

RESEARCH ARTICLE

Open Access

Genome-wide identification, phylogeny and expression analysis of the *SPL* gene family in wheat



Ting Zhu¹, Yue Liu², Liting Ma¹, Xiaoying Wang¹, Dazhong Zhang¹, Yucui Han¹, Qin Ding³ and Lingjian Ma^{1*}

Abstract

Background: Members of the plant-specific *SPL* gene family (squamosa promoter-binding protein -like) contain the SBP conserved domain and are involved in the regulation of plant growth and development, including the development of plant flowers and plant epidermal hair, the plant stress response, and the synthesis of secondary metabolites. This family has been identified in various plants. However, there is no systematic analysis of the *SPL* gene family at the genome-wide level of wheat.

Results: In this study, 56 putative *TaSPL* genes were identified using the comparative genomics method; we renamed them *TaSPL001* - *TaSPL056* on their chromosomal distribution. According to the un-rooted neighbor joining phylogenetic tree, gene structure and motif analyses, the 56 *TaSPL* genes were divided into 8 subgroups. A total of 81 *TaSPL* gene pairs were designated as arising from duplication events and 64 interacting protein branches were identified as involve in the protein interaction network. The expression patterns of 21 randomly selected *TaSPL* genes in different tissues (roots, stems, leaves and inflorescence) and under 4 treatments (abscisic acid, gibberellin, drought and salt) were detected by quantitative real-time polymerase chain reaction (qRT-PCR).

Conclusions: The wheat genome contains 56 *TaSPL* genes and those in same subfamily share similar gene structure and motifs. *TaSPL* gene expansion occurred through segmental duplication events. Combining the results of transcriptional and qRT-PCR analyses, most of these *TaSPL* genes were found to regulate inflorescence and spike development. Additionally, we found that 13 *TaSPLs* were upregulated by abscisic acid, indicating that *TaSPL* genes play a positive role in the abscisic acid-mediated pathway of the seedling stage. This study provides comprehensive information on the *SPL* gene family of wheat and lays a solid foundation for elucidating the biological functions of *TaSPLs* and improvement of wheat yield.

Keywords: *SPL* gene family, Phylogenetic analysis, Expression patterns analysis, Wheat

Background

Gene families, consisting of multiple genes that share similar structures and functions, act in important roles for a given organism. Now, all kinds of gene families have been found in eukaryotes, such as *bHLH* [1], *TCP* [2], and *Prrx* [3]. Among all gene families, *SPL* (squamosa promoter

binding protein, SBP) is a plant-specific family that is widely distributed in green plants. The plant *SPL* proteins, bind to the *SQUAMOSA* promoter of *MADS*-Box genes, were first identified in a cDNA library of the *Antirrhinum majus* inflorescence [4]. These proteins contain a specific SBP domain that comprises approximately 70 amino acids and possesses two zinc finger sites (Cys-Cys-His-Cys and Cys-Cys-Cys-His) [5, 6]. In this domain, the 4 amino acid residues coordinate one zinc ion, playing an important role in maintaining the stability of the protein's

* Correspondence: malingjian@nwfau.edu.cn

¹College of Agronomy, Northwest A&F University, Yangling 712100, China
Full list of author information is available at the end of the article



© The Author(s). 2020 **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.

configuration. Besides, a conserved nuclear localization signal located at the C-terminus of the SBP domain overlaps with the second zinc finger structure [7] and directs the protein to the nucleus for regulating the transcription of downstream genes.

In recent years, a large number of *SPL* genes have been successfully cloned in various plants, and these *SPL* genes distribute from *C. reinhardtii* to *P. patens* and higher plants [8]. At present, the identification and evolution of the *SPL* gene family have been studied in *A. thaliana* [9, 10], rice [5], soybean [11], *Populus* [12], *Petunia* [13], *Gossypium* [14], and tobacco [15]. Based on the analysis of sequence homology and phylogeny, this family is usually classed into 6–9 subgroups. Guo [8] reported that 120 *SPL* genes from nine species divided into 3 groups, all *SPL* genes from land plants have been classed into two distinct groups and 7 subgroups. Yang [5] divided 35 *SPLs* from rice and *A. thaliana* into three categories: A, B, and C, with all *SPL* genes classified into 9 subfamilies. In *Populus* [12], the 28 *PtSPLs* were separated into 8 subgroups. In *Salvia miltiorrhiza*, 15 *SPL* genes were divided into 6 subgroups [16]. Unfortunately, little research has been conducted to show the classification of the *SPL* gene family in wheat.

The expression levels of *SPL* genes can be controlled by miR156/157. Rhoades [17] and Schwab [18] found that 11 of 17 *SPLs* in *A. thaliana* contain miR156/157 recognition sites. Schwab [18] reported that overexpressing miR156 in *A. thaliana* downregulated the expression of multiple *SPLs*, however, *SPLs* without this recognition site were unaffected. After mutating the miR156/157 recognition site of *AtSPL3*, the transcription level of *AtSPL3* was significantly increased [19]. Salinas observed the complementary expression fashion in apices and stamens of *SlySPL* genes with miR156/157, revealing that *SlySPL* genes are regulated by miR156/157 [20].

The *SPL* genes, whose expression levels are governed by miR156/157, play critical roles in various aspects of plant growth and physiology, including plant embryos, tissues, and vegetative-phase changes, transduction of gibberellin and light signal, and drought stress and salt stress responses. For instance, Inhibiting post-translational modification of the *AtSPL3* protein caused early flowering in *Arabidopsis* [19]. *AtSPL3*, *-4*, and *-5* regulated the expression of the key downstream genes *AtAPI*, *AtFUL*, and *AtLFY* to be involved in floral meristem development [10]. Zhang has shown that *SPL8* gene positively regulated GA signaling in the flower, however, negative roles were found in the seedling, implying that *AtSPL8* modulates GA signaling in plant development [21]. Wang purported that *OsSPL16* can be effectively improved the grain quality and yield, by increasing cell division and grain filling, implying that this gene has positive consequences for rice grain development [22]. *OsSPL14* inhibited the number of tillers

in rice, but promote panicle branching to increase grain weight together with stronger stems [23]. At low temperature (5 °C), both *VvSBP3* and *VvSBP5* were upregulated but *VvSBP4* and *VvSBP7* were downregulated, indicating that *VvSBP3* and *VvSBP5* are associated with low-temperature responses in *Vitis vinifera L.* [24]. The *BpSPL9* respond to the salt and drought stress, Overexpressed the *BdSPL9* in birch leaves can scavenge of ROS to improve the tolerance of salt and drought stress [25]. Much is know about the functions of *SPL* genes in *A. thaliana* and rice, however, such information for wheat is limited.

The plant-specific *SPL* family plays an important role in growth and development. With the progress of genome-sequencing techniques, various transcription factors or gene families in plants have been identified. Although many plant *SPL* gene family members have been identified, this is not the case for wheat, and knowledge of *SPL* gene functions in this species is limited. In this study, we used bioinformatics methods to identify *SPL* genes in wheat, and we analyzed *TaSPL* characteristics based on the results of gene structures, motifs, cis-elements, phylogenetic relationships, gene duplications, GO annotation, protein-protein interactions, and gene expression patterns, and predicted their functions. Our results provided information on the functional elucidation and evolution of *SPL* genes in wheat.

Results

Genome-wide identification and chromosomal location of *TaSPL* genes

To obtain *TaSPL* genes from the whole genome of wheat, two methods, HMM and BLASTP, were used for identification, and two online websites, Pfam and CDD, were used for confirmation. Ultimately, 56 genes were designated *TaSPL* genes; all are unevenly distributed on the wheat chromosomes. The 56 putative *TaSPL* genes were renamed *TaSPL001* to *TaSPL056* based on chromosomal locations (Additional file 1: Table S1). Among them, 34% (19) mapped to the A and D genomes, respectively; 32% (18) located on the B genome. TaChr7A and TaChr7D contain a large number of *SPL* genes, but no *TaSPL* genes are located on chromosomes 4B and 4D.

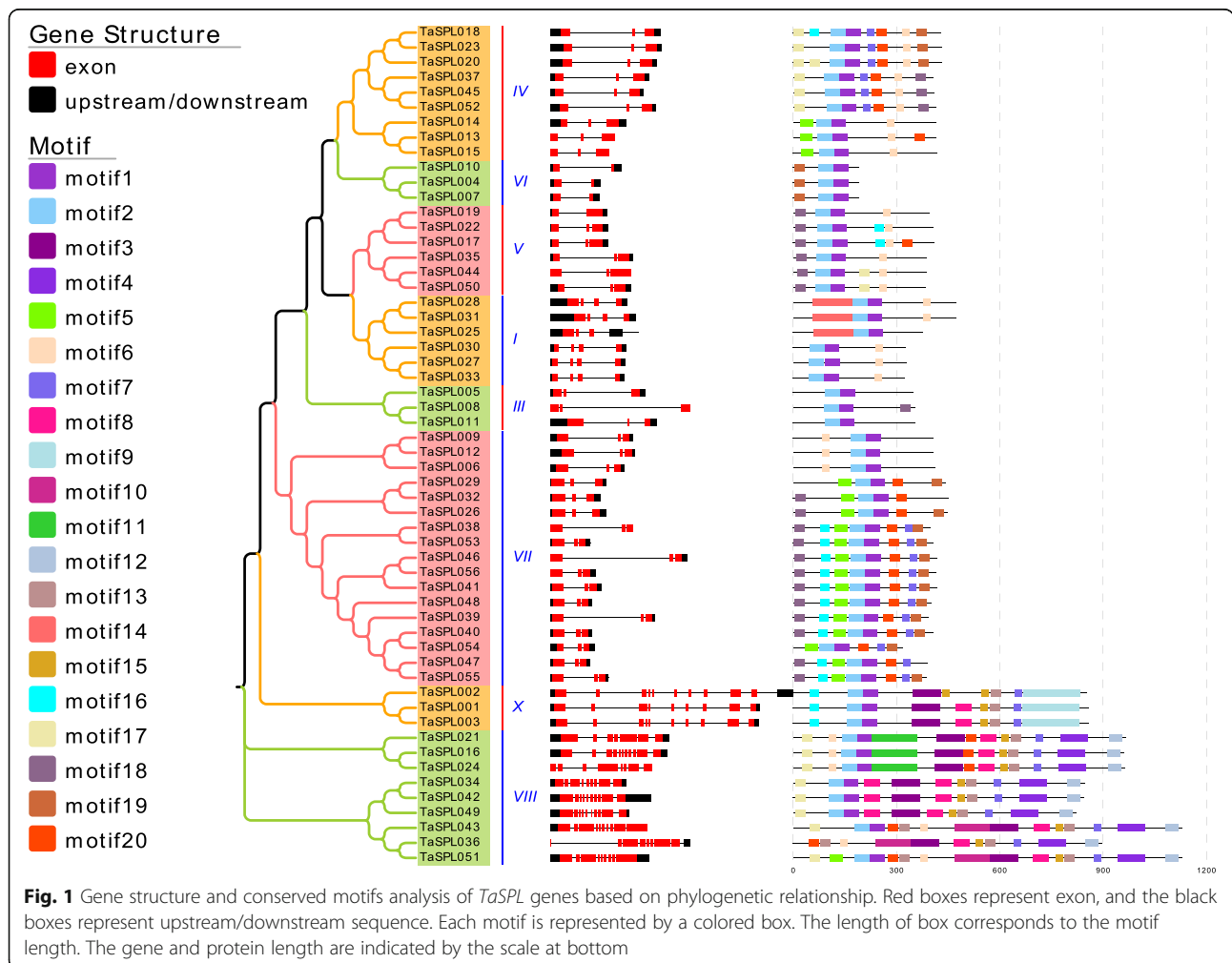
The 56 *SPL* proteins were submitted to the cello website to predict their subcellular localizations; as shown in Table S1, all of the *SPL* proteins were localized to the nucleus, and 3 *SPL* (*TaSPL034*, *TaSPL049*, *TaSPL052*) also distribute to the plasma membrane, suggesting that the 56 *TaSPL* proteins perform their functions in the nucleus. The 56 predicted *TaSPL* proteins exhibit diversity in amino acid length and protein molecular weights (MWs). The MWs of the *TaSPL* proteins range from 20.11 to 123.69 kDa. The lengths of the *TaSPL* proteins vary from 192 to 1129

aa, with the longest protein being TaSPL043 and the shortest being TaSPL004, TaSPL007, and TaSPL010.

Gene structure and motif compositions of the SPL gene family in wheat

Gene structure and motif diversity is a mechanism that promotes the evolution of gene families. We constructed a phylogenetic tree of *TaSPL* genes and analyzed the exons, introns, and conserved motifs (Fig. 1). The results showed that despite being different with regard to exon position, *TaSPL* genes from the same subfamily share a similar genetic structure, and the number of exons in most *TaSPL* genes is conserved. Approximately 59% (33) of the genes contain 3 exons, but 12 genes in subgroup VIII and subgroup X possess the largest number of exons, at 8–11. For example, *TaSPL001* in subfamily X contains 10 exons, and *TaSPL042* in subfamily VIII contains 11 exons.

We submitted the entire protein sequences of TaSPL to MEME software to detect the motif compositions and explore the motif diversity. Twenty specific motifs were defined and named motif 1 to 20. The TaSPL proteins exhibit similar conserved motif compositions, in particular, all TaSPL proteins except TaSPL036 contain motif 1 and motif 2, suggesting that these two motifs are important components for TaSPL protein sequences. Not only the common motifs exist in all SPL proteins, but also SPL members in separate subgroups also contain their specific motifs. For example, motif 18 is present in all SPL members of subfamily VII and subfamily IV; motif 20 is found in subfamily VII. In short, members in the same subfamily share similar gene structure and motif compositions, while different subgroups contain the specific structure, implying that the *TaSPL* gene family presents the functional conservation and diversity during evolution.



Multiple sequence alignment, phylogenetic analysis of *TaSPLs*

Multiple sequence alignment of the full-length *TaSPL* proteins was carried out. The results showed that the specific SBP domain is conserved among the 56 *TaSPLs* (Fig. 2). Approximately 76 aa comprise an SBP domain that contains two zinc finger motifs with the structure Cys-Cys-Cys-His and Cys-Cys-His-Cys, and one nuclear localization signal (NSL) at the C-terminus. Three conserved sequences, CQQC, SCR, and RRR, appear in the *TaSBP* domain, except for *TaSPL036*, which does not possess the CQQC sequence.

To evaluate the evolutionary relationships of *SPL* genes in wheat, rice, and *A. thaliana*, a neighbor-joining phylogenetic tree was constructed using the full-length *SPL* proteins (Fig. 3A). Ninety-one *SPL* genes from the above three species were divided into 10 subfamilies (I to X). According to this analysis, the *SPL* family of wheat is closer to *SPL* genes in rice than *A. thaliana*. Except for subfamilies II, III, VI, and IX, the remaining subfamilies contain 3 species *SPL* genes. For example, subfamily I contains 3 *AtSPLs*, 4 *OsSPLs*, and 6 *TaSPLs*. 2 *AtSPLs*, 2 *OsSPLs*, and 6 *TaSPLs* comprise subfamily V. This indicates that *TaSPL* genes did not evolve with the characteristics of monocotyledons and dicotyledons, and that the *SPL* gene family was formed before the differentiation of these two classes of the plant. We also found 22 paralogous gene pairs in the NJ-tree, among them, 4 gene pairs are from *A. thaliana* and 18 genes from wheat, indicating that the *TaSPL* genes expanded in a species-specific manner.

The same methods were used to identify 17, 13, and 17 *SPL* genes from *B. distachyon*, *T.urartu*, and *A. tauschii* genomes and downloaded the 17 *SPL* gene sequences of barley [26]. Then we constructed a phylogenetic tree (Fig. 3B). The 120 *SPL* genes from 5 species

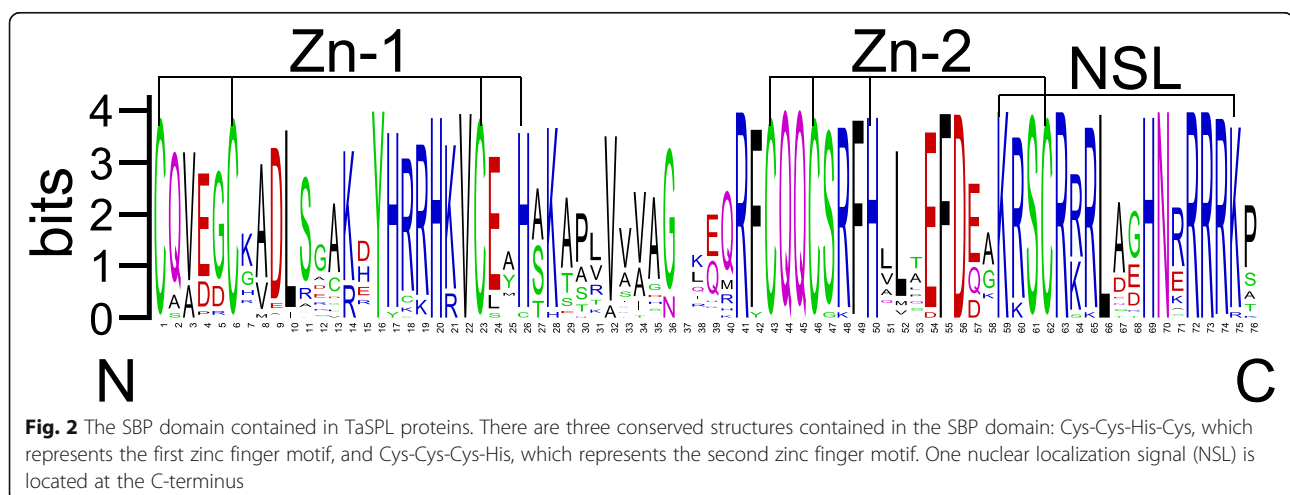
were still divided into 8 subgroups. We detected 13 orthologous *SPL* gene pairs from wheat and *A. tauschii*, 3 from wheat and *T.urartu*, and 1 from wheat and barley. Conversely, no orthologous gene pairs were found in wheat and *B. distachyon*. It is implied that the *TaSPL* genes share a strong evolutionary relationship with *AetSPLs* and *TuSPLs*.

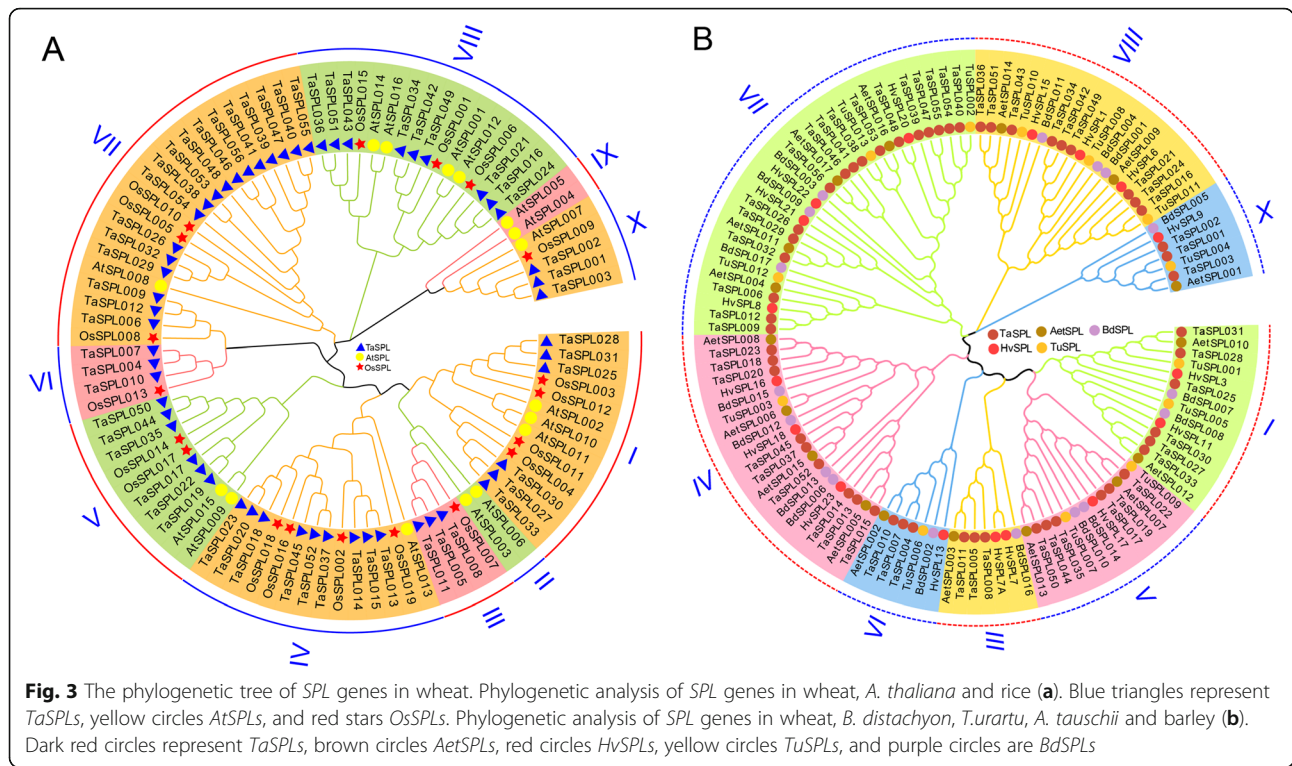
Gene duplication and synteny analysis of *TaSPL* genes

Gene duplication, an indispensable mechanism, can expand new genes that share similar or different functions; therefore, we analyzed the duplication events that occurred in the *TaSPL* gene family. Eighty-one *SPL* gene pairs from wheat were detected as duplicated gene pairs (Fig. 4A, Additional file 1: Table S3), with 8 being tandem duplications, and 73 were associated with a segmental duplication event. The number of segmental duplication events in the *TaSPL* gene family was found to be higher than that of tandem duplications, suggesting that the former was the main route for expanding *SPL* genes in wheat and that many homologous genes on the different wheat chromosomes support the high conservation of the family.

Ka/Ks, the nonsynonymous to synonymous substitution ratio, determines the selection pressure of duplicated genes. Thus, the protein and CDS sequences of each duplicated gene pair were compared, and Ka/Ks ratios and divergence times were calculated. According to the results (Additional file 1: Table S3), the Ka/Ks ratios of all *TaSPL* gene pairs are < 1, indicating that the evolution of *TaSPL* genes was accompanied by intense purifying selection. We also predicted that the divergence time of *TaSPL* gene pairs to be approximately 27 Mya.

We detected the colinearity of wheat with other species at a genome-wide level to better understand the origin of the *TaSPL* gene family. Most *TaSPL* genes have





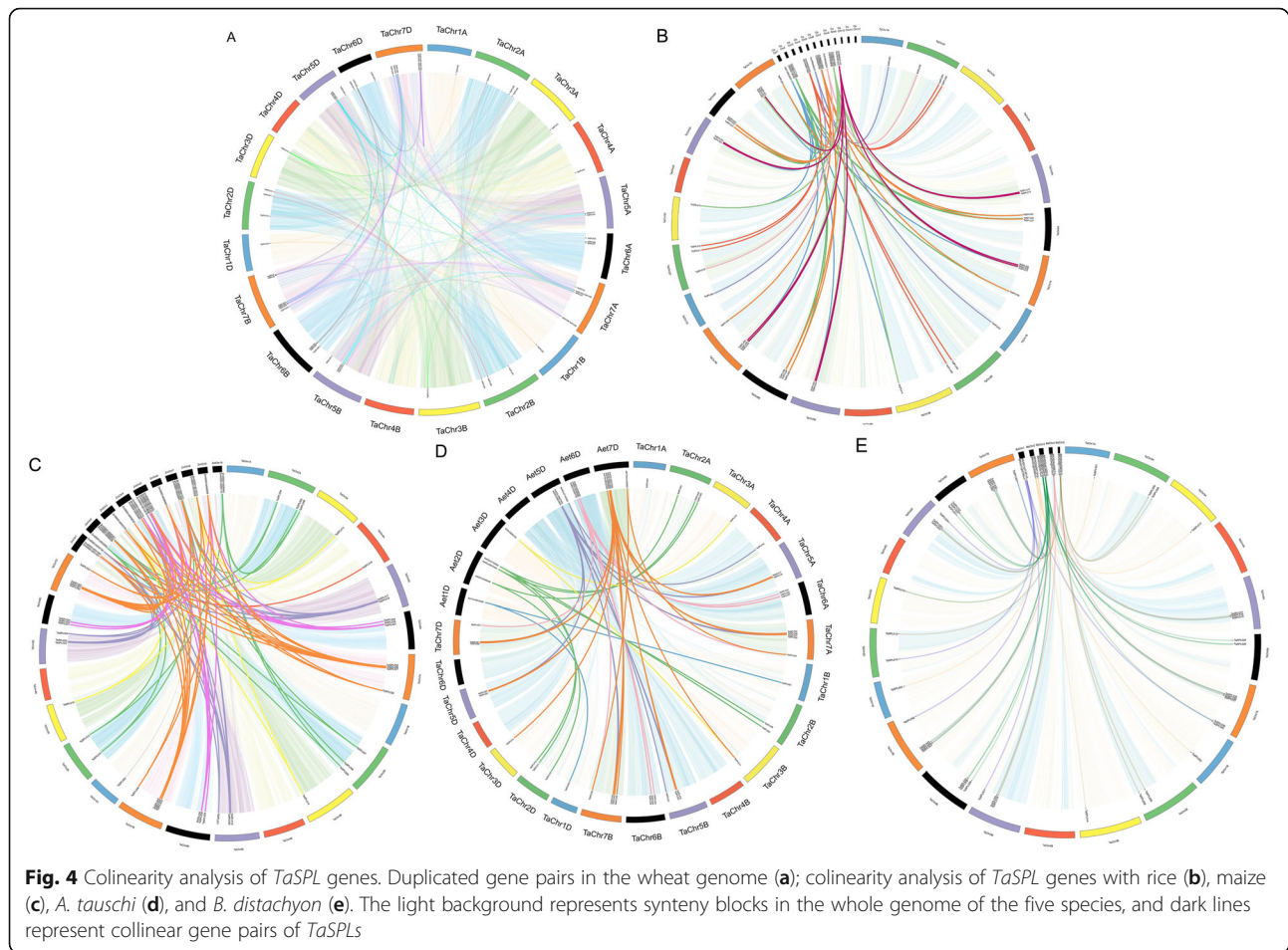
orthologous genes in rice, maize, *A. tauschii*, and *B. distachyon* (Fig. 4B-4E, Additional file 1: Table S4-S7). We detected 67, 110, 52, and 56 gene pairs, respectively. The *Ka/Ks* ratios between wheat and rice, maize, *A. tauschii*, and *B. distachyon* were calculated to be 0.38, 0.39, 0.29, and 0.31. All of the collinear gene pairs showed a value lower than 1, confirming that the evolution of the *SPL* gene family in wheat has undergone a strong purifying selection. The divergence time of collinear gene pairs in rice was the same as that in maize, at approximately 52 Mya, and earlier than that in *A. tauschii* (39 Mya) and *B. distachyon* (41Mya), indicating that *TaSPL* gene family shares an intimate correlation with those in rice, maize, *A. tauschii*, and *B. distachyon*.

GO annotation analysis and protein-protein interaction network of *TaSPLs*

We performed GO annotation analysis on the 56 proteins, revealing that they may participate in a range of cellular components, molecular functions and biological processes (Fig. 5, Additional file 1: Table S8). The 56 *TaSPL* proteins were assigned a total of 20 GO terms; although only a few GO terms belong to cellular component, many proteins were enriched in this category. Under the cellular component category, the most highly enriched categories are related to cell part, intracellular part, and organelle. All of the *TaSPLs* can participate in

these three processes, whereas less than 30% of *TaSPLs* are involved in membrane formation. Under molecular function, all proteins have the capacity to bind to other molecules, such as organic cyclic compound binding and ion binding. Among them, 100% of the *TaSPLs* can bind to heterocyclic compounds and 80% to ions. Regarding biological processes, less than 20% of the *TaSPLs* participate in biological processes, such as nitrogen metabolism, and regulation of cellular processes.

To understand the protein-protein interactions between *TaSPLs* and other proteins in wheat, we constructed a protein-protein interaction network (Fig. 6, Additional file 1: Table S9). Thirty-two *TaSPL* proteins and a total of 64 interacting protein branches were detected. Among them, 10% of *TaSPLs* can interact with at least 6 proteins, such as *TaSPL001*, *TaSPL002*, and *TaSPL003*, suggesting that these 3 *TaSPL* proteins play a significant role in the regulation of protein networks. The sequences of additional wheat proteins that interact with *TaSPL* proteins were submitted to the CDD database to predict their conserved domains, which we revealed various transcription factor families, such as AP2, Cu-Zn_Superoxide_Dismutase, and MADS_MEF2_like. Among them, 32 interaction branches were identified as the interaction of *TaSPLs* with the AP2 family. Therefore, we suggest that highly possible that *TaSPL* proteins interact with proteins of the AP2 family to regulate wheat development.



Promoter analysis of *TaSPL* genes

The specific cis-acting elements contained in promoter region can regulate the gene expressions. We analyzed the 1.5-kb DNA sequence upstream of 56 *TaSPL* genes and identify potential cis-acting elements (Figure S1, Additional file 1: Table S10). A variety of cis-elements in the promoter regions of *TaSPL* genes clustered into three subdivisions: hormone and stress response elements, light-responsive elements, and plant growth and development elements. Among the three types, most *TaSPL* genes contain light-responsive elements, such as ACE, Sp1, and G_box. In particular, the G_box was found in 60% (34) of the *TaSPL* genes, suggesting that *TaSPL* expressions may be regulated by light and that the G_box is the most important for wheat development. In addition, the ABRE element, which is associated with abscisic acid responsiveness were distributed widely throughout 80% of the *TaSPL* genes, suggesting that *TaSPLs* are associated with regulation of the ABA metabolic pathway. At the same time, some elements related to plant growth and development, such as CAT_box, O2_site, and HD_ZIP 1, were found in *TaSPL* genes,

indicating that the *TaSPL* gene family plays roles in regulating physiological processes.

Expression patterns of *TaSPL* genes in different wheat tissues

The transcriptional data were downloaded from the wheat expression browser website to determine the tissue-specific expression patterns of *TaSPL* genes (Fig. 7A). The results showed that 53 *TaSPLs* exhibited different expression levels in shoots/leaves, roots, and spikes of wheat. For example, 46 *TaSPLs* displayed high expression in spikes, and 7 genes were found to be highly expressed in roots, implying that *TaSPLs* are primarily expressed in spikes and participate in the development of spikes. In addition, genes in the same subgroup shared similar expression patterns. All *TaSPL* genes in subfamily V showed high expressions in roots and expressions of subfamily VIII members in shoots/leaves, and roots were observed to be low, revealing that *TaSPL* in the same family share conserved functions.

Expressions of *TaSPL* genes in five organs (ovary_early_anthesis, ovary_late_anthesis, spike_booting, spike_

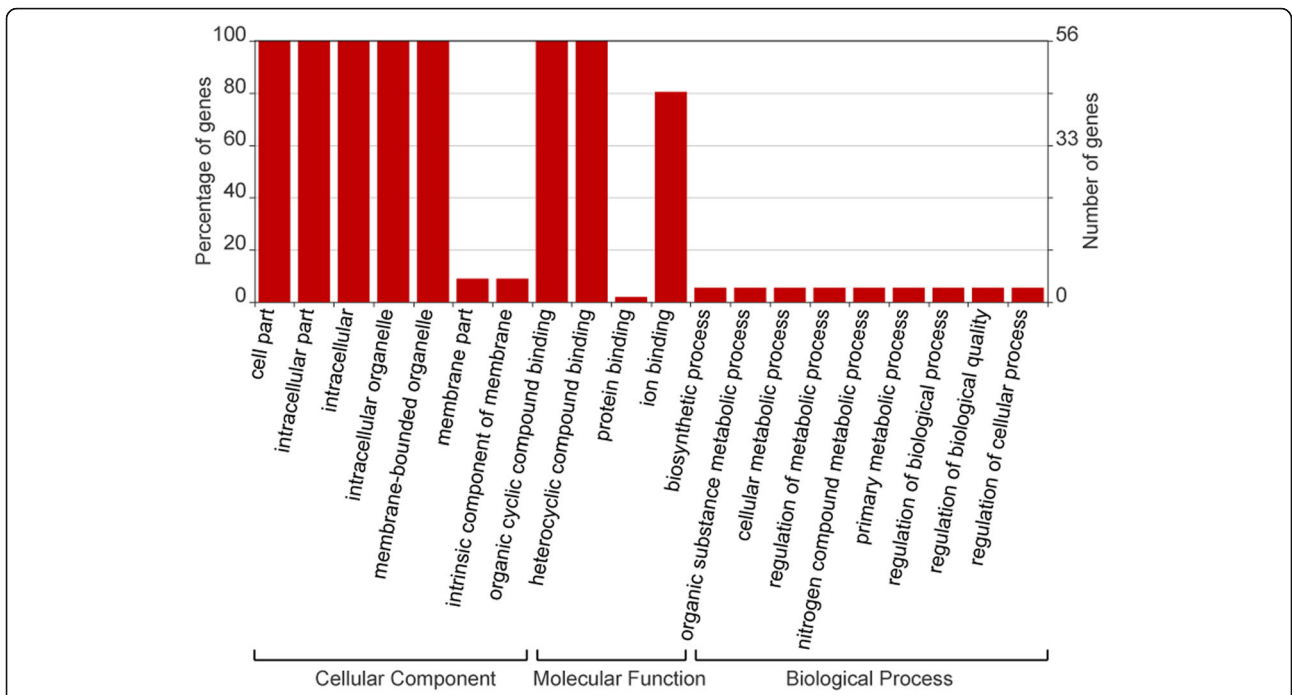


Fig. 5 GO annotations for TaSPL proteins. The GO annotation is divided into three main categories: cellular component, molecular function and biological process. The y-axis on the left represents the percentage of genes in each category, and the number of genes is showed on the right

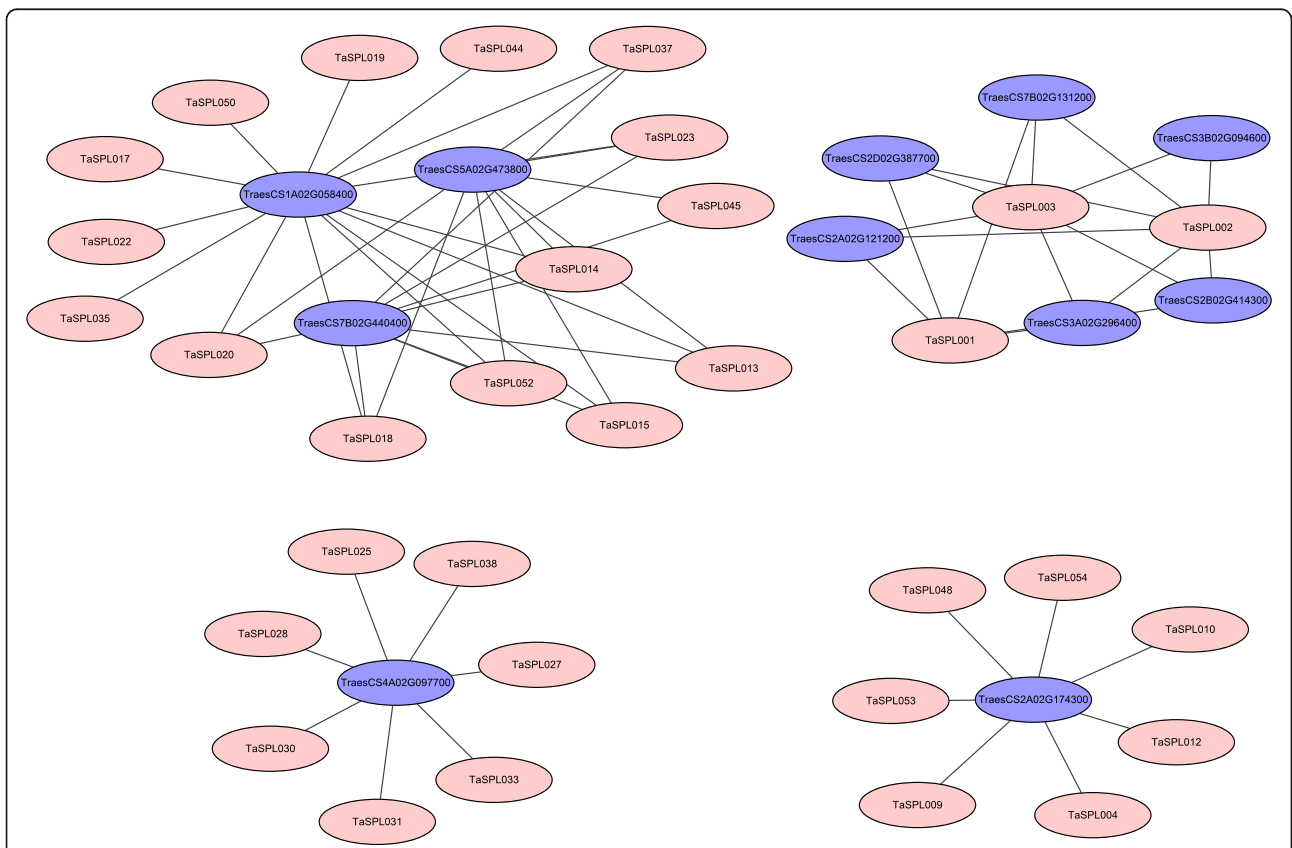
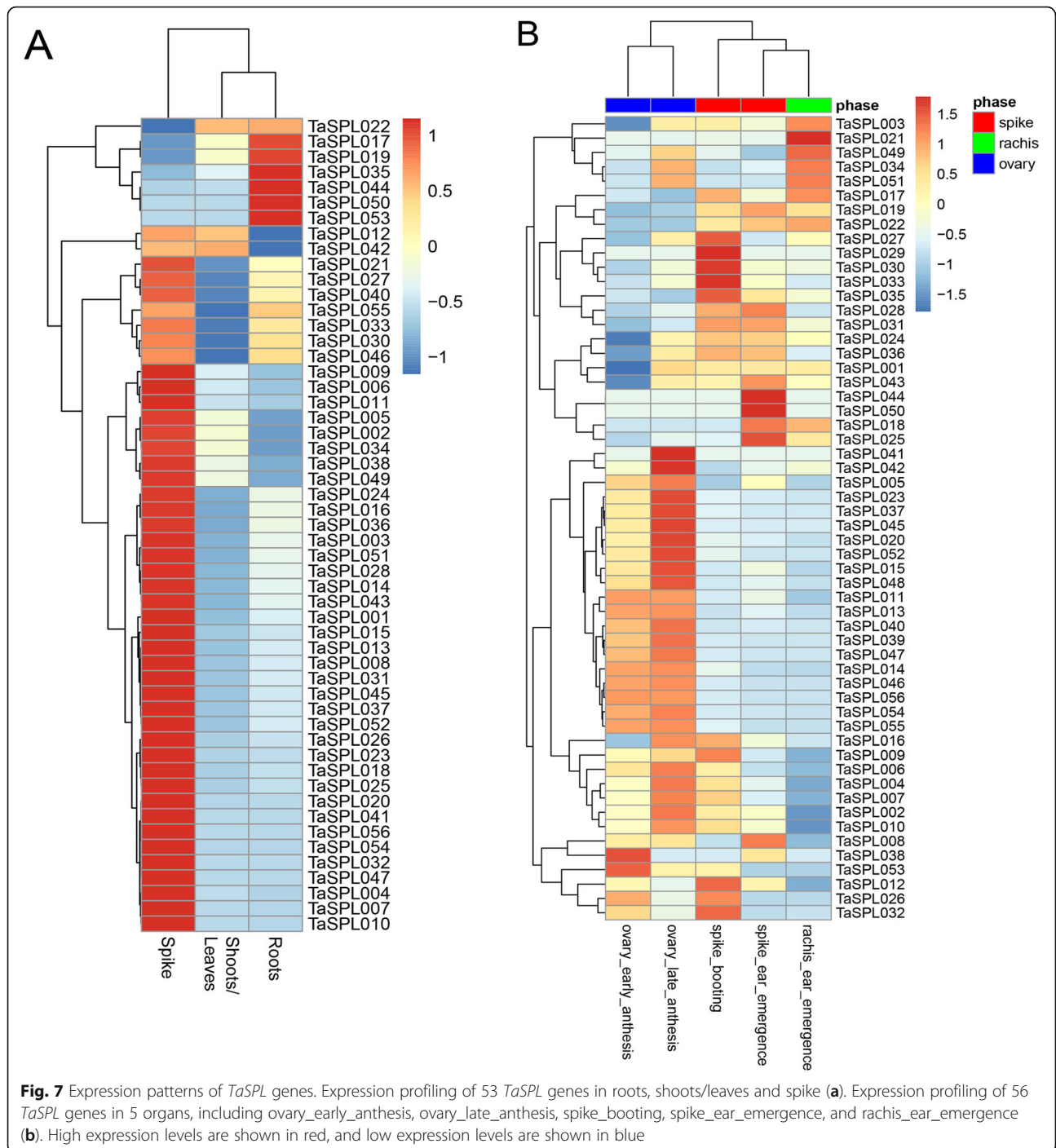


Fig. 6 The protein-protein interaction network between TaSPLs and other wheat proteins. The TaSPL proteins are in pink, and the other wheat proteins are in purple



ear_emergence, and rachis_ear_emergence) were also analyzed (Fig. 7B). The 56 genes were expressed as follows: 75% (42) were highly expressed in ovary_anthesis, 55% (31) were highly expressed in spikes, and only 14% (8) showed high expressions in rachis ear emergence. These findings suggest that *TaSPL* genes are primarily expressed in organs related to flower development. Among them, 55% (32) were mainly found in ovary_early_anthesis. Additionally, 68% (38) showed high

expression in ovary_late_anthesis, especially, 12 *TaSPL* genes, such as *TaSPL005*, *-038*, and *-052*, expressed at higher levels in ovary_anthesis, suggesting that *TaSPL* genes are involved in the development of wheat flowers and that 12 *TaSPL* genes are essential for flower development. Moreover, 46% (26) of the *TaSPLs* displayed elevated expressions in spike_booting, and 29% (16) were expressed at higher levels in spike_ear_emergence. Therefore, we inferred that *TaSPL* genes are related to

the development of flowers and spikes, influencing the formation of flowers and spikes by participating in the development of anthers and spikes.

Quantitative RT-PCR analysis of 21 *TaSPL* genes

To further investigate the possible functions of *TaSPLs*, qRT-PCR was used to measure the expression patterns of 21 randomly selected *TaSPL* genes in four tissues (roots, leaves, stems, and inflorescences) (Fig. 8A). Twenty-one *TaSPLs* were expressed in the four tissues, with strong tissue-specific expression patterns. The number of *TaSPL* genes that were highly expressed in inflorescences and stems was higher than that in roots and leaves. For example, 8 genes (*TaSPL009*, *-020*, *-034*, *-035*, *-037*, *-041*, *-044*, and *-052*) showed high expression levels in inflorescences, and 8 genes (*TaSPL008*, *-012*, *-016*, *-019*, *-024*, *-027*, *-028*, and *-031*) were expressed in stems; however, only 3 (*TaSPL001*, *-002*, and *-003*) were mainly found in roots, and 2 (*TaSPL005*, and *-011*) showed high expressions in leaves. This suggests that *TaSPL* genes are primarily expressed in inflorescences and stems and play critical roles in the development of wheat flowers and stems.

As *SPL* genes may be associated with plant stress, hormone responses and other signaling processes, the expression levels of 21 *TaSPLs* under 4 treatments were analyzed (Fig. 8B). Compared to CK, 13 genes, such as *TaSPL001*, *-002* and *-019* were upregulated by ABA treatment, suggesting that *TaSPL* genes are involved in the ABA pathway. The expression levels of more than half of the genes decreased gradually under the other 3 treatments. For example, 14 genes, such as *TaSPL009*, *-005*, and *-008* was downregulated by GA treatment; 19 genes, such as *TaSPL011*, *-012*, *-016*, and *-019* was downregulated by salt treatment; and 14 genes were downregulated by PEG treatment. These results indicate that *TaSPL* genes play negative roles under salt, GA and PEG treatments.

Discussion

Characteristics of *SPL* family genes in wheat

The *SPL* genes encode plant-specific zinc finger proteins play critical roles in regulating plant growth and development. These genes have been identified in many green plants, such as rice [27], *A. thaliana* [10], maize [28], and buckwheat [29]. Wheat is a major crop, and the complete genome sequencing of Chinese Spring would provide detailed information to study *SPL* gene families. In this study, we used comparative genomics methods to examine 56 *TaSPLs*, 13 *TuSPLs*, 17 *AetSPLs*, and 17 *BdSPLs*. The number of *SPL* genes in *T.urartu*, *A. tauschii* and *B. distachyon* is similar to that in *A. thaliana* (16) [10], and rice (19) [27] but lower than that in maize (31) [28] and buckwheat (24) [29]. However, the

number of *TaSPLs* is higher than that of the above species; in particular, there are threefold as many *TaSPLs* than *TuSPLs* and *AetSPLs*. This relationship of *TaSPL* with *TuSPL* and *AetSPL* is consistent with previous studies on the *Prx* and *PK* gene family in wheat, *T.urartu*, and *A. tauschii* [30]. It is that wheat is allohexaploid, whose origin involved two polyploidization events, and it appears that many gene duplication events occurred in the wheat genome, allowing the number of *TaSPLs* to increase compared to other species. All *TaSPL* proteins were found to localize to the nucleus; however, during the evolution process of the *TaSPL* gene family, the insertion or loss of exon, and functional diversity and subfunctionalization of *SPL* proteins in different subfamilies caused a large difference in amino acid length and type. Therefore, the isoelectric points and MWs of *TaSPL* proteins are significantly different. This result agrees with that for other grass crops, indicating that the *SPL* genes are conserved among species.

Evolution of *SPL* family genes in wheat

It is suggested that *SPL* genes existed only in green plants and predated the divergence of the green algae [8]. The ancestor *SPL* originally formed into two different lineages in land plants, named clade I and clade II. The clade I has conserved structure characteristics that possess more exons and longer protein sequences [31]. Our results showed that all *SPL* genes from wheat contained 1–11 exons. The number of exons in most *SPL* genes was 3, with only two subgroups (VIII and X) having more than 10 exons. Furthermore, the protein sequences of *TaSPLs* in subgroup VIII and X are greater than *TaSPL* proteins in the remaining subgroups. We inferred that *TaSPL* genes in subgroup X and VIII belong to clade I and that the remaining *TaSPL* genes belong to clade II. In accordance with Zhang's [32] study, we found that the means of K_a and K_s for subgroup X and VIII were lower than those for the other subgroups. Further analysis revealed that subgroup VIII and X genes have lower K_a/K_s ratio (0.25) compared to other duplicated genes. We consider that the evolutionary mechanism of the *TaSPL* gene family is similar to that in other land plants can be clustered into two clades and the subgroup VIII and X genes have evolved slower than other genes.

To explore the evolutionary and functional divergence of the *SPL* gene family in wheat, we identified and determined the phylogenetic relationships of this family. Seventeen orthologous gene pairs of *TaSPL* were detected in this phylogenetic tree, including 13 gene pairs from wheat orthologous with *A. tauschii*, 3 gene pairs orthologous with *T.urartu*, and 1 gene pair orthologous with barley. No gene pair orthologous with rice, *A. thaliana* or *B. distachyon*, suggesting that *TaSPLs* have a

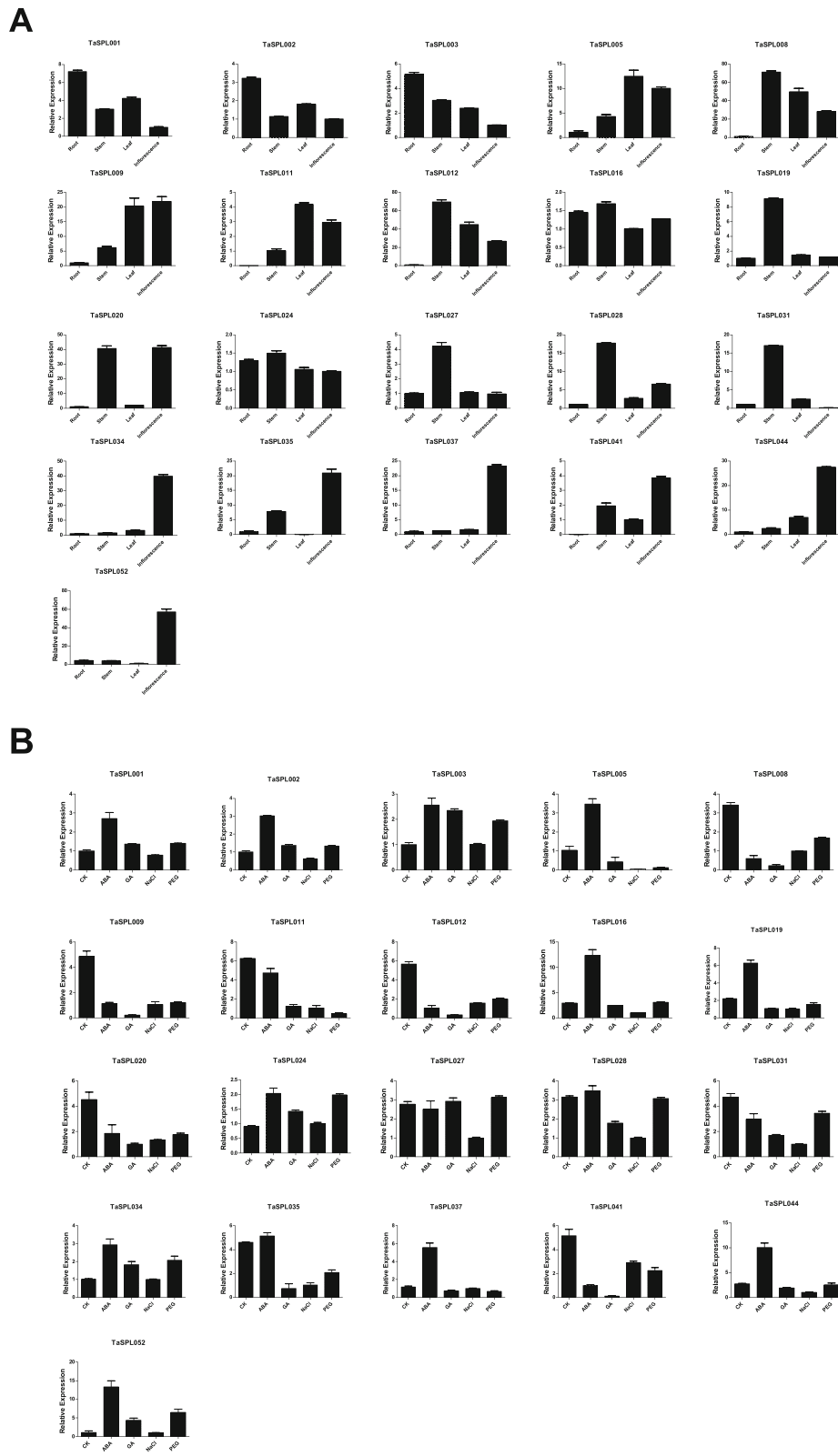


Fig. 8 Quantitative RT-PCR analysis of 21 *TaSPL* genes. Relative expression levels of 21 genes in 4 tissues (roots, stems, leaves and inflorescence) (a); relative expression levels of 21 genes under 4 different treatments (ABA, GA, NaCl and PEG) (b)

close relationship with *SPL* genes in *A. tauschii* and *T. urartu*. From gene duplication results, 81 *TaSPL* gene pairs are duplicated genes and the high levels of collinear gene pairs were observed among rice, maize, *A. tauschii*, and *B. distachyon*. The number of segmental duplication events was greater than that of tandem duplication events, indicating that segmental duplication contributed to *TaSPL* genes expansion. Gene duplication and syntenic analysis confirmed that no positive selection occurred in *TaSPLs*, additionally, *TaSPL* genes underwent strong purifying selection. Our results are consistent with the study of *SPL* genes in buckwheat [29] and rice [31], indicating that the evolution of *TaSPL* genes is comparable with that of other plants.

In brief, we inferred that the evolution of the *SPL* gene family in wheat is similar to that in other land plants. Compared with *A. thaliana* and *B. distachyon*, *TaSPL* genes share a strong relationship with *AetSPLs* and *TuSPLs*.

Expression patterns and potential function of the *SPL* gene family in wheat

In general, genes perform a function based on their expression in the organism, and the expression patterns of genes reflect the gene functions [13]. In this study, we examined the expression patterns of 21 *TaSPL* genes, with 38% showing high expression in stems and inflorescences, implying *TaSPL* genes mainly perform functions in wheat inflorescences and stems. In *A. thaliana*, Chao [33] reported that the *AtSPL1* and the *AtSPL12* showed higher expression levels in inflorescences, overexpressed the two genes enhanced the inflorescence thermotolerance. Xu [10] found that *AtSPL2*, -9, -10, -11, -13, and -15 play dominant roles in the transformation of plants from vegetative to reproductive stages. Except *AtSPL10*, the remaining 5 genes promote both floral meristem identity and floral induction. Homologous genes may share similar functions. The results of evolutionary analysis showed that *AtSPL1* and *AtSPL12* are highly orthologous to *TaSPLs* in the VIII subgroup, including, *TaSPL034*; *AtSPL9*, and -15 share close relationships with *TaSPL044* and -035. *AtSPL13* is orthologous to *TaSPL* genes belonging to subfamily IV, such as *TaSPL020*, -037 and *TaSPL052*. These 6 *TaSPL* genes were expressed at higher levels in inflorescences. We inferred that these genes are involved in the development of inflorescences. Three genes (*TaSPL001*, -002, -003) containing 10 exons were expressed at higher levels in wheat roots than in three other tissues. However, the remaining genes, containing fewer or more than 10 exons, have lower expression in roots. Therefore, we inferred that these 3 *TaSPL* genes are essential for the development of roots and the insertion of exons will affect *SPL* gene expressions in wheat tissues.

Function analyses have uncovered that the *SPL* gene family governs many aspects of plant growth and development in environmental stresses [32]. In this study, we found that expression of *SPL* genes was low in the seedling stage under 4 treatments. Wu [34, 35] and Wang [36] reported that miR156 presents the highest expression levels at seedling stage. As the plant progresses from a seedling to more mature stages, its expression gradually decreases. In contrast, the expression levels of the target genes *AtSPL3* and *AtSPL9* were lower in the seedling stage and gradually increased throughout the vegetative growth stage. Therefore, we deduced that miR156 inhibits the expression of *TaSPL* genes of the seedlings stage. In our study, we found 21 *TaSPL* genes were up- or downregulated by 4 different treatment, with 13 *TaSPL* genes being upregulated by ABA, most of them contained the ABRE element in their promoters, suggesting a positive role in ABA-mediated plant growth. Fourteen *TaSPL* genes were downregulated by GA, suggesting that these 14 genes play a negative role in GA-mediated plant growth. Cui [37] reported that *AtSPL9* was downregulated by salt and drought treatments, regulating *A. thaliana* flowering, implying that *AtSPL9* responds to these two stress treatment. We found that *TaSPL035* and -044, which were downregulated by NaCl and PEG treatments, share high homology with *AtSPL9*, suggesting that these two genes can respond to salt and drought stress. In brief, we inferred that the *TaSPL* gene family plays critical roles in wheat development, especially, this family is involved in stems and flowers development, and that members of the family can participate in abiotic stress.

Conclusions

Phylogeny and diversification of *SPL* genes in wheat were investigated from different levels, including gene structures, evolutionary relationships, synteny analyze, GO annotation, protein-protein interaction, and expression patterns. The *SPL* gene family in wheat was expanded by segmental duplication and purifying selection, during the evolution. All 56 *TaSPL* genes were divided into 8 subgroups, genes in same subgroup share similar evolutionary features and expression patterns, implying similarity function potentially of *TaSPL* genes. Twenty-one gene expressions were detected and that the *TaSPL* genes have tissue-specific expression patterns in different tissues and these genes exhibited higher expression levels at stems and other tissues related to spike development, such as anthers and inflorescence, suggesting that *TaSPL* genes regulate the development of stems and influence the formation of flowers and spikelets. Here, we first reported the identification, phylogenetic, and expression analysis of the *TaSPL* gene family, which will serve as a foundation for further elucidation of the potential biological functions of *SPLs* in plants.

Methods

Identification of *SPL* genes in wheat

Whole-genome data of wheat were obtained from the Ensembl plant database (<http://plants.ensembl.org/info/website/ftp/index.html>), and the SBP domain (PF03110) was downloaded from the PFAM database (<https://pfam.xfam.org/>). Then, the SBP domain was used as the query sequence to find proteins containing an SBP domain in wheat using the HMMER 3.0 program with the threshold of $e < 1e^{-5}$ [38]. The SBP protein sequences from rice [27] and *A. thaliana* [10] were used as query sequences to search against the wheat protein dataset using the BLASTP program, and the threshold was set at $e < 1e^{-5}$ and 50% identity. Putative wheat *SPL* genes were preliminarily identified by analyzing the results from HMM and BLASTP. Then, the NCBI-CDD web server (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) and the PFAM database were used to further confirm the candidate *SPL* genes of wheat. The protein sequences of *TaSPLs* were computed in the ExPASy server (https://web.expasy.org/compute_pi/) [39] to obtain the theoretical isoelectric point (PI) and molecular weight (MW), and the cello web server (<http://cello.life.nctu.edu.tw/>) [40] was used to predict the subcellular localization of these proteins.

Chromosomal location, gene structure and conserved motif analysis of *TaSPL* genes

Physical locations of *TaSPL* genes were obtained from the genome annotation information (gff3) of wheat, and the *TaSPL* genes were mapped to the wheat chromosomes. The structure of genes and coding sequences (CDS) of *TaSPLs* were analyzed to investigate the exon-intron organization, and the results were displayed in Gene Structure Display Serve (<http://gsds.cbi.pku.edu.cn/>) [41]. Conserved motifs of *TaSPL* proteins were identified using MEME v4.9.0 (<http://meme-suite.org/tools/meme>) with the optimum motif set at ≥ 10 and ≤ 200 amino acids and the maximum number of motif set at 20 [42].

Phylogenetic analyses

SPL protein sequences of wheat, *B. distachyon*, *A. tauschii*, *T. urartu*, rice, and *A. thaliana* were used to perform the phylogenetic analysis. ClustalX 2.0 software [43] with the default parameters was utilized to conduct multiple sequence alignment. An un-rooted neighbor joining (NJ) tree was constructed using MEGA 6.0 [44] software with 1000 bootstrap replications.

Gene duplication and synteny analysis of wheat with rice, maize, *A. tauschii*, and *B. distachyon*

MCSanX [45] software was used to detect collinear regions between *TaSPL* genes, as well as collinear blocks

of *TaSPL* genes with 4 other species (*B. distachyon*, rice, *A. tauschii*, and maize). Gene duplication events of *TaSPLs* and synteny relationships between the aforementioned species were visualized using the Circos 0.67 tool [46]. CDSs and protein sequences of collinear gene pairs were compared, and KaKs ratios were calculated using KaKs_Calculator software [47]. Finally, the divergence time of collinear gene pairs was calculated using the method of Wang [$T = Ks / (2\lambda \times 10^{-6})$ Mya ($\lambda = 6.5 \times 10^{-9}$)] [3].

GO annotation and protein-protein interaction network analysis of *TaSPLs*

GO annotation of *SPL* proteins in wheat was available from the PLAZE database (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_monocots/) [48] and the Plant Transcriptional Regulatory Map database (<http://plantregmap.gao-lab.org/>). The GO annotation results were visualized using the WEGO (<http://wego.genomics.org.cn/>) online tool [49].

Based on orthogonal genes between wheat and *A. thaliana*, the ArenaNet V2 tool [50] and the String database (<https://string-db.org/cgi/input.pl>) [1] were used to construct an interaction network between *TaSPL* proteins and other wheat proteins. A trusted value of ≥ 0.8 in the String database was used to confirm the interaction network, which was graphically displayed by Cytoscape software [51].

Promoter analysis of *TaSPLs*

The upstream 1.5-kb DNA sequences of the *TaSPL* genes were downloaded from the Ensembl plant database, and then submitted to the PLACE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) to predict cis-regulatory elements in promoter regions [40].

Gene expression, materials and qRT-PCR analysis of *TaSPLs*

The transcriptional data of *TaSPL* genes in 3 tissues (roots, shoots/leaves, spike) and 5 organs (ovary_early_anthesis, ovary_late_anthesis, spike_booting, spike_ear_emergence, and rachis_ear_emergence) of wheat were obtained from wheat expression browser website (<http://www.wheat-expression.com/download>) to detect the expression profile [52]. R was used to display the expression patterns in a heat map.

For qRT-PCR, the Chinese Spring cultivar of wheat was grown in a greenhouse in 2019. Roots, leaves, stems, and inflorescences were collected during the heading stage. Three-week-old seedlings were exposed to 200 mM NaCl, 100 μ M gibberellin acids (GA), 100 μ M abscisic acid (ABA), or 20% PEG-600 for 2 h, and samples were then collected. All samples were stored at -80°C for RNA extraction using RNAiso Reagent (TaKaRa,

Beijing, China) with three biological replicates. cDNA was synthesized using the RT Master Mix Perfect Real-Time kit (TaKaRa, Beijing, China). QuantStudio™ Real-Time PCR software was used to perform qRT-PCR, and the $2^{(-\Delta\Delta Ct)}$ analysis method was used to determine the relative expression levels of 21 randomly selected *TaSPLs*. We chose *TraesCS6B02G243700* as the reference gene to normalize the expression levels of the *TaSPL* genes. The primers used are listed in Additional files 1: Table S2.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12870-020-02576-0>.

Additional file 1: Table S1. Characteristic features of the *SPL* gene family identified in wheat. **Table S2.** Primers used for qRT-PCR. **Table S3.** Ka/Ks ratios and estimated divergence time for duplicated *TaSPL* gene pairs. **Table S4.** Ka/Ks ratios and estimated divergence time for orthologous *TaSPL* genes between wheat and rice. **Table S5.** Ka/Ks ratios and estimated divergence time for orthologous *TaSPL* genes between wheat and maize. **Table S6.** Ka/Ks ratios and estimated divergence time for orthologous *TaSPL* genes between wheat and *A. tauschii*. **Table S7.** Ka/Ks ratios and estimated divergence time for orthologous *TaSPL* genes between wheat and *B. distachyon*. **Table S8.** GO annotations of *TaSPL* proteins. **Table S9.** The protein-protein interaction between *TaSPLs* and other proteins in wheat. **Table S10.** Cis-elements contained in the *TaSPL* genes promoter region. **Figure S1.** Cis-regulatory elements in the promoter region of *TaSPL* genes.

Abbreviations

TaSPL: Wheat SPL; ABA: Abscisic acid; GA: Gibberellic acid; CDS: Coding sequence; HMM: Hidden Markov Model; qRT-PCR: Quantitative real-time polymerase chain reaction

Acknowledgements

We are very grateful to Professor Ma and professor Ding for their guidance. We thank the Instrument sharing platform of Northwest Agricultural and Forestry University and lab members for their assistance in this study. We thank the English language editors of Nature Research Editing Service.

Authors' contributions

TZ arranged and designed the study, analyzed the data, and wrote the manuscript. YL and LTM completed the visualization of the data. XYW, DZZ and YCH performed the experiments. QD and LJM supervised the research. All authors have read and approved the manuscript.

Funding

This study was supported by the National Key Research and Development Program of China (No. 2016YFD0101602). The funding body was not involved in the design of the study, analysis or interpretation of data or writing the manuscript.

Availability of data and materials

The Chinese Spring cultivar used in the experiment is supplied by Professor Ma Lingjian of Northwest Agricultural and Forestry University. The datasets supporting the results of this article are included in the article and Additional files.

Ethics approval and consent to participate

The Chinese Spring cultivar used in the experiment is supplied by Professor Ma Lingjian of Northwest Agricultural and Forestry University. These plant materials are widely used globally. This article does not contain any studies with human participants or animals and did not involve any endangered or protected species.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Agronomy, Northwest A&F University, Yangling 712100, China. ²College of Life Science, Wuhan University, Wuhan 430072, China. ³College of Horticulture, Northwest A&F University, Yangling 712100, China.

Received: 3 March 2020 Accepted: 26 July 2020

Published online: 11 September 2020

References

- Guo XJ, Wang JR. Global identification, structural analysis and expression characterization of bHLH transcription factors in wheat. *BMC Plant Biol.* 2017;17(1):90.
- Zheng A, Sun F, Cheng T, Wang Y, Xie K, Zhang C, Xi Y. Genome-wide identification of members of the TCP gene family in switchgrass (*Panicum virgatum* L.) and analysis of their expression. *Gene.* 2019;702:89–98.
- Wang Y, Wang Q, Zhao Y, Han G, Zhu S. Systematic analysis of maize class III peroxidase gene family reveals a conserved subfamily involved in abiotic stress response. *Gene.* 2015;566(1):95–108.
- Huijser P, Klein J, Lonig W, Meijer H, Saedler H, Sommer H. Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MAD5-box gene *squamosa* in *Antirrhinum majus*. *EMBO J.* 1992;11(4):1239–49.
- Yang Z, Wang X, Gu S, Hu Z, Xu H, Xu C. Comparative study of SBP-box gene family in Arabidopsis and rice. *Gene.* 2008;407(1–2):1–11.
- Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Aoki M, Seki E, Matsuda T, Nunokawa E, et al. A novel zinc-binding motif revealed by solution structures of DNA-binding domains of Arabidopsis SBP-family transcription factors. *J Mol Biol.* 2004;337(1):49–63.
- 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(3):585–96.
- Guo AY, Zhu QH, Gu X, Ge S, Yang J, Luo J. Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family. *Gene.* 2008;418(1–2):1–8.
- Cardon G, Hohmann S, Klein J, Nettessheim K, Saedler H, Huijser P. Molecular characterisation of the Arabidopsis SBP-box genes. *Gene.* 1999;237(1):91–104.
- Xu M, Hu T, Zhao J, Park MY, Earley KW, Wu G, Yang L, Poethig RS. Developmental functions of miR156-regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) genes in Arabidopsis thaliana. *PLoS Genet.* 2016;12(8):e1006263.
- Tripathi RK, Goel R, Kumari S, Dahuja A. Genomic organization, phylogenetic comparison, and expression profiles of the SPL family genes and their regulation in soybean. *Dev Genes Evol.* 2017;227(2):101–19.
- Li C, Lu S. Molecular characterization of the SPL gene family in *Populus trichocarpa*. *BMC Plant Biol.* 2014;14:131.
- Zhou Q, Zhang S, Chen F, Liu B, Wu L, Li F, Zhang J, Bao M, Liu G. Genome-wide identification and characterization of the SBP-box gene family in *Petunia*. *BMC Genomics.* 2018;19(1):193.
- Zhang X, Dou L, Pang C, Song M, Wei H, Fan S, Wang C, Yu S. Genomic organization, differential expression, and functional analysis of the SPL gene family in *Gossypium hirsutum*. *Mol Genet Genom.* 2015;290(1):115–26.
- Han YY, Ma YQ, Li DZ, Yao JW, Xu ZQ. Characterization and phylogenetic analysis of fifteen NtabSPL genes in *Nicotiana tabacum* L. cv. Qinyan95. *Dev Genes Evol.* 2016;226(1):1–14.
- Zhang L, Wu B, Zhao D, Li C, Shao F, Lu S. Genome-wide analysis and molecular dissection of the SPL gene family in *Salvia miltiorrhiza*. *J Integr Plant Biol.* 2014;56(1):38–50.
- Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP. Prediction of plant microRNA targets. *Cell.* 2002;110(4):513–20.
- Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D. Specific effects of microRNAs on the plant transcriptome. *Dev Cell.* 2005;8(4):517–27.
- Gandikota M, Birkenbihl RP, Hohmann S, Cardon GH, Saedler H, Huijser P. The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. *Plant J.* 2007;49(4):683–93.

20. Salinas M, Xing S, Hohmann S, Berndtgen R, Huijser P. Genomic organization, phylogenetic comparison and differential expression of the SBP-box family of transcription factors in tomato. *Planta*. 2012;235(6):1171–84.
21. Zhang Y, Schwarz S, Saedler H, Huijser P. SPL8, a local regulator in a subset of gibberellin-mediated developmental processes in Arabidopsis. *Plant Mol Biol*. 2007;63(3):429–39.
22. Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, Zeng R, Zhu H, Dong G, Qian Q, et al. Control of grain size, shape and quality by OsSPL16 in rice. *Nat Genet*. 2012;44(8):950–4.
23. Wang L, Zhang Q. Boosting Rice yield by fine-tuning SPL gene expression. *Trends Plant Sci*. 2017;22(8):643–6.
24. Hou H, Li J, Gao M, Singer SD, Wang H, Mao L, Fei Z, Wang X. Genomic organization, phylogenetic comparison and differential expression of the SBP-box family genes in grape. *PLoS One*. 2013;8(3):e59358.
25. Ning K, Chen S, Huang H, Jiang J, Yuan H, Li H. Molecular characterization and expression analysis of the SPL gene family with BpSPL9 transgenic lines found to confer tolerance to abiotic stress in *Betula platyphylla* Suk. *Plant Cell Tissue Organ Culture (PCTOC)*. 2017;130(3):469–81.
26. Tripathi RK, Bregitzer P, Singh J. Genome-wide analysis of the SPL/miR156 module and its interaction with the AP2/miR172 unit in barley. *Sci Rep*. 2018;8(1):7085.
27. Xie K, Wu C, Xiong L. Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol*. 2006;142(1):280–93.
28. Peng X, Wang Q, Zhao Y, Li X, Ma Q. Comparative genome analysis of the SPL gene family reveals novel evolutionary features in maize. *Genet Mol Biol*. 2019;42(2):380–94.
29. Liu M, Sun W, Ma Z, Huang L, Wu Q, Tang Z, Bu T, Li C, Chen H. Genome-wide identification of the SPL gene family in Tartary buckwheat (*Fagopyrum tataricum*) and expression analysis during fruit development stages. *BMC Plant Biol*. 2019;19(1):299.
30. Yan J, Su P, Li W, Xiao G, Zhao Y, Ma X, Wang H, Nevo E, Kong L. Genome-wide and evolutionary analysis of the class III peroxidase gene family in wheat and *Aegilops tauschii* reveals that some members are involved in stress responses. *BMC Genomics*. 2019;20(1):666.
31. Zhong H, Kong W, Gong Z, Fang X, Deng X, Liu C, Li Y. Evolutionary analyses reveal diverged patterns of SQUAMOSA promoter binding protein-like (SPL) gene family in *Oryza* genus. *Front Plant Sci*. 2019;10:565.
32. Zhang SD, Ling LZ, Yi TS. Evolution and divergence of SBP-box genes in land plants. *BMC Genomics*. 2015;16:787.
33. Chao LM, Liu YQ, Chen DY, Xue XY, Mao YB, Chen XY. Arabidopsis transcription factors SPL1 and SPL12 confer plant Thermotolerance at reproductive stage. *Mol Plant*. 2017;10(5):735–48.
34. Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS. The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. *Cell*. 2009;138(4):750–9.
35. Wu G, Poethig RS. Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development (Cambridge, England)*. 2006;133(18):3539–47.
36. Wang JW, Czech B, Weigel D. miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. *Cell*. 2009;138(4):738–49.
37. Cui LG, Shan JX, Shi M, Gao JP, Lin HX. The miR156-SPL9-DFR pathway coordinates the relationship between development and abiotic stress tolerance in plants. *Plant J*. 2014;80(6):1108–17.
38. Cui L, Feng K, Wang M, Wang M, Deng P, Song W, Nie X. Genome-wide identification, phylogeny and expression analysis of AP2/ERF transcription factors family in *Brachypodium distachyon*. *BMC Genomics*. 2016;17(1):636.
39. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res*. 2012;40(Web Server issue):W597–603.
40. Zhu T, Xin F, Wei S, Liu Y, Han Y, Xie J, Ding Q, Ma L. Genome-wide identification, phylogeny and expression profiling of class III peroxidases gene family in *Brachypodium distachyon*. *Gene*. 2019;700:149–62.
41. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDB 2.0: an upgraded gene feature visualization server. *Bioinformatics (Oxford, England)*. 2015;31(8):1296–7.
42. Wang J, Hu Z, Zhao T, Yang Y, Chen T, Yang M, Yu W, Zhang B. Genome-wide analysis of bHLH transcription factor and involvement in the infection by yellow leaf curl virus in tomato (*Solanum lycopersicum*). *BMC Genomics*. 2015;16:39.
43. Larkin MA, Blackshields G, Brown NP, Chenna R, PA MG, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*. 2007;23(21):2947–8.
44. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725–9.
45. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40(7):e49.
46. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639–45.
47. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*. 2010;8(1):77–80.
48. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res*. 2017;45(D1):D1040–d1045.
49. Ye J, Zhang Y, Cui H, Liu J, Wu Y, Cheng Y, Xu H, Huang X, Li S, Zhou A, et al. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res*. 2018;46(W1):W71–w75.
50. Lee T, Yang S, Kim E, Ko Y, Hwang S, Shin J, Shim JE, Shim H, Kim H, Kim C, et al. AraNet v2: an improved database of co-functional gene networks for the study of Arabidopsis thaliana and 27 other nonmodel plant species. *Nucleic Acids Res*. 2015;43(Database issue):D996–1002.
51. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
52. Ramirez-Gonzalez RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, et al. The transcriptional landscape of polyploid wheat. *Science (New York, NY)*. 2018;361:6403.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

