Multiple allele loci; LRS, Loci with reading support less than 4; GCLMP, Genotypically congruent loci in mixed pools; UPFL, Use of parentally filtered loci; HSNP, High-quality SNPs.

3.5 Candidate area screening
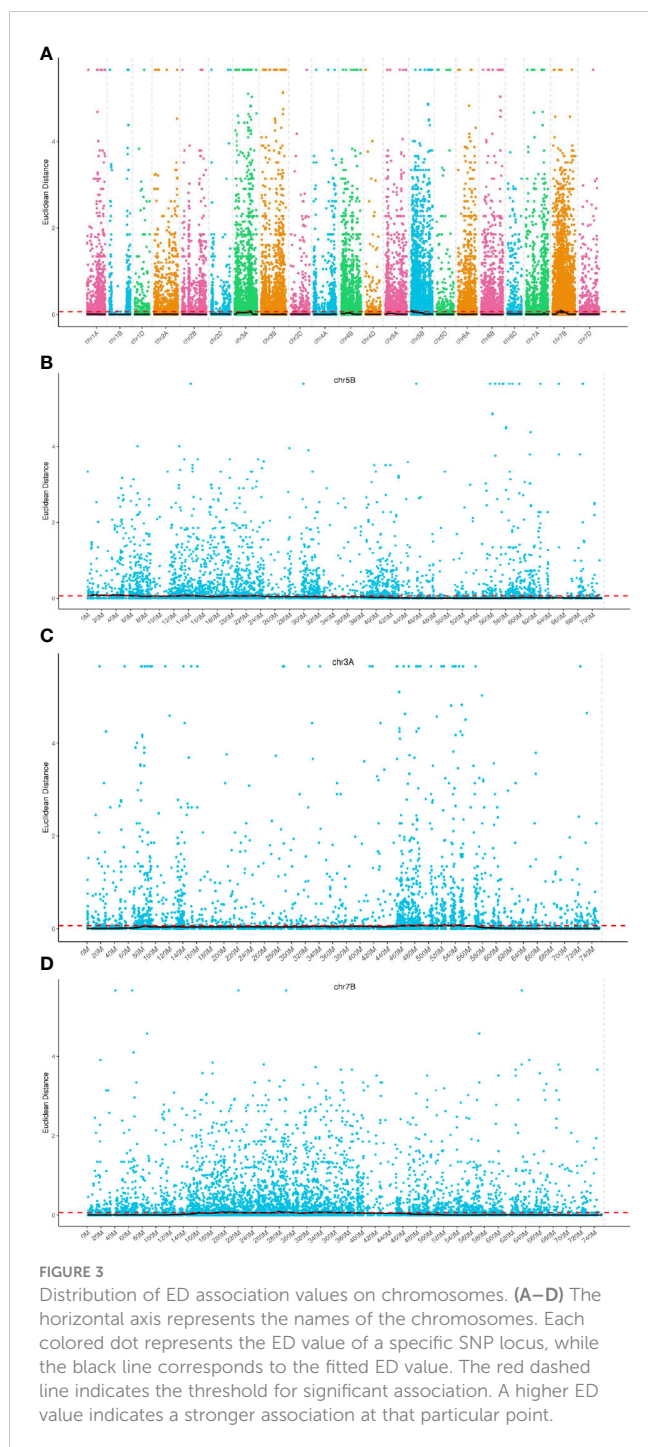
3.5.1 SNP annotation of candidate regions

A comparison was conducted between the association results obtained from the ED method and the SNP-index method, and the intersecting genes were subjected to annotation. Within this intersection, three SNPs were identified, exhibiting non-synonymous mutations between the parental lines, thus directly linking them to the target traits (Supplementary Table 4). These three genes were identified as *TraesCS3A01G260600*, *TraesCS3A01G261000*, and *TraesCS3A01G310500*, all of which exhibited up-regulation in wheat.

3.6 Functional notes on candidate areas

The candidate genes within the identified intervals underwent comprehensive annotation using various databases such as Gene Ontology (GO, https://geneontology.org/), Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg/), and others, facilitated by the BLAST software. This thorough annotation allowed for efficient screening of candidate genes, resulting in the annotation of 399 genes within the candidate region (Figure 5, Please refer to Supplementary Table 5 for additional details). Supplementary Table 5 presents the annotated genes that exhibited non-synonymous mutant SNPloci. Among them, three genes were annotated in theNon-Redundant Protein Database (NR, https://www.ncbi.nlm.nih.gov/refseq/about/nonredundantproteins/) andTranslation of EMBL (trEMBL, https://www.ebi.ac.uk/Tools/st/) databases, while no annotation was found in the KEGG database. *TraesCS3A01G310500* was annotated as a B3 domain-containing protein in the SwissProt database (https://ngdc.cncb.ac.cn/databasecommons/database/id/5614), and *TraesCS3A01G260600.1* was annotated as being involved in lipid transport and metabolism, according to the Clusters of Orthologous Groups of proteins (COG, https://www.ncbi.nlm.nih.gov/research/cog/) database.

GO annotation was performed for a total of 399 genes located within the candidate region. These genes were annotated based on their classification into cellular component (CC), biological process (BP), and molecular function (MF) categories. Among the CC annotations, the most frequently assigned positions included cell (188), cell part (188), organelle (177), intracellular organelle (177), cytoplasm (169), membrane (86), and others. In terms of BP, the commonly annotated processes were metabolic process (185), cellular process (175), and single-organism process (133). Regarding MF annotations, the predominant functional categories were binding (169), catalytic activity (153), and transferase activity (75) (Supplementary Table 6).

Furthermore, GO enrichment analysis was performed on the 10 genes with the smallest KS (Kolmogorov-Smirnov) values in CC, BP, and MF. These genes exhibited enrichment in specific GO terms. In CC, the enriched terms included intracellular organelles (177), nucleus (56), cytoplasmic membrane-bounded vesicles (32), interchromatin granules (3), perinuclear region of cytoplasm (2),

Distribution of ED association values on chromosomes. **(A–D)** The horizontal axis represents the names of the chromosomes. Each colored dot represents the ED value of a specific SNP locus, while the black line corresponds to the fitted ED value. The red dashed line indicates the threshold for significant association. A higher ED value indicates a stronger association at that particular point.

and others. In BP, the enriched terms comprised oxidation-reduction processes (14), negative regulation of DNA-templated transcription (4), recognition of pollen (3), arginyl-tRNAaminoacylation (2), and others. In MF, the enriched terms were primarily related to pyridoxal phosphate binding (8), diacylglycerol O-acyltransferase activity (3), arginine-tRNA ligase activity (2), and others. These genes were primarily associated with the activation of specific enzymatic activities, redox processes, and energy transfer. The nucleus and cytoplasm were identified as the main sites of action for these genes.

Genes in an organism collectively contribute to the execution of various biological functions, while pathways represent a coordinated sequence of interactions involving multiple genes. In order to gain insights into the functional characteristics of the 399 genes within the candidate regions, KEGG annotation was conducted (Supplementary Table 7). The analysis revealed that the majority of genes (191) were annotated within the categories of Environmental Information Processing, Genetic Information Processing, and Metabolism. Within the Environmental Information Processing category, 12 genes were assigned to the ABC transporters pathway. In the Genetic Information Processing category, 9 genes were associated with the Spliceosome pathway, and 5 genes were linked to the Ribosome pathway. In the Metabolism category, 9 genes were annotated in the Carbon metabolism pathway, 6 genes were involved in the Biosynthesis of amino acids pathway, and 4 genes were associated with each of the following pathways: Glycine, serine and threonine metabolism; Carbon fixation in photosynthetic organisms; and Glutathione metabolism.

4 Discussion

In wheat breeding programs, the identification of genes or QTL associated with yield-related traits is of great importance, as these traits often exhibit higher heritability compared to yield itself. This genetic information is crucial for understanding the underlying mechanisms of yield and facilitating the genetic improvement of wheat varieties (Li et al., 2022). Among the yield-related traits, spike-related traits hold promise for enhancing seed yield. However, conventional screening of these traits in large isolated populations can be laborious, costly, and inefficient. As a result, the adoption of molecular-assisted selection (MAS) or genetic selection has gained popularity as an efficient breeding approach. A comprehensive understanding of the genetic basis underlying yield traits, including wheat spike morphology, is essential for enhancing wheat yield. In this regard, the utilization of genetic resources such as wheat genome sequences and SNP platforms for wheat breeding plays a pivotal role in the development of improved wheat varieties (Malik et al., 2021).

SLAF-seq, a cost-effective technique, has proven to be highly valuable in generating a large number of polymorphic markers, thereby enabling the construction of high-density genetic maps for plant species with extensive genomes (Sun et al., 2013). Compared to traditional PCR-based methods, numerous studies have demonstrated that SLAF-seq significantly enhances the accuracy and efficiency of QTL mapping (Zheng et al., 2018). To improve mapping precision and narrow down candidate regions, BSA has been employed, which involves the selection of individuals with extreme phenotypes. BSA has been successfully utilized as a molecular marker in various organisms, and protocols have been developed specifically in model plants such as Arabidopsis and rice (Song et al., 2017). Classical BSA analysis reduces the cost of second-generation sequencing by pooling equimolar amounts of DNA from individuals exhibiting the same trait (Zhu et al., 2021).
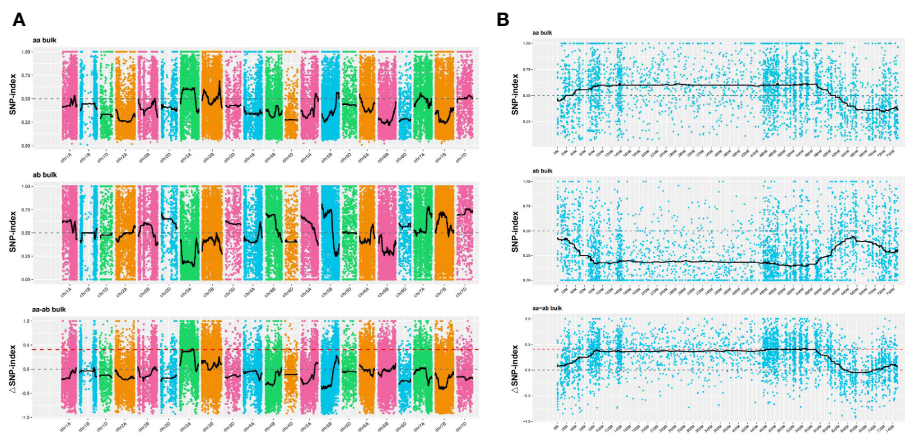
FIGURE 4

Distribution of SNP-index association values on chromosomes. **(A, B)** The x-axis corresponds to the chromosome names, while the y-axis represents the calculated ΔSNP-index values. The colored dots on the graph represent the calculated ΔSNP-index values, while the black lines indicate the fitted ΔSNP-index values. The top panel displays the distribution of SNP-index values specifically for the recessive mixed pool, whereas the middle panel depicts the distribution of SNP-index values for the dominant mixed pool. Lastly, the bottom panel illustrates the distribution of ΔSNP-index values, with the red line indicating the threshold line set at the 99th percentile.

In our study, we performed SLAF-seq and obtained a total of 2,810,474 SLAF tags, resulting in the identification of 187,489 SNPs between the parental lines and 164,018 SNPs between the mixed pools. Using the SNP-index association algorithm, we identified 24 trait-associated marker regions spanning a cumulative length of 44.33 Mb. Additionally, employing the ED association algorithm led to the identification of 70 trait-associated regions. Within these association regions, a total of 399 genes were annotated. Notably, three genes harbored non-synonymous mutant SNP loci, suggesting a direct association with the trait under investigation. These genes were considered as promising candidates for subsequent analyses.

Further analysis of the 399 annotated genes revealed that the KEGG enrichment results highlighted the enrichment of ABC transporters, which are responsible for facilitating the transportation of various substances across membranes. Among these transporters, certain inward transporters play a vital role in

the transport of nutrients, including amino acids and sugars, from the extracellular environment to the intracellular matrix. This process promotes cell growth and viability (Theodoulou and Kerr, 2015). Additionally, several amino acid metabolite pathways, such as Glycine, serine, and threonine metabolism, Alanine, aspartate, glutamate metabolism, Arginine and proline metabolism, and Arginine biosynthesis, were identified through KEGG enrichment analysis. Furthermore, specific metabolites associated with cell wall, respiratory, and protective functions were found to be correlated with genotypic superiority and yield stability (Vergara-Diaz et al., 2020).

Environmental factors, including temperature, photoperiod, water, and mineral availability, have been shown to regulate spikelet differentiation rate, apical spikelet formation, and ultimately determine the final grain number. Carbon fixation and metabolic pathways were also enriched in our analysis. Gámez et al. (2020)
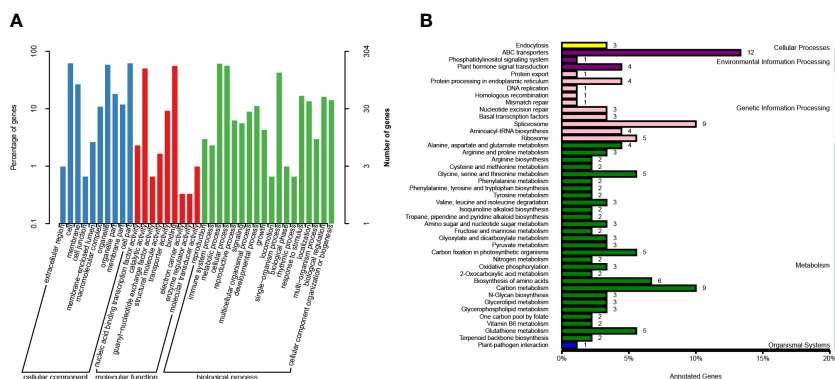

FIGURE 5

Gene KEGG and GO annotations in candidate regions. **(A)** Gene annotations based on the GO database. The proportion of genes belonging to the three main categories of GO classification is represented on the left Y-axis, while the number of genes in each category is depicted on the right Y-axis. **(B)** Gene annotations derived from the KEGG database. The Y-axis represents the top 50 metabolic pathways, whereas the X-axis displays the number of genes annotated to these pathways and their ratio in relation to the total number of annotated genes.

demonstrated that a significant portion of the carbon accumulated in wheat seeds is derived from the spikelet. They further found that the spikelet exhibits twice the capacity of the flag leaf lamina in re-fixing respired CO_2 and shows higher rates of gross photosynthesis and respiration compared to the flag leaf.

The GO analysis revealed the enrichment of terms related to hormones, kinases, and floral development, particularly in response to growth hormone (GO:0060416). This finding underscores the role of growth hormone, a key plant hormone, in promoting cell polarity establishment, cell elongation, and influencing spike type and the number of grains per spike (Gallavotti et al., 2008). Moreover, auxin transport proteins were identified as critical for auxin distribution during wheat spike development, suggesting their significance in regulating this process (Li et al., 2018).

The inflorescence branching pattern in wheat spikes is a crucial trait that significantly impacts floret formation and overall productivity. Transcription factor annotation of genes within candidate regions has revealed a noteworthy similarity to members of the APETALA2 (AP2)-like transcription factor family, particularly the gene *TraesCS3A01G259900*, which bears a close relationship to the Q gene (Simons et al., 2006). The Q gene is considered a major gene in wheat domestication, and its q allele is prevalent in cultivated wheat varieties. Extensive research has highlighted the q allele's influence on various traits, including subcompact spike phenotype, rachis fragility, gum toughness, plant height, and flowering time. Moreover, it is associated with deltoid spike morphology and non-free-threshing grains (Faris et al., 2005).Loss-of-function alleles at the q locus in wheat exhibit an extended developmental stage of the rachilla, leading to a significant increase in the number of florets per spikelet. MicroRNA172 (miRNA172) interactions with its target genes, particularly the genes encoding AP2-like transcription factors such as the Q gene, play a pivotal role in cereal inflorescence development and structure (Debernardi et al., 2017; Greenwood et al., 2017). The AP2 transcription factor family, including the Q gene (*AP2L5*) and its associated paralog *AP2L2*, is regulated by miRNA172 and demonstrates critical and redundant functions in specifying axillary floral meristems (Debernardi et al., 2020). The proper balance of AP2-like gene expression, located on chromosomes 5AL (Q), 2AL (*AP2L2-2A*), 2BL (*AP2L2-2B*), and 2DL (*AP2L2*), is essential for spikelet and floret development (Adonina et al., 2021). Structural reorganization of a chromosome during hybridization can disrupt this balance, thereby offering an opportunity to modify spikelet structure and potentially improve grain yield by manipulating this regulatory module.The*WFZP* gene, a member of the *AP2/ERF* transcription factor family, has been experimentally demonstrated to regulate spikelet development and exhibits conserved functionality between wheat and Brachypodium. Mutations in *WFZP* can affect spikelet development by activating *TaGW5* or directly or indirectly inhibiting *TaGW8*, thereby influencing grain width and weight. Enhancing wheat grain yield remains a challenging task in breeding programs, and the identification of valuable genes and favorable alleles through screening represents a crucial approach towards achieving this goal (Li T. et al., 2021).

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: National Genomics Data Center, accession number CRA015054.

Author contributions

JW: Conceptualization, Formal analysis, Funding acquisition, Investigation, Software, Supervision, Validation, Writing – original draft. EW: Data curation, Investigation, Writing – review & editing. SC: Data curation, Validation, Writing – review & editing. AM: Investigation, Validation, Visualization, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1361621/full#supplementary-material

References

Adonina, I. G., Shcherban, A. B., Zorina, M. V., Mehdiyeva, S. P., Timonova, E. M., and Salina, E. A. (2021). Genetic features of Triticale-wheat Hybrids with vaviloid-type spike Branching. *Plants (Basel)* 11, 58. doi: 10.3390/plants11010058

Alqudah, A. M., Haile, J. K., Alomari, D. Z., Pozniak, C. J., and Börner, A. (2020). Genome-wide and SNP network analyses reveal genetic control of spikelet sterility and yield-related traits in wheat. *Sci. Rep.* 10, 2098. doi: 10.1038/s41598-020-59004-4

Chai, L., Chen, Z. Y., Bian, R. L., Zhai, H. J., Cheng, X. J., Peng, H. R., et al. (2019). Correction to: Dissection of two quantitative trait loci with pleiotropic effects on plant height and spike length linked in coupling phase on the short arm of chromosome 2D of common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 132, 1815–1831. doi: 10.1007/s00122-019-03420-2

Chen, G. F., Wu, R. G., Li, D. M., Yu, H. X., Deng, Z. Y., and Tian, J. C. (2017). Genomewide association study for seeding emergence and tiller number using SNP markers in an elite winter wheat population. *J. Genet.* 96, 177–186. doi: 10.1007/s12041-016-0731-1

Chen, Z. Y., He, Y. C., Iqbal, Y., Shi, Y. L., Huang, H. M., and Yi, Z. L. (2022). Investigation of genetic relationships within three miscanthus species using SNP markers identified with SLAF-seq. *BMC Genomics* 23, 43. doi: 10.1186/s12864-021-08277-8

Clifton-brown, J., Harfouche, A., Casler, M. D., Jones, H., Macalpine, W. J., Murphy-Bokern, D., et al. (2019). Breeding progress and preparedness for the mass-scale deployment of perennial lignocellulosic biomass crops switchgrass, miscanthus, willow, and poplar. *Glob Change Biol. Bioenergy* 11, 118–151. doi: 10.1111/gcbb.12566

Cui, F., Ding, A., Li, J., Zhao, C. H., Wang, L., Wang, X. Q., et al. (2012). QTL detection of seven spike-related traits and their genetic correlations in wheat using two related RIL populations. *Euphytica* 186, 177–192. doi: 10.1007/s10681-011-0550-7

Das, S., Upadhyaya, H. D., Srivastava, R., Bajaj, D., Gowda, C. L. L., Sharma, S., et al. (2015). Genome-wide insertion-deletion (InDel) marker discovery and genotyping for genomics-assisted breeding applications in chickpeas. *DNA Res.* 22, 377–386. doi: 10.1093/dnares/dsv020

Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., and Blaxter, M. L. (2013). Special features of RAD Sequencing data: implications for genotyping. *Mol. Ecol.* 22, 3151–3164. doi: 10.1111/mec.12084

Debernardi, J., Greenwood, J. R., Finnegan, E. J., and Dubcovsky, J. (2020). APETALA 2-like genes AP2L2 and Q specify lemma identity and axillary floral meristem development in wheat. *Plant J.* 101, 171–187. doi: 10.1111/tpj.14528

Debernardi, J., Lin, H. Q., Faris, J., and Dubcovsky, J. (2017). microRNA172 plays a crucial role in wheat spike morphogenesis and grainthreshability. *Development* 144, 1966–1975. doi: 10.1242/dev.146399

Du, H. S., Wen, C. L., Zhang, X. F., Xu, X. L., Yang, J. J., Chen, B., et al. (2019). Identification of a major QTL (*qRRs-10.1*) that confers resistance to *ralstoniasolanacearum* in pepper (*Capsicum annuum*) using SLAF-BSA and QTL mapping. *Int. J. MolSci* 20, 5887. doi: 10.3390/ijms20235887

Faris, J. D., Simons, K. J., Zhang, Z., and Gill, B. S. (2005). The wheat superdomestication gene Q. *Front. Wheat Bioscience* 100, 129–148.

Faris, J. D., Zhang, Z., Garvin, D. F., and Xu, S. S. (2014). Molecular and comparative mapping of genes governing spike compactness from wild emmer wheat. *Mol. Genet. Genomics* 289, 641–651. doi: 10.1007/s00438-014-0836-2

Fekih, R., Takagi, H., Tamiru, M., Abe, A., Natsume, S., Yaegashi, H., et al. (2013). MutMap+: genetic mapping and mutant identification without crossing in rice. *PloS One* 8, e68529. doi: 10.1371/journal.pone.0068529

Gallavotti, A., Yang, Y., Schmidt, R. J., and Jackson, D. (2008). The relationship between auxin transport and maize branching. *Plant Physiol.* 147, 1913–1923. doi: 10.1104/pp.108.121541

Gámez, A. L., Vicente, R., Sanchez-Bragado, R., Jauregui, I., Morcuende, R., Goicoechea, N., et al. (2020). Differential flag leaf and ear photosynthetic performance under elevated (CO₂) conditions during grain filling period in durum wheat. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.587958

Geng, X., Jiang, C. H., Yang, J., Wang, L. J., Wu, X. M., and Wei, W. H. (2016). Rapid identification of candidate genes for seed weight using the SLAF-Seq method in Brassica napus. *PloS One* 11, e0147580. doi: 10.1371/journal.pone.0147580

Graner, A., Tuberosa, R., and Frison, E. (2013). *Genomics of plant genetic resources* (Berlin: Springer).

Greenwood, J. R., Finnegan, E. J., Watanabe, N., Trevaskis, B., and Swai, S. M. (2017). New alleles of the wheat domestication gene Q reveal multiple roles in growth and reproductive development. *Development* 144, 1959–1965. doi: 10.1242/dev.146407

Guo, J., Shi, W., Zhang, Z., Cheng, J. Y., Sun, D. Z., Yu, J., et al. (2018). Association of yield-related traits in founder genotypes and derivatives of common wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 18, 38. doi: 10.1186/s12870-018-1234-4

Hill, J. T., Demarest, B. L., Bisgrove, B. W., Gorsi, B., and Yi, C. Y. (2013). MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. *Genome Res.* 23, 687–697. doi: 10.1101/gr.146936.112

Jain, M., Moharana, K. C., Shankar, R., Kumari, R., and Garg, R. (2014). Genomewide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. *Plant Biotechnol. J.* 12, 253–264. doi: 10.1111/pbi.12133

Li, T., Deng, G. B., Su, Y., Yang, Z., and Long, H. (2021). Identification and validation of two major QTLs for spike compactness and length in bread wheat (*Triticum aestivum* L.) showing pleiotropic effects on yield-related traits. *Theor. Appl. Genet.* 134, 3625–3641. doi: 10.1007/s00122-021-03918-8

Li, Y. P., Fu, X., Zhao, M. C., Zhang, W., Li, B., An, D. G., et al. (2018). A genome-wide view of transcriptome dynamics during early spike development in bread wheat. *Sci. Rep.* 8, 15338. doi: 10.1038/s41598-018-33718-y

Li, T., Li, Q., Wang, J. H., Yang, Z., Tang, Y. Y., Su, Y., et al. (2022). High-resolution detection of quantitative trait loci for seven important yield-related traits in wheat (*Triticum aestivum* L.) using a high-density SLAF-seq genetic map. *BMC Genom Data* 23, 37. doi: 10.1186/s12863-022-01050-0

Li, Y. P., Li, L., Zhao, M. C., Guo, L., Guo, X. X., Zhao, D., et al. (2021). Wheat FRIZZY PANICLE activates VERNALIZATION1-A and HOMEOBOX4-A to regulate spike development in wheat. *Plant Biotechnol J.* 19, 1141–1154. doi: 10.1111/pbi.13535

Liang, F., Xin, X. Y., Hu, Z. J., Xu, J. D., Wei, G., Qian, X. Y., et al. (2011). Genetic analysis and fine mapping of a novel semi-dominant dwarfing gene LB₄D in rice. *J. Integr. Plant Biol.* 53, 312–323. doi: 10.1111/j.1744-7909.2011.01031.x

Malik, P., Kumar, J., Sharma, S., Sharma, R., and Sharma, S. (2021). Multi-locus genome-wide association mapping for spike-related traits in bread wheat (*Triticum aestivum* L.). *BMC Genomics.* 22, 597. doi: 10.1186/s12864-021-07834-5

Michenlmore, R. W., and Kesseli, I. R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. U. S. A* 88, 9828–9832. doi: 10.1073/pnas.88.21.9828

Shen, C., Jin, X., Zhu, D., and Lin, Z. X. (2017). Uncovering SNP and indel variations of tetraploid cotton by SLAF-seq. *BMC Genomics* 18, 247. doi: 10.1186/s12864-017-3643-4

Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y. S., Gill, B. S., et al. (2006). Molecular characterization of the major wheat domestication gene Q. *Genetics* 172, 547–555. doi: 10.1534/genetics.105.044727

Slafer, G. A., Elia, M., Savin, R., García, G. A., Terrile, I. I., Ferrante, A., et al. (2015). Fruiting efficiency: an alternative trait to further rise wheat yield. *Food Energy Secur.* 4, 92–109. doi: 10.1002/fes3.59

Song, J., Li, Z., Liu, Z. L., Guo, Y., and Qiu, L. J. (2017). Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00919

Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., et al. (2013). SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PloS One* 8, e58700. doi: 10.1371/journal.pone.0058700

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105

Theodoulou, F. L., and Kerr, I. D. (2015). ABC transporter research: going strong 40 years on. *BiochemSoc Trans.* 43, 1033–1040. doi: 10.1042/bst20150139

Vergara-Diaz, O., Vatter, T., Vicente, R., Obata, T., Nieto-Taladriz, M. T., Aparicio, N., et al. (2020). Metabolome profiling supports the key role of the spike in wheat yield performance. *Cells* 9, 1025. doi: 10.3390/cells9041025

Wang, X., Ren, G., Li, X., Tu, J., Lin, Z., and Zhang, X. (2012). Development and evaluation of intron and insertion–deletion markers for *Gossypium* barbadense. *Plant Mol. Biol. Rep.* 30, 605–613. doi: 10.1007/s11105-011-0369-3

Wang, H. Y., Wang, X. Y., Yan, D. W., Sun, H., Chen, Q., Li, M. L., et al. (2022). Genome-wide association study identifying genetic variants associated with carcass backfat thickness, lean percentage, and fat percentage in a four-way crossbred pig population using SLAF-seq technology. *BMC Genomics* 23, 594. doi: 10.1186/s12864-022-08827-8

Xu, X. M., Chao, J., Cheng, X. L., Wang, R., Sun, B. J., Wang, H. M., et al. (2016). Mapping of a novel race specific resistance gene to phytophthora root rot of Pepper (*Capsicum annuum*) using bulked segregant analysis combined with specific length amplified fragment sequencing strategy. *PloS One* 11, e0151401. doi: 10.1371/journal.pone.0151401

You, J. N., Liu, H., Wang, S. R., Luo, W., Gou, L. L., Tang, H. P., et al. (2021). Spike density quantitative trait loci detection and analysis in tetraploid and hexaploidwheat recombinant inbred line populations. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.796397

Yu, Q., Feng, B., Xu, Z. B., Fan, X. L., Zhou, Q., Ji, G. ,. S., et al. (2022). Genetic dissection of three major quantitative trait loci for spike compactness and length in bread wheat (*Triticum aestivum* L.). *Front. PlantSci* 13. doi: 10.3389/fpls.2022.882655

Zhang, X. F., Wangm, G. Y., Chenm, B., Dum, H. S., and Gengm, S. S. (2018). Candidate genes for first flower node identified in pepper using combined SLAF-seq and BSA. *PloS One* 13, e0194071. doi: 10.1371/journal.pone.0194071

Zheng, Y. T., Xu, F., Li, Q. K., Wang, G. J., Liu, N., Gong, Y. M., et al. (2018). QTL mapping combined with bulked segregant analysis identify SNP markers linked to leaf shape traits in Pisum sativum using SLAF sequencing. *Front. Genet.* 9. doi: 10.3389/fgene.2018.00615

Zhu, Y., Han, L. Y., Li, P., Kang, X. L., and Shi, Y. G. (2021). Investigation of genetic markers for intramuscular fat in the hybrid Wagyu cattle with bulked segregant analysis. *Sci. Rep.* 11, 11530. doi: 10.21203/rs.3.rs-277670/v1