Genome-Wide Identification, Characterization, and Expression Profiling Analysis of SPL Gene Family During The Inflorescence Development in *Trifolium Repens*

Jieyu Ma
Sichuan Agricultural University
Gang Nie
Sichuan Agricultural University
Zhongfu Yang
Sichuan Agricultural University
Sainan Ma
Sichuan Agricultural University
Jinwan Fan
Sichuan Agricultural University
Ruchang Hu
Sichuan Agricultural University
Feifei Wu
Sichuan Agricultural University
Xinquan Zhang (zhangxq@sicau.edu.cn)
Sichuan Agricultural University

Research Article

Keywords: Trifolium repens, SPL gene family, Phylogenetics, Flowering regulation, Expression patterns

Posted Date: March 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1430343/v1

Abstract

Trifolium repens is the most widely cultivated perennial legume forage in temperate region around the world. It has rich nutritional value and good palatability, seasonal complementarity with grasses, and can improve the feed intake and digestibility of livestock. However, flowering time and inflorescence development directly affects the quality and yield of *T. repens*, as well as seed production. Squa promoter binding protein–like (*SPL)* gene family is a plant specific transcription factor family, which has been proved to play a critical role in regulating plant formation time and development of flowers. In this study, a total of 37 *TrSPL* genes were identified from the whole genome of *T. repens* and were divided into 9 clades based on phylogenetic tree. The conserved motif of squamosa promoter binding protein (SBP) contains two zinc finger structures and one NLS structure. Gene structure analysis showed that all *TrSPL* genes contained SBP domain, while ankyrin repeat region was just distributed in part of genes. 37 *TrSPL* genes were relatively dispersedly distributed on 16 chromosomes, and 5 pairs of segmental repeat genes were found, which indicated that segmental duplication was the main way of gene expansion. Furthermore, the gene expression profiling showed that *TrSPL11, TrSPL13, TrSPL22* and *TrSPL26* were highly expressed only in the early stage of inflorescence development, while *TrSPL1* and *TrSPL6* are highly expressed only in the mature inflorescence. Significantly, the expression of *TrSPL4* and *TrSPL12* increased gradually with the development of inflorescences. The results of this study will provide valuable clues for candidate gene selection and elucidating the molecular mechanism of *T. repens* flowering regulation.

1 Introduction

Trifolium repens is the most important perennial legume forage in temperate regions. It is mainly cultivated in perennial pasture together with other forage, and utilization through directly grazed or mechanically harvested into hay and silage (Birte et al., 2014). *T. repens* was a kind of high quality forage with rich protein and mineral content, good palatability and high nitrogen fixation ability, which was beneficial to improve grassland quality, complement seasonal growth patterns of commonly forages and promote the intake and digestibility of livestock (Caradus et al., 1995; Abberton et al., 2005). However, flowering always reduces the number of axillary buds growing into branch stolons and leaf production (Gibson et al., 1985), which lead to sharply dropped of the nutritional value and digestibility (Kilcher et al., 1981). Therefore, delaying the flowering time and prolonging the vegetative growth period of *T. repens* will greatly improve the forage quality and yield (Fiorella et al., 2017). Also, *T. repens* flowering at the suitable time and inflorescence well developed are not only an important guarantee for seed production, but also increasing the opportunity of resowing in the pasture under adverse conditions and enhancing the persistence (Pederson et al., 2000; Christian et al., 2017). Indeed, it is important to study the flowering time regulation and inflorescence development for furture *T. repens* genetic improvement and breeding program, nevertheless, the molecular mechanism remain unknow due to the absence of omics data and limited availability of genome resource. The newly published genome article provides a possibility platform for the further study of *T. repens* (Griffiths et al., 2019).

Squa Promoter binding protein–Like (SPL) proteins are plant specific transcription factor family and operate on the characteristic genes of flower meristem, participate in the formation and later development of flowers. The most leading feature of SPL proteins is that SBP-box (squamosa promoter binding protein) encodes a conserved protein domain with 76 amino acids (Cardon et al., 1999). The conserved SBP domain contains two zinc finger structures and one NLS structure, which are the structural basis for sequence-specific gene binding (Yamasaki et al., 2004; Birkenbihl et al., 2005). Although *SPL* gene family had been widely studied in lots of species and the function of this kind of genes was also confirmed, there is no systematic exploration in *T. repens*. In Snapdragon (*Antirrhinum majus*), *SBP1* and *SBP2* acted on the promoter region of flower meristem characteristic gene *SQUAMOSA* to control early flower development (Klein et al., 1996). A total of 16 *AtSPL* genes were identified in *Arabidopsis thaliana*, and *AtSPL3* directly activates *LEAF*, *FRUITFULL* and *APETALA1* to control the formation time of flowers (Ayako et al., 2009). Besides, overexpressing *AtSPL10* showed an early flowering phenotype, and the triple loss-of-function mutants with its homologous genes *AtSPL2* and *AtSPL11* showed flowering later than wild type (Tao et al., 2019). In terms of inflorescence development, *AtSPL8* showed significant functions in controlling pollen sac development and it is necessary for the normal development of spore tissue (Unte et al., 2003). The loss of *AtSPL9* and *AtSPL15* function can lead to the change of inflorescence structure (Schwarz et al., 2008). In rice (*Oryza sativa*), 19 *OsSPL* genes were identified (Xie et al., 2006), with function of promoting the grain development to increase yield at the optimal expression level, and preventing the reversal of rice inflorescence during flowering (Lei et al., 2017; Francesca et al., 2021). There were 56 *TaSPL* genes in wheat (*Triticum aestivum*), and most of which regulated the development of inflorescence and spike (Ting et al., 2020). In addition, 17 putative *DgSPL* genes were identified in orchardgrass (*Dactylis glomerata*), a common companion species of *T. repens* mixed grassland, which were mainly involved in vegetative to reproductive growth transition, flower development and flowering regulation (Guangyan et al., 2021). In Tribulus alfalfa (*Medicago truncatula*), a total of 23 *MtSPL* genes were identified which are involved in the development of seed pods, especially the formation of thorns on pods (Wang et al., 2019).

It is critical to explore the white clover flowering time regulation and inflorescence development mechanism for its genetic improvement and utilization. Genome-wide identification, characterization, and expression profiling analysis of *SPL* gene family in *T. repens* will be an important basic work for providing valuable clues of candidate gene selection and breeding program. In this study, *TrSPL* gene family were identified and comprehensively analyzed, including SBP conserved domain comparation, conserved motif composition, gene structure annotation and chromosome location distribution. The evolution of *TrSPL* genes were preliminarily predicted by studying the intraspecific replication events, phylogenetic analysis and collinearity analysis with other plant species. Also, an expression profiling analysis of *TrSPL* genes at different flower development stages were constructed.

2 Materials And Methods

2.1 Genome-wide identification of *TrSPL* genes

The *T. repens* genome resource information came from previous study (Griffiths et al., 2019) and all files were provided by Stig Uggerhøj Andersen from Aarhus University. The hidden markov model (HMM) file of SBP domain (pf03110) was download from Pfam database (Finn et al., 2014) (pfam.xfam.org/). Take the downloaded HMM file as the query sequence to search the protein sequence data of *T. repens* by using HMMER3.0. The obtained proteins were aligned by ClustaW (E-value < 1^{e-20}) and the SBP HMM file was rebuilt by hmmbuild in HMMER 3.0. Finally, SBP HMM of *T. repens* was used to identify SPL protein in *T. repens* genome, and the cut-off value was setted to 0.01(Finn et al., 2011). In order to ensure that all candidate genes contain SBP domain, the NCBI Conserved Domain Search website was used for further confirmation (Conserved Domains Database (CDD) and Resources (nih.gov)). The isoelectric point and relative molecular weight data were obtained using the Expasy website (Compute pI/MW - SIB Swiss Institute of Bioinformatics | Expasy). The subcellular localization information of *TrSPL* genes were analyzed and predicted online by BUSCA tool (BUSCA - Bologna Biocomputing Group (unibo.it)). Cis acting element analysis was predicted and visualized by using online tools PlantCARE and TBtools software (http://bioinformatics. psb.ugent.be/webtools/plantcare/html/) (Chen et al., 2020).

2.2. Phylogenetic analyses and classification of *TrSPL* genes

Genes and protein sequences of *Arabidopsis thaliana*, *Medicago truncatula*, *Trifolium pratense and Glycine max SPL* gene family are downloaded from TAIR (https://www.arabidopsis.org/down load/index-auto.jsp?dir=%2Fdownload_files%2FSequences%2FAraport11_blastsets), the *Medicago truncatula* Genome Database (http://www.medicagogenome.org/), Ensembl website (https://plants. ensembl. org/index.html) and the *Glycine max* Genome Database (https://www. soybase.org/), respectively. Multiple sequence alignment of SPL protein sequences was performed by Log-Expectation (MUSCLE) in MEGA v6.0 (Tamura et al., 2013). Eventually, the phylogenetic tree was constructed by the neighbor-joining method of MEGA v6.0 software and the bootstrap replications value was 1000.0.

2.3. Gene Structure and Motif Analysis of *TrSPL* genes

Jalview software for SBP conservative sequence alignment of *T. repens* (http://www.jalview.org/), two zinc finger structures and one NLS structure are marked in the figure. The conserved motifs of *T. repens* are analyzed by MEME Suite 5.14 (Introduction - MEME Suite (meme-suite.org)) (Bailey et al., 2009), and the details of each conserved motif are also derived from the same website. The maximum number of motifs are set as 10. Use the program Visualize MEME/MAST Motif Pattern of Tbtools software for conservative motif visual analysis. Gene structure analysis uses the gene sequence file of *TrSPL* genes to analyze and visualize in TBtools software Visualize Gene Structure function, showing the CDS sequence, SBP conserved domain, ANK conserved domain and intron.

2.4. Chromosomal locations and synteny analysis of *TrSPL* genes

The chromosome location information of *TrSPL* genes were obtained from the genome annotation file. The Gene Location Visualize From GTF/GFF function of Tbtools software was used for gene chromosome mapping and visual analysis. The intraspecific collinearity circle map of *TrSPL* genes were analyzed by using the Dual Synergy Plot for McscanX function of TBtools and visualized by the Advanced Circos function. The genome data of *Arabidopsis thaliana* was downloaded from TAIR (https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FSequences%2FAraport11_blastsets). The genome data of *Medicago truncatula* was downloaded from the *Medicago truncatula* Genome Database (http:// www.medicagogenome.org/). The genome data of *Trifolium pratense* was downloaded from Ensembl website (https://plants. ensembl. org/ index.html). The genome data of *Glycine max* was downloaded from the *Glycine max* Genome Database (https://www.soybase.org/). Drawing with Dual Synteny Ploter function of TBtools.

2.5. Material culture, sampling and qRT-PCR

T. repens Seed Haifa is provided by Beijing MAMMOTH SEED company, with the registration number of 249. The same genotype material Haifa is used for asexual propagation by cutting in pot (The diameter is 16.5cm, the height is 10cm, and the bottom diameter is 12cm). The growth environment of *T. repens* is a growth chamber with a Photoperiod of 14h at 22 ℃ and a dark period of 10h at 20 ℃. *T. repens* after flowering, fresh inflorescences in different developmental states are taken and stored in the freezing tube. (T1: Immature inflorescence, T2: Inflorescences in which no floret was open, T3: Inflorescences in which outermost circle of florets were open, T4: Inflorescences in which 50% of florets were open, T5: Mature inflorescence). Three biological replicates per sample. Put the fresh sample into liquid nitrogen immediately and long-term preservation in -80 ℃ refrigerator. Total RNA was extracted using the Hipure HP Plant RNA Mini Kit (Magen). The obtained RNA was reverse transcribed into cDNA using the MonScript™ RTIII ALL-in-One Mix with dsDNase (Monad) kit. Primer design used primer5 software and Tr β- Actin as internal parameter (Jia et al., 2021). All primers and internal paramenter information have been given in the attachment (**Table S1**). qRT-PCR was performed with Bio-Rad CFX96 instrument, and used SYBR® green real-time PCR master Mix test kit. The qRT-PCR procedure as follow:30s at 95°C, denaturation (95°C for 10s), anneal/extension (55°C for 30s), for 40 cycles, melting curve detection (65°C-95°C). $2^{(-\Delta\Delta Ct)}$ analysis method was used to calculate the relative expression of 16 *TrSPL* genes, and finally the expression maps of 16 *TrSPL* genes in 5 different flower development stages were obtained.

3 Results

3.1 Genome-wide Identification of *T. repens SPL* Genes

Based on the *T. repens* genome resources, putative *TrSPL* genes were preliminarily identified by performing HMM (hidden Markov model) search (SBP domain, PF03110) from Pfam database. Subsequently, 38 *TrSPL* genes were identified after removing redundant sequences preliminarily. However, owing to the incomplete of SBP structure in gene *chr13.jg763.t1*, 37 genes with highly conserved SBP domain were retained (**Table. 1**). According to the phylogenetic tree of *T. repens* and *Arabidopsis* (Fig. 1),

37 *TrSPL* genes were named as *TrSPL1* - *TrSPL37* respectively (**Table S2**). The isoelectric point (PI) of 40.5% of proteins was less than 7, with the lowest PI of *TrSPL5* was 5.72, and the highest PI of *TrSPL14* and *TrSPL33* were 9.41. The protein sequence length (aa) of all TrSPL proteins ranged from 1053 (*TrSPL6*) to 124 (*TrSPL12*) amino acids with an average of 474. The relative molecular weight (MW) ranged from 116204.16Da to 14269.05Da and the corresponding gene was consistent with the length of protein sequence. Subcellular localization results showed that 37 *SPL* genes in *T. repens* were located in the nucleus (30 genes), chloroplast (2 genes), plasma membrane (2 genes) and endomembrane system (3 genes), respectively. The basic data of *TrSPL* gene family varied widely, which indicated that diverse function of these genes.

3.2 Phylogenetic Analyses and Classification of the *TrSPL* Gene Family

For the classification of the *TrSPL* gene family, a Neighbor-joining (NJ) tree (with 1,000 bootstraps) of 16 *Arabidopsis SPL* genes and 37 identified *TrSPL* genes was constructd (Fig. 1). The results showed that 53 genes were divided into 9 clades, and all 16 *SPL* genes in *Arabidopsis thaliana* were distributed in 8 main *SPL* evolution clades named as to (Salinas et al., 2012; Preston et al., 2013). There were two *TrSPL* genes grouped into *T. repens*-specific clade named in , which indicated that potentially emerged after the divergence between two species. Moreover, in order to explore the evolutionary relationship of *SPL* gene families among the related species of *T. repens*, 24 *SPL* genes of red clover (*Trifolium pratense*) and 23 *SPL* genes of tribulus alfalfa (*Medicago truncatula*) were together analzed with white clover and Arabidopsis (Fig. 2). A total of 100 *SPL* genes are divided into 7 clades (-), and each evolutionary clade containd all 4 species. The results showed that *SPL* gene distance between three legumes species were closer, while farther than *Arabidopsis*. Multi species phylogenetic tree revealed that *SPL* gene family was relatively conservative in evolutionary direction.

Cis-acting elements are important in regulating gene expression. Analysis cis acting elements of 37 *TrSPL* genes (upstream 2000 bp) showed that *TrSPL* genes contained a large number of action elements in response to light (light response element, circadian rhythm regulation element, phytochrome down-regulation response element) and hormones (abscisic acid response element, gibberellin response element, plant auxin response element, salicylic acid response element, methyl jasmonate response element). In addition, there were some other elements in response to external stress, such as low temperature response element, defense and stress response element, plant trauma response element. (**Table S2; Fig S1**).

3.3 Sequence feature and gene structure of *TrSPL* genes

The full-length protein sequence of *T. repens SPL* genes were used for sequence alignment. The SBP domain was highly conserved in the *TrSPL* gene family (Fig. 3). All SBP domains contained two zinc finger structures and a nuclear localization signal (NLS), along with absence of some small fragments from *TrSPL12*, *TrSPL 20*, *TrSPL 34* and *TrSPL 35*. Besides, the motif of first *TrSPL* gene zinc finger for

clade Ⅱ (Cys-Cys-Cys-Cys) was different from that in the other clades (Cys-Cys-Cys-His, which was consistent with *Tribulus alfalfa* and other species (Li et al., 2019; Wang et al., 2019; Tong et al., 2020).

In order to analyze the diversity and similarity of *TrSPL* gene structure, 10 kinds of motifs were identified in the MEME website (Fig. 4). Among them, motif1 and motif2 contained a complete SBP domain and the length of motif ranges from 21 to 50. *TrSPL12* and *TrSPL35* only contained one motif, whereas *TrSPL6* contained 11 motifs. The relatively conserved parts were motif1, motif2 and motif10 in all *TrSPL* genes. Motif 4–7 were only appeared in the yellow clade which reveal that these motifs were the main factors for the evolution and even functional conservation of this branch. *TrSPL* genes in same branch have similar conserved motifs, indicating they may be having similar function. Sequence information for each motif is provided in **Table S3**. In the analysis of gene structure, all the genes in the yellow clade have Ankyrin repeat regions (Ank-2 and Ank-2 superfamily), which could be involved in protein-protein interaction (Riese, Höhmann et al. 2007). All *TrSPL* genes had at least one intron and *TrSPL34* had the most introns (with 20 introns).

3.4 Chromosomal locations and synteny analysis of *TrSPL* genes

37 *TrSPL* genes were accurately mapped onto *T. repens* chromosomes (Fig. 5, **Table S4**). *TrSPL* genes were relatively evenly distributed on all 16 chromosomes, and the number of *TrSPL* gene on each chromosome ranged from one (Chr6O, chr8O, chr5P and chr7P) to four (chr3O, chr2P and chr3P).

Gene duplication event is an important way to produce new genes with similar or different functions. We visualized the intraspecific replication events of *TrSPL* genes in Fig. 6. A total of 5 pairs of segmental duplication genes were found, while there was no tandem duplication in the *TrSPL* gene family (**Table S5**), indicating that segmental duplication was the main way of *TrSPL* gene family expansion.

To further explore the evolution of the *TrSPL* gene family, 4 comparative syntenic maps consisted of *Arabidopsis thaliana*, *Trifolium pratense*, *Medicago truncatula*, and *Glycine max* were constructed based on collinearity analysis (Fig. 7). The number of homologous pairs between *T. repens* and other 4 species was 10 (*Arabidopsis*), 14 (Red clover), 28 (Tribulus alfalfa) and 42 (soybean). The details of homologous pairs are given in **Table S6**. The comparison results showed that there are more homologous genes between *T. repens* and leguminosae species.

3.5 Expression patterns of *TrSPL* genes in different inflorescence development stage

In order to further forecast the function of *TrSPL* genes, 16 representative *TrSPL* genes were selected based on phylogenetic tree (Fig. 1). By constructing expression profiles in 5 different inflorescence development stages, preliminarily predicted function of genes was detected (Fig. 8). *TrSPL11*, *TrSPL13*, *TrSPL22* and *TrSPL26* had high expression only in the first development stage (T1), and and expression

decreased in subsequent stages. *TrSPL33* was highly expressed at T1 and T2, and decreased sharply at three stages after florets bloom. These results suggested that these genes may play an important role in the early development of *T. repens* inflorescence. Of course, some genes, such as *TrSPL1* and *TrSPL6*, were highly expressed only at inflorescence maturity (T5). With the development of inflorescence, the expression level of *TrSPL4* and *TrSPL12* gradually increased and reached the highest at inflorescence maturity (T5). It was worth noting that the relative expression of *TrSPL12* was the highest among the 16 genes, which may be closely related to the regulation of inflorescence development. The expression level of *TrSPL24* and *TrSPL25* increased sharply at T2 stage, and then decreased gradually with the development of inflorescences. The gene expression profile of *TrSPL* genes provided important information to determine the potential regulatory function of *T. repens SPL* gene family in inflorescence development.

4 Discussion

T. repens is high-quality leguminous forage, and has important economic value in temperate agricultural system (Frame et al., 1998). However, fowering directly affects the quality and yield of *T. repens*, and inflorescence development directly affects seed production. *SPL* gene family is a plant-specific transcription factor family containing a highly conserved SBP domain (76 amino acids), which can bind DNA in a sequence-specific manner and regulate transcription. *SPL* genes can specifically bind related motifs in *SQUAMOSA* promoter of snapdragon and *AP1* promoter of *Arabidopsis*, which have been proved to play an important role in regulating plant growth and development (Huijser et al., 1992; Mandel et al., 1992; Klein et al., 1996). In this study, 37 *TrSPL* genes were identified in *T. repens*, and much more than 16 in *Arabidopsis*, 19 in rice (Xie, Wu et al. 2006), 14 in barley (*Hordeum vulgare*) (Tong et al., 2020), and 27 in apple (Malus domestica Borkh.) (Li et al., 2013), but less than 56 in wheat (Ting et al., 2020), 57 in mustard (*Brassica juncea*) (Li et al., 2020), 48 in walnut (*Juglans regia*) (Zhou et al., 2019), 77 in euphorbiaceae (Li et al., 2019), and 58 in oilseed rape (*Brassica napus*) (Cheng et al., 2016). Generally, the number of gene family is partly affected by the genome size of species. Although the genomes of *Arabidopsis* (125Mb) (Schneeberger et al., 2011), rice (389Mb) (Takuji et al., 2005) and apple (632.4Mb) (Xuewei et al., 2016) are much smaller than *T. repens* (1174Mb) (Griffiths et al., 2019), the genome of mustard (1056.53Mb) (Li et al., 2020; Kang et al., 2021) and walnut (620Mb) (Annarita et al., 2020) were also smaller than *T. repens*. Gene replication events are also very important in determining the evolution and expansion of gene family, and species-specific gene replication is an important reason for determining the size of *SPL* gene family. *T. repens* is an allotetraploid leguminous forage. It is predicted that heterologous polyploidization event occurred in the last great glacier period (Warren et al., 2012), which may have an essential impact on the size of *TrSPL* gene family. Additionally, 5 pairs of segmental repeat genes were found while no tandem repeat gene pairs, which indicated that segmental repeat is more conducive to the evolution and population expansion of *T. repens SPL* gene family.

The isoelectric point, relative molecular weight and protein sequence length analysis of *TrSPL* genes showed that rich variation within this gene family. A large number of cis-acting elements related to light, hormone and stress response were found, which speculated that the functions of this gene family in *T.*

repens are diverse and may play a regulatory role in this physiological process. Furthermore, *TrSPL* gene showed similar gene structure and conserved motifs in the same clade, but there were significant differences among clades. Ankyrin repeat regions were found in all genes of the yellow clade, indicating that these genes may play an important role in protein-protein interaction. Owing to the ancestor *SPL* originally formed into two different lineages, named clade □ and clade □ (Hua et al., 2019). Based on the phylogenetic trees of *SPL* gene families of white clover, red clover, tribulus alfalfa and Arabidopsis further reveal the phylogenetic relationship between them. Besides, there were only ten pairs of homologous pairs between *T. repens* and *Arabidopsis*, while more homologous pairs were found in leguminous species such as red clover, Tribulus terrestris, alfalfa and soybean, indicating that the evolution of *SPL* gene in leguminous also had high conservation and homology.

Generally, genes in the same branch of the phylogenetic tree have the similar function. Gene expression patterns can provide crucial information for determining gene function prediction (Zhou et al., 2018). Previous studies have shown that the *Arabidopsis SPL* gene in clade V (*AtSPL3*) and clade VI (*AtSPL2*, *AtSPL10* and *AtSPL11*) could regulate flowering time (Cardon et al., 1997; Tao et al.,2019), and it was speculated that *TrSPL19-25* located in the same clade may have similar functions in regulating the flowering time of *T. repens*. Interestingly, the light response elements were detected in all of these genes. Similarly, the *Arabidopsis SPL* genes (*AtSPL8*, *AtSPL9* and *AtSPL15*) in clade III and clade VIII have been proved to affect inflorescences development (Schwarz et al., 2008; Unte et al., 2003), and *TrSPL10* to14 and *TrSPL32* to *35* (assigned into clade III and clade VIII) were possible relevance to inflorescences development of *T. repens*. Among these genes, *TrSPL11* and *TrSPL13* was highly expressed only at T1 stage, and *TrSPL33* was highly expressed at T1 and T2 stages, indicating that these genes play an important regulatory role in the early development of *T. repens* inflorescences. Specially, with the development of *T. repens* inflorescences, the expression of *TrSPL12* gradually increased and peaked at T5, indicating that *TrSPL12* may play an important effect with the development of inflorescences. *AtSPL* genes (*AtSPL1*, *AtSPL12* and *AtSPL14*) has been proved to be involved in regulating the development and its sensitivity to fumonisin B1 of *Arabidopsis*. Similarly, Vpsbp5 in the same clade has also been proved to prevent powdery mildew in grapes (Hongmin et al., 2013). In this study, results showed that *TrSPL* genes in cade □ may play an important role in enhancing disease resistance during the development of *T. repens* inflorescences. In brief, *TrSPL* gene family is such important in *T. repens* flowering regulation, especially in inflorescence development.

5 Conclusions

Trifolium repens is the most widely cultivated perennial legume forage in temperate region around the world. However, flowering time and inflorescence development directly affects the quality and yield, as well as seed production. *SPL* gene family is a plant specific transcription factor family, which has been proved to play a critical role in regulating plant formation time and development of flowers. In this study, a total of 37 *TrSPL* genes were identified from the whole genome of *T. repens* and were divided into 9 clades based on phylogenetic tree. The basic information of 37 *TrSPL* genes was obtained, including isoelectric point (PI), relative molecular weight (MW), protein sequence length (aa) and subcellular

localization. The result of cis acting element anlysis showed that a large number of action elements in response to light were identified and potential flowering regulation function was predicted. Besides, 37 *TrSPL* genes were relatively dispersedly distributed on 16 chromosomes, and 5 pairs of segmental repeat genes were found, which indicated that segmental duplication was the main way of gene expansion. Furthermore, the gene expression profiling showed that *TrSPL11*, *TrSPL13*, *TrSPL22* and *TrSPL26* were highly expressed only in the early stage of inflorescence development, while *TrSPL1* and *TrSPL6* are highly expressed only in the mature inflorescence. The results of this study will provide valuable clues for candidate gene selection and elucidating the molecular mechanism of *T. repens* flowering regulation.

Abbreviations

SPL
Squa promoter binding protein–like
SBP
squamosa promoter binding protein
CDS
Coding sequence
HMM
Hidden Markov model
MW
Molecular weight
PI
isoelectric point
NLS
nuclear localization signal.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the China Agriculture Research System of MOF and MARA, and the Sichuan Province Breeding Research grant (2021YFYZ0013).

Availability of data and materials

All data generated or analyzed during this study are included in this article and its attached documents. The *Trifolium repens* genome resources were downloaded from the EMBL/GenBank data libraries (Bioproject PRJNA523044, https://www.ncbi.nlm. nih.gov/genbank/) (Griffiths et al., 2019). The genome data used for comparative syntenic analysis were obtained from open database. The genome data of *Arabidopsis thaliana* was downloaded from TAIR (https://www. arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FSequences%2FAraport11_blastsets). The genome data of *Medicago truncatula* was downloaded from the *Medicago truncatula* Genome Database (http:// www.medicagogenome.org/). The genome data of *Trifolium pratense* was downloaded from Ensembl website (https://plants. ensembl. org/ index.html). The genome data of *Glycine max* was downloaded from the *Glycine max* Genome Database (https://www.soybase.org/).

Authors' contributions

GN and XZ conceived and designed the experiments. Bioinformatics analysis was performed by JM and ZY. SM and RH carry out planting and management of materials. JM, FW and JF complete the experimental part. JM and GN complete the manuscript. GN and XZ revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We gratefully acknowledge Stig Uggerhøj Andersen from Aarhus University for providing *Trifolium repens* genomic resources.

References

1. Abberton, M. T. and A. H. Marshall (2005). Progress in breeding perennial clovers for temperate agriculture. The Journal of Agricultural Science 143(2-3): 117-135.
2. Annarita, M., B. Monica, A. Z. Paulo, V. Z. Aleksey, E. W. Rachael, P. Daniela, B. Luca, P. Erica Adele Di, J. A. Brian, C. Sandeep, T. Michela, A. L. Charles, T. Winston, D. Abhaya, L. S. Steven and B. N. David (2020). High-quality chromosome-scale assembly of the walnut (Juglans regia L.) reference genome. GigaScience 9(5): giaa050.
3. Ayako, Y., W. Miin-Feng, Y. Li, W. Gang, R. S. Poethig and W. Doris (2009). The MicroRNA-Regulated SBP-Box Transcription Factor SPL3 Is a Direct Upstream Activator of LEAFY, FRUITFULL, and APETALA1. Developmental Cell 17(2): 268-278.
4. Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li and W. S. Noble (2009). MEME SUITE: tools for motif discovery and searching. Nucleic acids research 37(suppl_2): W202-W208.

5. Birkenbihl, R. P., G. Jach, H. Saedler and P. Huijser (2005). Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. Journal of molecular biology 352(3): 585-596.
6. Birte, B., J. Bernadette, K. Đura and H. John (2014). Legume Seed Production Meeting Market Requirements and Economic Impacts. Critical Reviews in Plant Sciences 34(1-3): 412-427.
7. Caradus, J., D. Woodfield and A. Stewart (1995). Overview and vision for white clover. NZGA: Research and Practice Series 6: 1-6.
8. Cardon, G., S. Hohmann, J. Klein, K. Nettesheim, H. Saedler and P. Huijser (1999). Molecular characterisation of the Arabidopsis SBP-box genes. Gene 237(1): 91-104.
9. Cardon, G. H., S. Hohmann, K. Nettesheim, H. Saedler and P. Huijser (1997). Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in the floral transition. The Plant journal : for cell and molecular biology 12(2): 367-377.
10. Chen, C., H. Chen, Y. Zhang, H. R. Thomas, M. H. Frank, Y. He and R. Xia (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Molecular Plant 13(8): 1194-1202.
11. Cheng, H. T., M. Y. Hao, W. X. Wang, D. S. Mei, C. B. Tong, H. Wang, J. Liu, L. Fu and Q. Hu (2016). Genomic identification, characterization and differential expression analysis of SBP-box gene family in Brassica napus. Bmc Plant Biology 16 (1): 1-17.
12. Christian, J. (2017). Flowering time regulation: Agrochemical control of flowering. Nature Plants 3(4): 1-2.
13. Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm and J. Mistry (2014). Pfam: the protein families database. Nucleic acids research 42(D1): D222-D230.
14. Finn, R. D., J. Clements and S. R. Eddy (2011). HMMER web server: interactive sequence similarity searching. Nucleic acids research 39(suppl_2): W29-W37.
15. Fiorella, D. B. N. and Y. Toshihiko (2017). Molecular Regulation of Flowering Time in Grasses. Agronomy 7(1): 17.
16. Frame, J., J. Charlton and A. S. Laidlaw (1998). Temperate forage legumes, Cab International.
17. Francesca, G. and F. Fabio (2021). SPL transcription factors prevent inflorescence reversion in rice. Molecular Plant 14(7): 1041-1043.
18. Gibson, P. and W. Cope (1985). White clover. Clover science and technology 25: 471-490.
19. Griffiths, A. G., R. Moraga, M. Tausen, V. Gupta, T. P. Bilton, M. A. Campbell, R. Ashby, I. Nagy, A. Khan, A. Larking, C. Anderson, B. Franzmayr, K. Hancock, A. Scott, N. W. Ellison, M. P. Cox, T. Asp, T. Mailund, M. H. Schierup and S. U. Andersen (2019). Breaking Free: The Genomics of Allopolyploidy-Facilitated Niche Expansion in White Clover. Plant Cell 31(7): 1466-1487.
20. Guangyan, F., H. Jiating, Y. Zhongfu, L. Qiuxu, S. Yang, X. Xiaoheng, N. Gang, H. Linkai, L. Wei and Z. Xinquan (2021). Genome-wide identification, phylogenetic analysis, and expression analysis of the

SPL gene family in orchardgrass (Dactylis glomerata L.). Genomics 113(4): 2413-2425.

21. Hongmin, H., Y. Qin, W. Xiping and X. Hui (2013). A SBP-Box Gene VpSBP5 from Chinese Wild Vitis Species Responds to Erysiphe necator and Defense Signaling Molecules. Plant Molecular Biology Reporter 31(6): 1261-1270.
22. Hua, Z., Weilong, K., Ziyun, G., Xinyi, F., Xiaoxiao, D., Chang, L., and Yangsheng, L., 2019, Evolutionary Analyses Reveal Diverged Patterns of SQUAMOSA Promoter Binding Protein-Like (SPL) Gene Family in Oryza Genus: Frontiers in Plant Science 10: 565.
23. Huijser, P., J. Klein, W. Lönnig, H. Meijer, H. Saedler and H. Sommer (1992). Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene squamosa in Antirrhinum majus. The EMBO journal 11(4): 1239-1249.
24. Jia, T., J. Hou, M. Z. Iqbal, Y. Zhang, B. Cheng, H. Feng, Z. Li, L. Liu, J. Zhou, G. Feng, G. Nie, X. Ma, W. Liu and Y. Peng (2021). Overexpression of the white clover TrSAMDC1 gene enhanced salt and drought resistance in Arabidopsis thaliana. Plant Physiology and Biochemistry 165: 147-160.
25. Kang, L., L. Qian, M. Zheng, L. Chen, H. Chen, L. Yang, L. You, B. Yang, M. Yan and Y. Gu (2021). "Genomic insights into the origin, domestication and diversification of Brassica juncea." Nature genetics 53(9): 1392-1402.
26. Kilcher, M. R. (1981). Plant Development, Stage of Maturity and Nutrient Composition. Journal of Range Management 34(5).
27. Klein, J., H. Saedler and P. Huijser (1996). A new family of DNA binding proteins includes putative transcriptional regulators of the Antirrhinum majus floral meristem identity gene SQUAMOSA. Molecular and General Genetics MGG 250(1): 7-16.
28. Lei, W. and Z. Qifa (2017). Boosting Rice Yield by Fine-Tuning SPL Gene Expression. Trends in Plant Science 22(8): 643-646.
29. Li, J., X. Y. Gao, S. Y. Sang and C. N. Liu (2019). Genome-wide identification, phylogeny, and expression analysis of the SBP-box gene family in Euphorbiaceae. Bmc Genomics 20 20(9): 1-15.
30. Li, J., H. Hou, X. Li, J. Xiang, X. Yin, H. Gao, Y. Zheng, C. L. Bassett and X. Wang (2013). Genome-wide identification and analysis of the SBP-box family genes in apple (Malus x domestica Borkh.). Plant Physiology and Biochemistry 70: 100-114.
31. Li, M. Y., Q. He, Y. Zhang, B. Sun, Y. Luo, Y. Zhang, Q. Chen, Y. Wang, F. Zhang, Y. T. Zhang, Y. X. Lin, X. R. Wang and H. R. Tang (2020). New insights into the evolution of the SBP-box family and expression analysis of genes in the growth and development ofBrassica juncea. Biotechnology & Biotechnological Equipment 34(1): 810-824.
32. Mandel, M. A., G.-B. Cindy, S. Beth and F. Y. Martin (1992). Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. Nature 360(6401): 273-277.
33. Pederson, G. A. and G. E. Brink (2000). Seed Production of White Clover Cultivars and Naturalized Populations when Grown in a Pasture. Crop Science 40(4): 1109-1114.
34. Preston, J. C. and L. C. Hileman (2013). Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. Frontiers in Plant Science 4: 13.

35. Riese, M., S. Höhmann, H. Saedler, T. Münster and P. Huijser (2007). Comparative analysis of the SBP-box gene families in P. patens and seed plants. Gene 401(1-2): 28-37.
36. Salinas, M., S. Xing, S. Höhmann, R. Berndtgen and P. Huijser (2012). Genomic organization, phylogenetic comparison and differential expression of the SBP-box family of transcription factors in tomato. Planta 235(6): 1171-1184.
37. Schneeberger, K., S. Ossowski, F. Ott, J. D. Klein, X. Wang, C. Lanz, L. M. Smith, J. Cao, J. Fitz and N. Warthmann (2011). Reference-guided assembly of four diverse Arabidopsis thaliana genomes. Proceedings of the National Academy of Sciences 108(25): 10249-10254.
38. Schwarz, S., A. V. Grande, N. Bujdoso, H. Saedler and P. Huijser (2008). The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis. Plant molecular biology 67(1-2): 183-195.
39. Takuji, S. and P. International Rice Genome Sequencing (2005). The map-based sequence of the rice genome. Nature 436(7052): 793-800.
40. Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution 30(12): 2725-2729.
41. Tao, Y., P. Bong Soo, M. Hui-Zhu, S. Jun Sung, O. Naohiko, L. Ying, Y. Niu, M. Nur Fatimah Binte, H. Chung-Hao and C. Nam-Hai (2019). Regulation of flowering time by SPL10/MED25 module in Arabidopsis. New Phytologist 224(1): 493-504.
42. Ting, Z., L. Yue, M. Liting, W. Xiaoying, Z. Dazhong, H. Yucui, D. Qin and M. Lingjian (2020). Genome-wide identification, phylogeny and expression analysis of the SPL gene family in wheat. BMC Plant Biology 20(1): 1-14.
43. Tong, T., Y. X. Fang, Z. L. Zhang, J. J. Zheng, X. L. Lu, X. Q. Zhang and D. W. Xue (2020). Genome-wide identification, phylogenetic and expression analysis of SBP-box gene family in barley (Hordeum vulgare L.). Plant Growth Regulation 90(1): 137-149.
44. Unte, U. S., A.-M. Sorensen, P. Pesaresi, M. Gandikota, D. Leister, H. Saedler and P. Huijser (2003). SPL8, an SBP-box gene that affects pollen sac development in Arabidopsis. The Plant Cell 15(4): 1009-1019.
45. Wang, H. F., Z. C. Lu, Y. T. Xu, L. C. Kong, J. J. Shi, Y. F. Liu, C. X. Fu, X. S. Wang, Z. Y. Wang, C. E. Zhou and L. Han (2019). Genome-wide characterization of SPL family in Medicago truncatula reveals the novel roles of miR156/SPL module in spiky pod development. Bmc Genomics 20: 14.
46. Warren, M. W., W. E. Nicholas, A. A. Helal, M. V. Isabelle and S. W. Hussain (2012). Experimental evidence for the ancestry of allotetraploid Trifolium repens and creation of synthetic forms with value for plant breeding. BMC Plant Biology 12(1): 1-10.
47. Xie, K., C. Wu and L. Xiong (2006). Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. Plant physiology 142(1): 280-293.
48. Xuewei, L., K. Ling, Z. Jing, X. Yinpeng, W. Liping, Y. Yan, W. Na, X. Jidi, L. Cuiying, W. Wen, N. Steve van, D. Yang, M. Fengwang and G. Qingmei (2016). Improved hybrid de novo genome assembly of

domesticated apple (Malus x domestica). GigaScience 5(1): s13742-016-0139-0.

49. Yamasaki, K., T. Kigawa, M. Inoue, M. Tateno, T. Yamasaki, T. Yabuki, M. Aoki, E. Seki, T. Matsuda and E. Nunokawa (2004). A novel zinc-binding motif revealed by solution structures of DNA-binding domains of Arabidopsis SBP-family transcription factors. Journal of molecular biology 337(1): 49-63.
50. Zhou, L., S. W. Quan, L. Ma, H. Xu, J. P. Yang and J. X. Niu (2019). Molecular characterization of SBP-box gene family during floral induction in walnut (Juglans regia L.). Tree Genetics & Genomes 16(1): 1-16.
51. Zhou, Q., S. Zhang, F. Chen, B. Liu, L. Wu, F. Li, J. Zhang, M. Bao and G. Liu (2018). Genome-wide identification and characterization of the SBP-box gene family in Petunia. BMC genomics 19(1): 1-18.

Tables

Table 1 The basic information and subcellular localization of all identified *SPL* genes in *T. repens*

Name	PI	MW (Da)	Length (aa)	Subcellular Localization
*TrSPL*1	6.12	86752.27	773	nucleus
*TrSPL*2	6.88	67153.17	591	nucleus
*TrSPL*3	8.18	109672.44	989	plasma membrane
*TrSPL*4	5.75	111233.87	1010	endomembrane system
*TrSPL*5	5.72	111303.94	1014	endomembrane system
*TrSPL*6	7.01	116204.16	1053	nucleus
*TrSPL*7	8.51	115211.08	1044	endomembrane system
*TrSPL*8	5.8	111853.78	1004	nucleus
*TrSPL*9	5.86	110745.6	995	nucleus
*TrSPL*10	8.92	37297.64	335	nucleus
*TrSPL*11	9.14	40922.71	363	nucleus
*TrSPL*12	6.9	14269.05	124	nucleus
*TrSPL*13	9.3	20662.94	182	nucleus
*TrSPL*14	9.41	20574.88	182	nucleus
*TrSPL*15	5.98	57844.27	527	nucleus
*TrSPL*16	6.46	55524.04	509	nucleus
*TrSPL*17	6.31	55948.26	500	nucleus
*TrSPL*18	8.97	30286.98	267	chloroplast
*TrSPL*19	8.88	48472.58	439	nucleus
*TrSPL*20	8.73	44554.3	406	nucleus
*TrSPL*21	8.17	45859.11	408	nucleus
*TrSPL*22	8.67	34247.9	313	nucleus
*TrSPL*23	8.67	34313.96	312	nucleus
*TrSPL*24	6.1	17499.07	149	nucleus
*TrSPL*25	6.1	17472.04	149	nucleus
*TrSPL*26	7.6	41040.79	367	nucleus
*TrSPL*27	7.61	40962.63	367	nucleus
*TrSPL*28	8.86	42684.84	383	nucleus

*TrSPL*29	8.62	42817.9	385	nucleus
*TrSPL*30	6.62	42612.13	382	nucleus
*TrSPL*31	6.78	42977.52	387	nucleus
*TrSPL*32	9.2	37707.11	339	chloroplast
*TrSPL*33	9.41	37953.47	340	nucleus
*TrSPL*34	6.99	37614.66	343	plasma membrane
*TrSPL*35	8.73	36551.32	335	nucleus
*TrSPL*36	7.04	16415.99	142	nucleus
*TrSPL*37	7.04	16430.02	142	nucleus

Figures

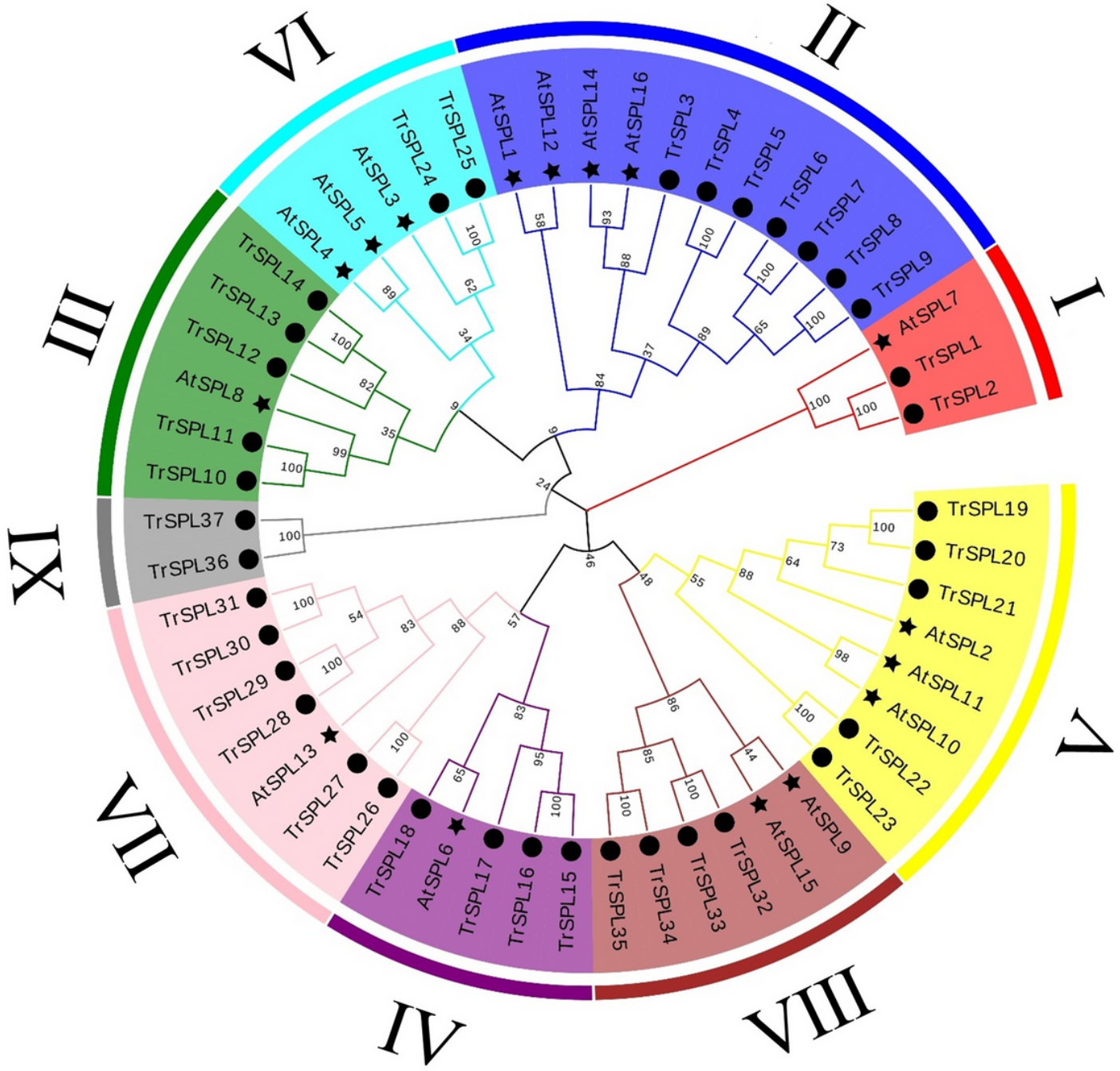

Figure 1

Phylogenetic analysis of SPL proteins in *Arabidopsis thaliana* and *T. repens*. Phylogenetic trees were plotted using the neighbor-joining (NJ) method with a bootstrap value of 1000. 53 genes were divided into 9 clades (Ⅰ-Ⅸ) and identified with different colors. The black circle represents *T. repens*, and the black pentacle represents *Arabidopsis*.

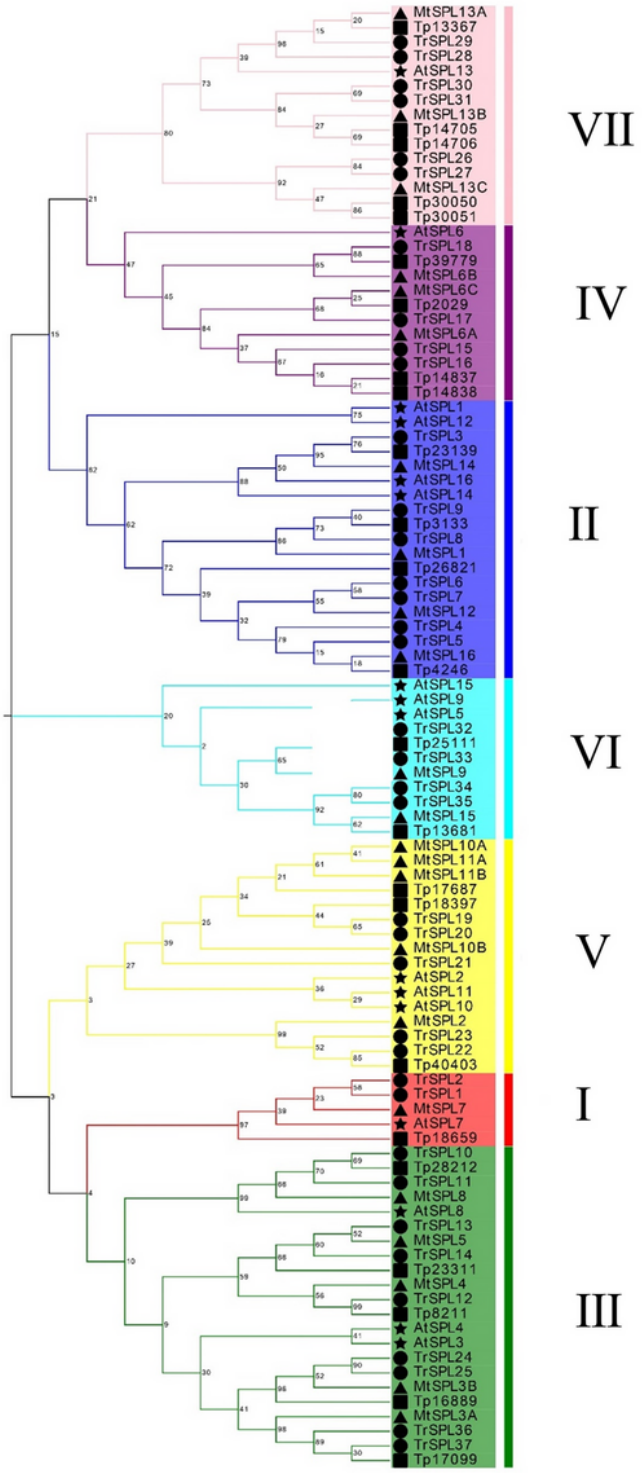

Figure 2

Phylogenetic analysis of SPL proteins in *Arabidopsis*, *Trifolium repens*, *Trifolium pratense* and *Medicago truncatula*. Phylogenetic trees were plotted using the neighbor-joining (NJ) method with a bootstrap value of 1000. A total of 100 genes were divided into 7 clades (Ⅰ–Ⅶ) and identified with different colors. The black circle, pentacle, square and triangle represent *Trifolium repens*, *Arabidopsis*, *Trifolium pratense* and *Medicago truncatula* respectively.

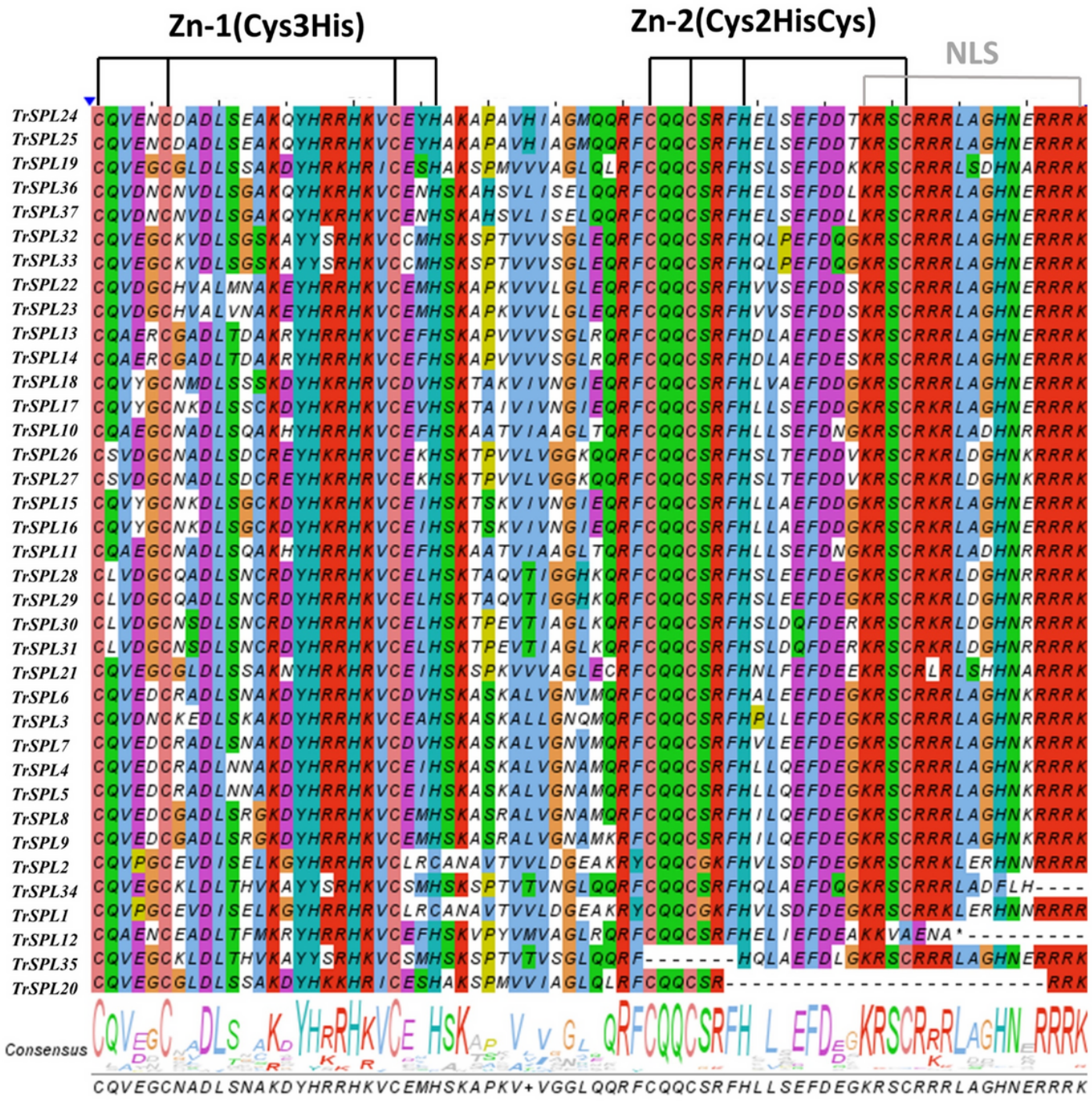

Figure 3

Sequence alignment and logo of SBP domain from *T. repens*. Two Zn-finger structures (Zn-1, Cys3 His; Zn-2, Cys2HisCys) and one NLS structure have been marked.

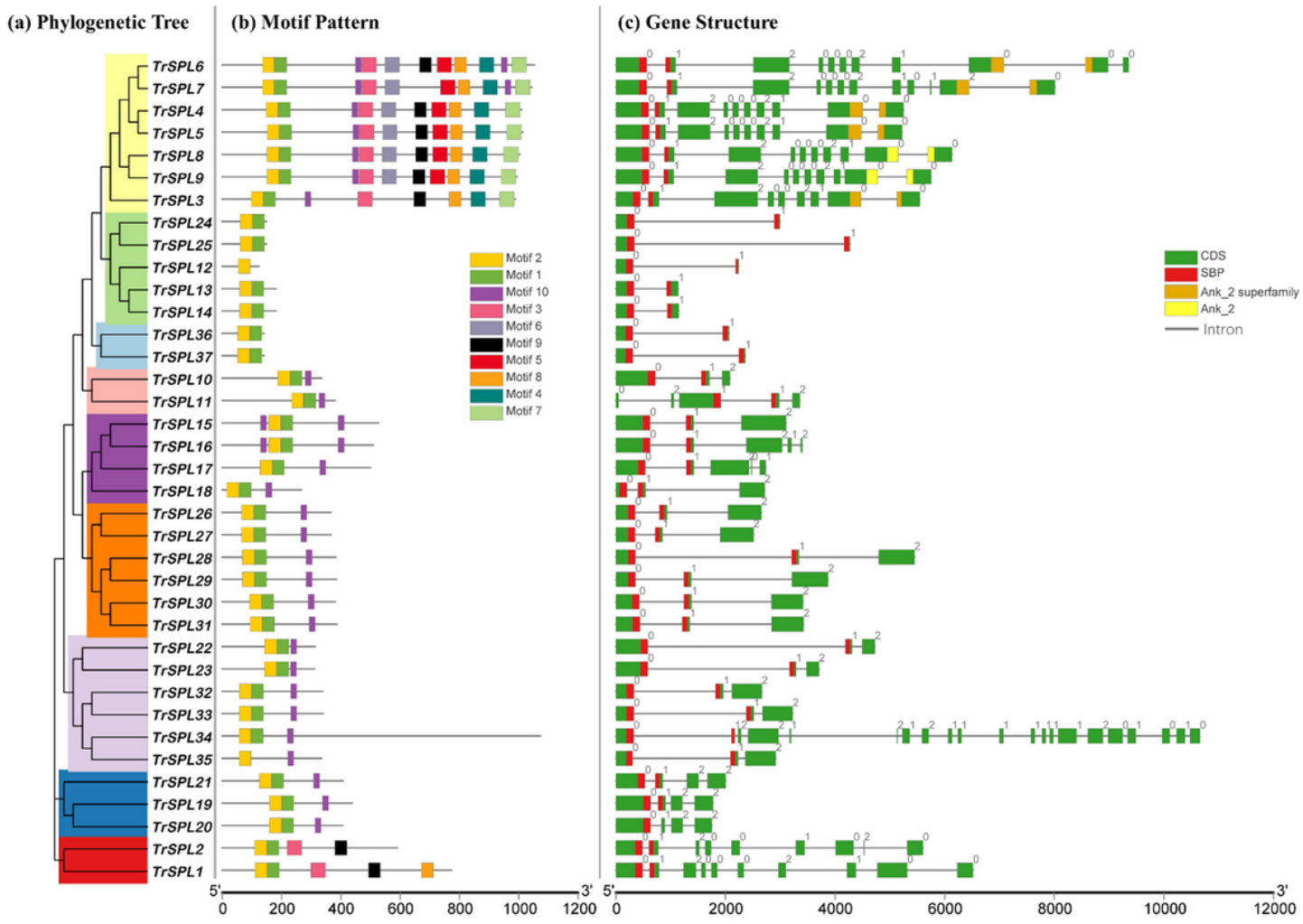

Figure 4

Analysis of conserved motifs and gene structure under the phylogenetic tree of *T. repens SPL* gene family. (a) A phylogenetic tree was constructed using the fulllength protein sequence of *T. repens SPL* gene family. Different evolutionary clades were marked with different colors. (b) Conserved motifs predicted in *TrSPL* proteins. The 10 motifs are represented by squares of different colors. (c) Exons and introns are represented by colored squares and black lines. The SBP conserved domain and Ankyrin repeat region were clearly marked and 0, 1 and 2 indicate intron phase.

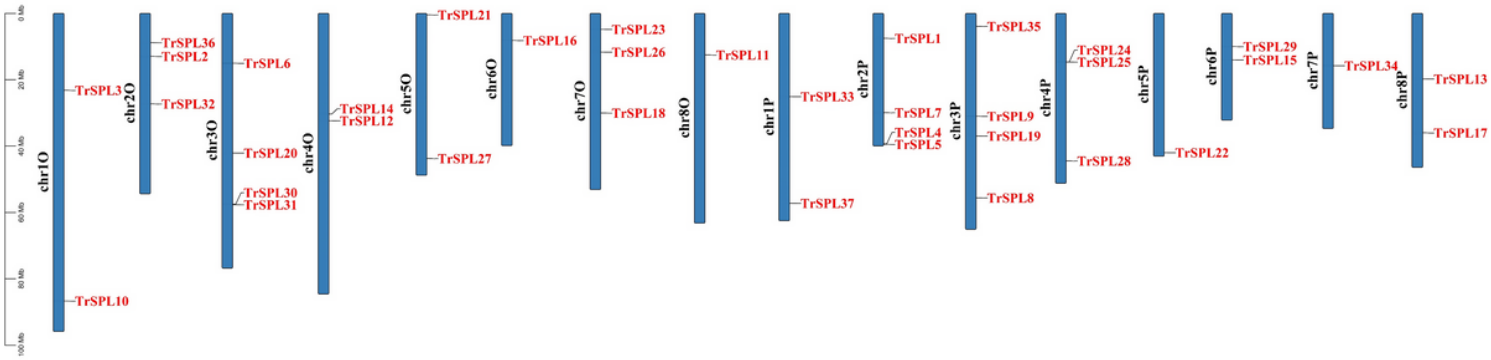

Figure 5

Distribution of *TrSPL* genes on chromosomes. The leftmost scale shows the length of chromosomes. Chromosomes are represented by blue bars.

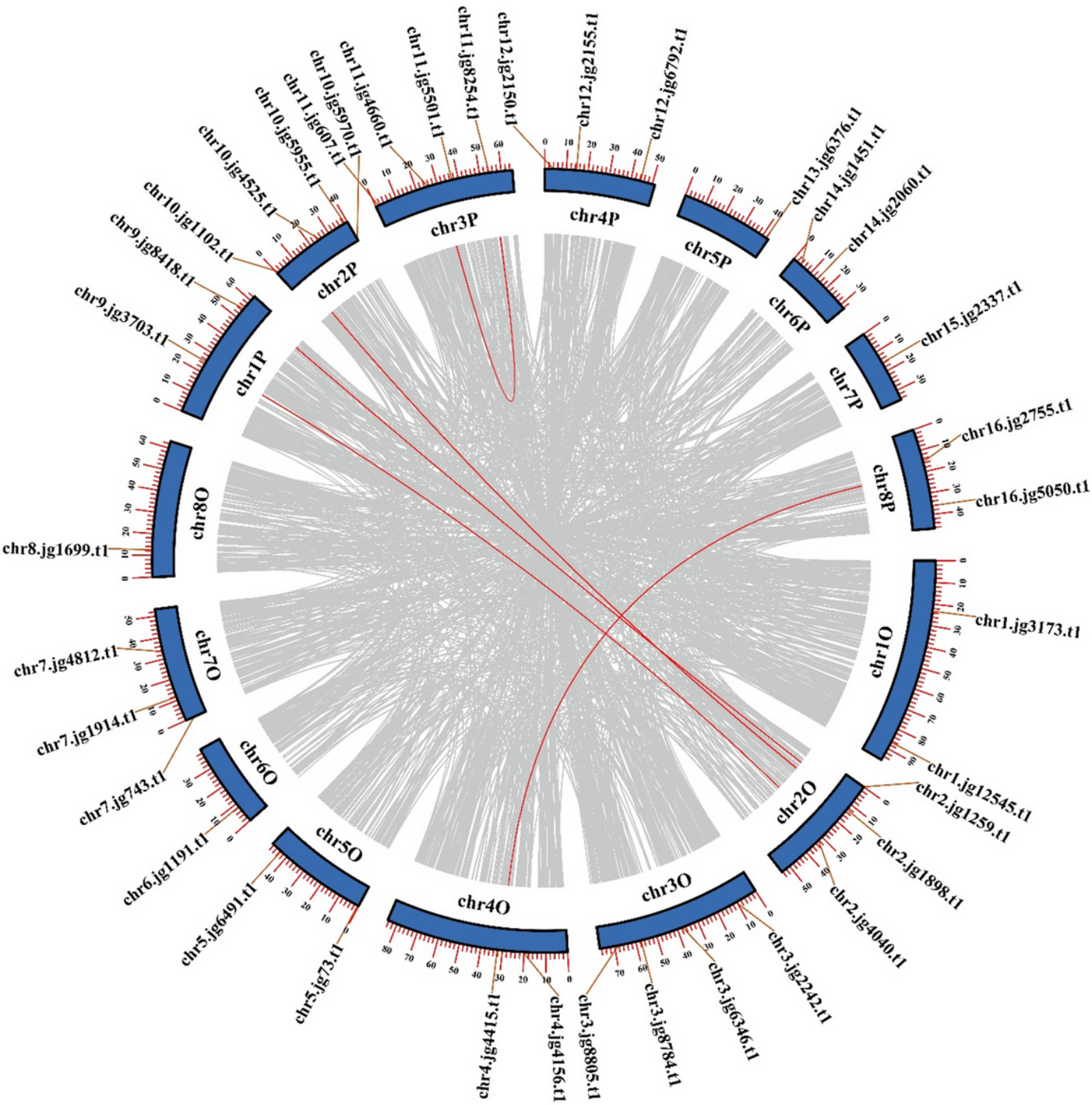

Figure 6

Segmental repeat analysis of *SPL* genes in *T. repens* genome. The 16 chromosomes form a circle, and the red lines represent the syntenic region of the 37 *TrSPL* genes. All synteny block produced by doubling the genome are indicated by gray lines.

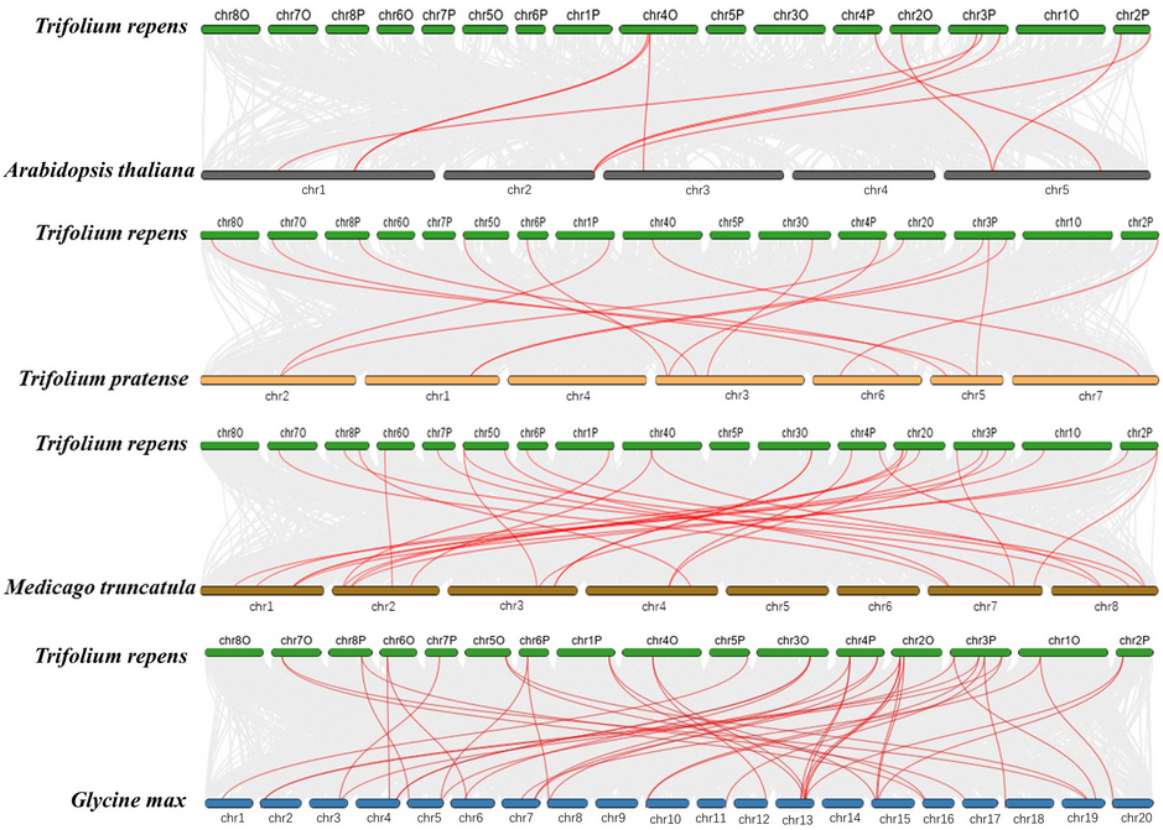

Figure 7

Collinearity analysis of *SPL* gene families between *T. repens* and representative species. The red line showed the collinearity of *SPL* gene family in *T. repens* and corresponding representative species. Other collinearity between genomes is indicated by gray lines.

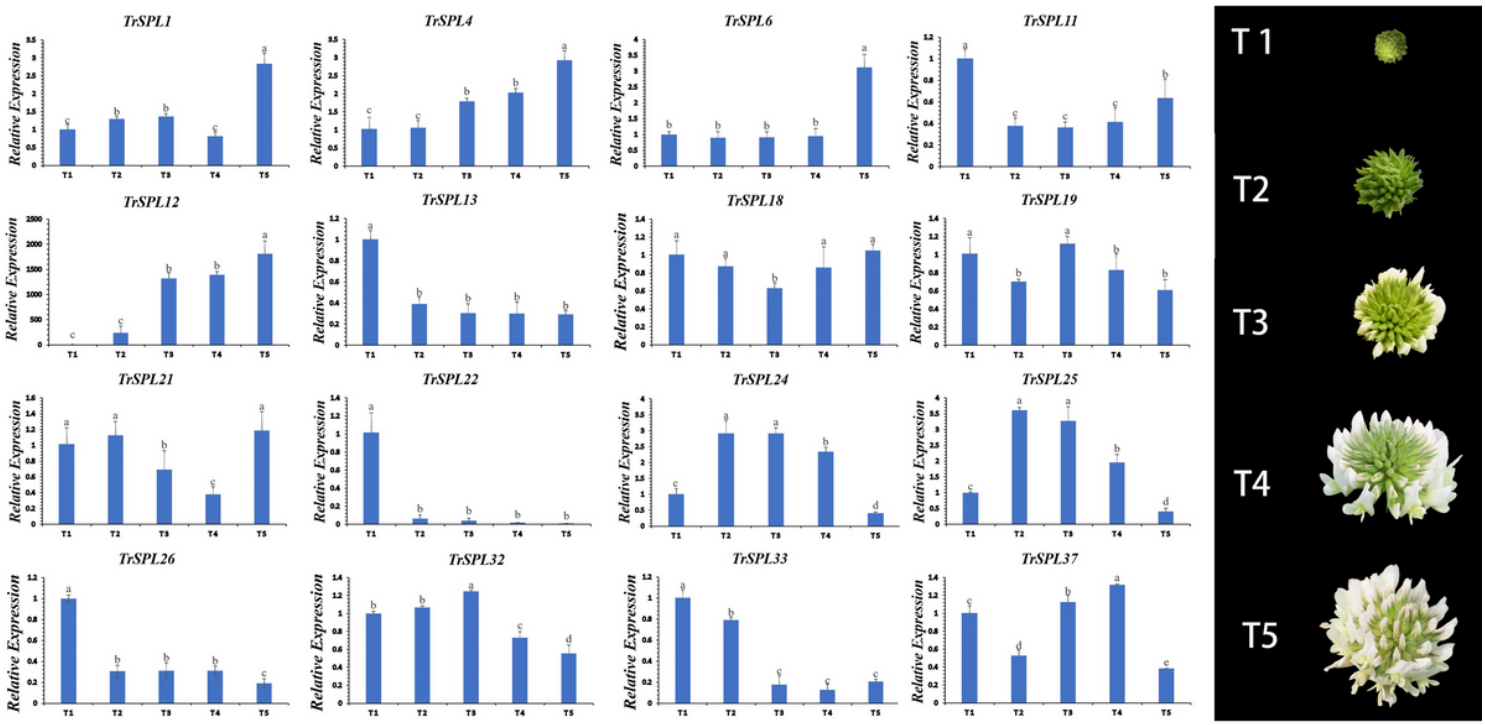

Figure 8

Expression profiles of 16 selected *TrSPL* genes at 5 different flower development stages in *T. repens*. T1: Immature inflorescence, T2: Inflorescences in which no floret was open, T3: Inflorescences in which outermost circle of florets were open, T4: Inflorescences in which 50% of florets were open, T5: Mature inflorescence. The vertical bar is the standard deviation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile.zip