RESEARCH ARTICLE

Genome-Wide Identification and Expression Profiling of the SBP-box Gene Family in *Medicago truncatula*

Maryam Noor[1], Muhammad Afzal[2], Ameena Dilshad[3], Muhammad Awais[4], Sial Haris[5], Iqra Mehar[1] and Hadia Hussain[1,6,*]

[1] Department of Biotechnology, University of Okara, Okara 56300, Pakistan
[2] Department of Plant Production, College of Food and Agricultural Sciences, King Saud University, Riyadh 11451, Saudi Arabia
[3] Department of Forest Protection, College of Forestry and Grassland, Nanjing Forestry University, Nanjing 210037, China
[4] Bahauddin Zakariya University, Multan 60800, Pakistan
[5] Shanxi Medical University (Zhongdu Campus), Jinzhong 030600, China
[6] State Key Laboratory of Tree Genetics and Breeding, College of Life Sciences, Nanjing Forestry University, Nanjing 210037, China

Abstract

The Squamosa Promoter Binding Protein (SBP) family comprises plant-specific transcription factors widely distributed across green plant lineages and holds significant value for agricultural improvement by regulating key traits including yield, flowering, and stress tolerance. Here, we performed a genome-wide analysis of *Medicago truncatula* and identified 22 SBP-box genes (*MtSBP*). These genes are irregularly distributed across seven chromosomes, with chromosome 6 devoid of any SBP-box gene. Sequence alignment confirmed that all MtSBP proteins share conserved nuclear localization signals and two zinc finger structures essential for transcriptional regulation. Phylogenetic analysis grouped these genes into seven subgroups, supported by conserved domain organization, motif composition, and gene structure. Gene Ontology analysis indicated that most *MtSBP* genes are involved in biological regulation. Expression profiling using public RNA-seq data revealed variable patterns across different nitrogen sources and symbiotic conditions, suggesting roles in nitrogen metabolism, symbiosis, and growth. Overall, this study provides a comprehensive overview of the SBP-box family in *M. truncatula* and identifies candidate genes for future functional studies on legume growth and nitrogen-use efficiency.

Keywords: *Medicago truncatula*, SBP-box gene family, phylogenetic analysis, Gene Ontology, expression analysis, nitrogen utilization, symbiotic interaction.

1 Introduction

Transcriptional factors are key regulators of gene expression, either by inhibiting or promoting

Submitted: 28 March 2026
Revised: 26 June 2026
Accepted: 28 June 2026
Published: 10 August 2026

Vol. 1, **No.** 2, 2026.
10.62762/PIJ.2026.672525

***Corresponding author:**
✉ Hadia Hussain
hadiahussain37@outlook.com

Citation
Noor, M., Afzal, M., Dilshad, A., Awais, M., Haris, S., Mehar, I., & Hussain, H. (2026). Genome-Wide Identification and Expression Profiling of the SBP-box Gene Family in *Medicago truncatula*. *Plant Innovation Journal*, 1(2), 99–117.

transcription. Transcription factor families are differentiated based on the presence of either binding domains for DNA or other conserved domains. The Squamosa Binding Protein (SBP) gene box family is a family of transcriptional factors present in plants. This family contains a highly conserved domain for DNA binding of approximately 80 amino acids along with a zinc finger-like structure of C3H and C2HC [1]. Domain analysis reveals nuclear localization interconnected with a zinc finger-like structure present at the C-terminal side. SBP domains are involved in two processes: nuclear import and binding to a consensus DNA sequence that contains the GTAC core motif [2]. First defined in *Antirrhinum majus* and subsequently identified across eukaryotes in genome-wide transcription factor surveys [3, 4], the SBP-box gene family has since been identified in various plant species, including angiosperms, green algae, gymnosperms, and mosses [5]. Moreover, genome-wide analysis was also recorded in *Arabidopsis thaliana*, *Oryza sativa* [6], *Nicotiana tabacum* [7], *Solanum lycopersicum* [8], *V. vinifera* [9], *Malus domestica* [10], and *Glycine max* [11]. SBP genes are believed to be involved in various functions like shoot development, flowering, fruit ripening, and stress responses [5].

Further research demonstrates their function in stress tolerance and disease resistance. *OsSPL10* was implicated in trichome formation and salt tolerance in *Oryza sativa* [12]. By regulating anthocyanin formation, *MsSPL9* reduces *Medicago sativa*'s resistance to drought [13]. Silencing of *CaSBP08* in *Capsicum annuum* enhanced resistance to *Phytophthora capsici* [14]. Accumulating evidence indicates that SPL (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE) genes, which belong to the plant-specific SBP-box transcription factor family, have undergone extensive neofunctionalization during evolution. For instance, the rice gene *GW8*/*OsSPL16*, an ortholog of *TGA1*, regulates grain shape, size, and quality [15, 16]. In tomato, an ortholog of *AtSPL3* at the colorless non-ripening locus is essential for fruit ripening [17]. *AtSPL7*, an SPL gene not targeted by miR156, modulates copper homeostasis and contributes to cadmium responses [18]. *AtSPL8* acts as a tissue-dependent regulator governing gynoecium differential patterning, early anther development, sepal trichome formation, and stamen development in *Arabidopsis* [19]. Additionally, *AtSPL14* has been shown to participate in plant development and modulate sensitivity to the mycotoxin fumonisin B1 [20].

Medicago truncatula is a small legume native to the Mediterranean Sea. It has garnered attention due to its genomic proximity to legumes such as soybeans and alfalfa, which are agriculturally significant. We used this plant in our study because of its small genome size (approximately 500Mb) and its ability to fix nitrogen. Self-pollination and high genetic transformation ability make it suitable for our analysis. Extensive studies have examined transcriptomic responses in *Medicago truncatula*, including cell- and tissue-specific analyses of root interactions with symbiotic and pathogenic microbes, supporting legume functional genomics research [21]. Although a prior study identified SPL genes in *M. truncatula* using an earlier genome assembly and revealed the role of the miR156/SPL module in spiky pod development [22], comprehensive functional and structural analyses of the entire SBP-box family based on the current Mt4.0v1 genome have not yet been conducted. Finding and examining the expression traits of the SBP gene family in *Medicago truncatula* is crucial, as these genes are extensively involved in development and plant growth, as well as in how the plant responds to various abiotic stresses.

In this study, we utilized bioinformatics tools and genomic resources specific to this plant to analyze the SBP gene family in *Medicago truncatula*. This analysis included identifying the genes, exploring their evolutionary relationships, and determining their chromosomal locations. We also demonstrated synteny analyses among *Medicago truncatula*, *Arabidopsis thaliana*, *Solanum lycopersicum*, and *Brassica rapa*. The results of this article will inform further investigation into the role of the SBP gene family in plants and its use for genetic improvement.

2 Materials and Methods

2.1 Identification and analysis of SBP properties in *Medicago Truncatula*

We searched the Plant Transcription Factor Database (PlantTFDB v4.0[1]) [23] for SBP genes in *Medicago truncatula*. Genome files for *Medicago truncatula* were downloaded from the Phytozome database (Phytozome v13[2]), and the *Arabidopsis* gene sequence was identified from TAIR[3]. Sequences of *Arabidopsis* were blasted utilizing protein–protein BLAST

[1] http://planttfdb.gao-lab.org/
[2] https://phytozome-next.jgi.doe.gov/
[3] https://www.arabidopsis.org/browse/genefamily/index.jsp/

(BLASTP) against the genomic sequences of *Medicago* with the following parameters: E-value $\leq$ 1e-5, identity $\geq$ 30%, query coverage $\geq$ 50%, and default BLOSUM62 substitution matrix [24]. Following the removal of redundant sequences, we obtained 22 final gene sequences containing SBP domains. Theoretical isoelectric point values were obtained using the ExPASy ProtParam tool (v2025[4]). Amino acid length and molecular weight were also identified along with their starting point and chromosome number.

2.2 Sequence alignment, conserved domain analysis, and 3D structure prediction of SBP genes in *Medicago truncatula*

BioEdit software was used to perform multiple sequence alignments based on the identified SBP proteins. To confirm the presence of the SBP domain, a conserved domain analysis was performed using the NCBI Batch Conserved Domain Database (CDD) tool[5]. In addition, three-dimensional (3D) protein structures were predicted using the Swiss-Model[6] server to study the structural properties of the discovered SBP proteins.

2.3 Phylogenetic relationship, motif analysis, and gene structure analysis of SBP genes in *Medicago truncatula*

A phylogenetic tree of MtSBP proteins was constructed using MEGA 11[7]. To investigate the evolutionary relationships among selected plant species, three well-studied species were chosen: *Medicago truncatula* Mt4.0v1, *Arabidopsis thaliana*, and *Oryza sativa* v7.0. A total of 22 SBP gene sequences from *Medicago truncatula*, 17 sequences from *Arabidopsis thaliana*, and 20 sequences from *Oryza sativa* were aligned using the ClustalW method within the MEGA 11 software. The phylogenetic relationships among all SBP genes from *Medicago*, *Arabidopsis*, and *Oryza* were analyzed using the Maximum Likelihood (ML) method, with a bootstrap value of 1000, using MEGA 11. MEME software v4.11.1[8] was used to identify conserved motifs with default parameters. A total of ten motifs were identified. Gene structure was analyzed using GSDS v2.0[9]. For gene structure analysis, we used genomic sequences and CDS (coding sequences) of the *MtSBP* genes.

2.4 SBP gene-box family chromosomal distribution, and *in silico* subcellular localization

Furthermore, we investigated the precise chromosomal positions of *Medicago* genes. All 22 *Medicago* SBP genes were located on their respective chromosomes. The location of *MtSBP* on the chromosome was examined and mapped using the MG2C v2[10]. *In silico* subcellular localization of these SBP genes was predicted using data from WolfPsort[11] and CELLO v.2.5[12], with scores $\geq$ 0.7 indicating high confidence. A heatmap was created using TBTools v1.120 to illustrate these localizations.

2.5 Synteny analysis of MtSBP genes with other plant species

To study the exact region and homologous genes between *Medicago truncatula* Mt4.0v1, *Arabidopsis thaliana*, *Solanum lycopersicum* ITAGv2.4, and *Brassica rapa* O_302V v1.1, we performed dual synteny analysis using the MCScanX v.2.010 Multiple Collinearity Scan Toolkit[13] with default scoring parameters.

2.6 GO analysis and expression pattern analysis

Gene Ontology analysis was performed from the Plant Transcriptional Regulatory Map database (PlantRegMap[14]) and the GO map was constructed using the SR plot online server[15].

Additionally, organ-specific data for *Medicago truncatula* genes were retrieved from the Phytozome V14 database using the BioMart/GeneAtlas v2 module[16]. The organism *Medicago truncatula* Mt4.0v1 was selected, and expression information was obtained by querying the identified MtSBP gene IDs. The attributes selected for data retrieval included Sample Group, Gene Name, Organism Name, Sample Name, Expression Value, Expression Class, and Expression Units. The downloaded dataset contained FPKM-based expression values for different tissues and treatment conditions, including leaf and root samples under ammonium, nitrate, urea, and symbiotic conditions. The expression matrix was extracted from the downloaded file and used for heatmap visualization of MtSBP gene expression patterns. A heatmap was generated using TBTools v1.120 based on $\log_{10}(\text{FPKM} + 1)$ transformed

[4] https://www.expasy.org/
[5] https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi
[6] https://beta.swissmodel.expasy.org/
[7] https://www.megasoftware.net/
[8] https://meme-suite.org/meme/
[9] https://gsds.gao-lab.org/
[10] http://mg2c.iask.in/mg2c_v2.0/
[11] https://wolfpsort.hgc.jp/
[12] http://cello.life.nctu.edu.tw/
[13] https://ngdc.cncb.ac.cn/biocode/tools/BT000733
[14] https://plantregmap.gao-lab.org/go.php
[15] https://www.bioinformatics.com.cn/en
[16] https://phytozome-next.jgi.doe.gov/geneatlas/

Table 1. Details of 22 *MtSBP* genes with their location on respective chromosomes

Transcript IDs	Genes	Chromosome	Start	End	Strand
Medtr3g099080.1	*MtSBP1*	chr3	45,410,077	45,413,482	Forward
Medtr3g085180.1	*MtSBP2*	chr3	38,492,622	38,497,369	Forward
Medtr2g014200.1	*MtSBP3*	chr2	3,964,614	3,967,481	Forward
Medtr2g461920.1	*MtSBP4*	chr2	25,606,880	25,611,792	Forward
Medtr2g046550.1	*MtSBP5*	chr2	20,453,788	20,462,574	Forward
Medtr2g078770.1	*MtSBP6*	chr2	32,971,839	32,974,296	Reverse
Medtr8g096780.1	*MtSBP7*	chr8	40,622,635	40,626,632	Forward
Medtr8g005960.1	*MtSBP8*	chr8	419,201	421,415	Reverse
Medtr8g080670.1	*MtSBP9*	chr8	34,719,932	34,722,421	Reverse
Medtr8g463140.1	*MtSBP10*	chr8	22,196,924	22,198,280	Reverse
Medtr8g080690.1	*MtSBP11*	chr8	34,729,478	34,731,697	Reverse
Medtr8g080680.1	*MtSBP12*	chr8	34,725,301	34,727,462	Reverse
Medtr7g028740.1	*MtSBP13*	chr7	9,871,681	9,875,095	Forward
Medtr7g110320.1	*MtSBP14*	chr7	45,210,189	45,220,920	Reverse
Medtr7g092930.1	*MtSBP15*	chr7	36,893,346	36,897,295	Reverse
Medtr7g444860.1	*MtSBP16*	chr7	15,012,334	15,016,587	Reverse
Medtr5g046670.1	*MtSBP17*	chr5	20,459,088	20,465,349	Reverse
Medtr1g053715.1	*MtSBP18*	chr1	22,678,197	22,682,537	Reverse
Medtr1g035010.1	*MtSBP19*	chr1	12,692,333	12,698,670	Forward
Medtr1g086250.1	*MtSBP20*	chr1	38,604,281	38,611,936	Forward
Medtr4g088555.1	*MtSBP21*	chr4	35,174,503	35,179,012	Reverse
Medtr4g109770.1	*MtSBP22*	chr4	45,646,471	45,650,584	Reverse

expression values to visualize gene expression patterns.

3 Results

3.1 Identification and physicochemical properties of SBP-box genes in *Medicago truncatula*

In this study, SBP-box genes were identified at the whole-genome level in *Medicago truncatula*. BLAST searches were performed to obtain candidate sequences, and SBP-related entries were retrieved from the plant transcription factor database (PlantTFDB). After removing redundant sequences, a total of 22 non-redundant SBP-box family genes were identified in *M. truncatula*, and designated as *MtSBP1–MtSBP22*. These genes were unevenly distributed across seven chromosomes. The genomic lengths of the identified genes varied considerably, ranging from 1,357 bp (*MtSBP10*) to 10,732 bp (*MtSBP14*), which indicates substantial structural diversity within the SBP gene family. Detailed chromosomal information of the 22 identified *MtSBP* genes is provided in Table 1.

The physicochemical properties of the 22 identified MtSBP proteins were analyzed using the ExPASy ProtParam tool. Protein length exhibited considerable variation, ranging from 143 amino acids in MtSBP3 to 1,025 amino acids in MtSBP5. Correspondingly, molecular weight varied substantially from 16.78 kDa (MtSBP3) to 112.97 kDa (MtSBP5). The theoretical isoelectric point (pI) values ranged from 5.80 (MtSBP20) to 9.20 (MtSBP10), indicating notable variation in charge properties among MtSBP proteins, with the majority exhibiting basic characteristics (pI > 7). GRAVY values were negative for all members, ranging from -1.483 (MtSBP21) to -0.356 (MtSBP5), indicating their overall hydrophilic nature. These results highlight substantial diversity in protein length, molecular weight, isoelectric point, and hydrophilicity among MtSBP family members, suggesting functional diversification within *Medicago truncatula*. Detailed physicochemical properties of the 22 identified MtSBP proteins are provided in Table 2. The complete protein sequences of all identified genes are provided in Supplementary File S1.

3.2 Multiple sequence alignment, domain analysis and structure prediction of the SBP gene family in *Medicago truncatula*

To investigate the evolutionary conservation of the SBP domain, a motif sequence logo was generated

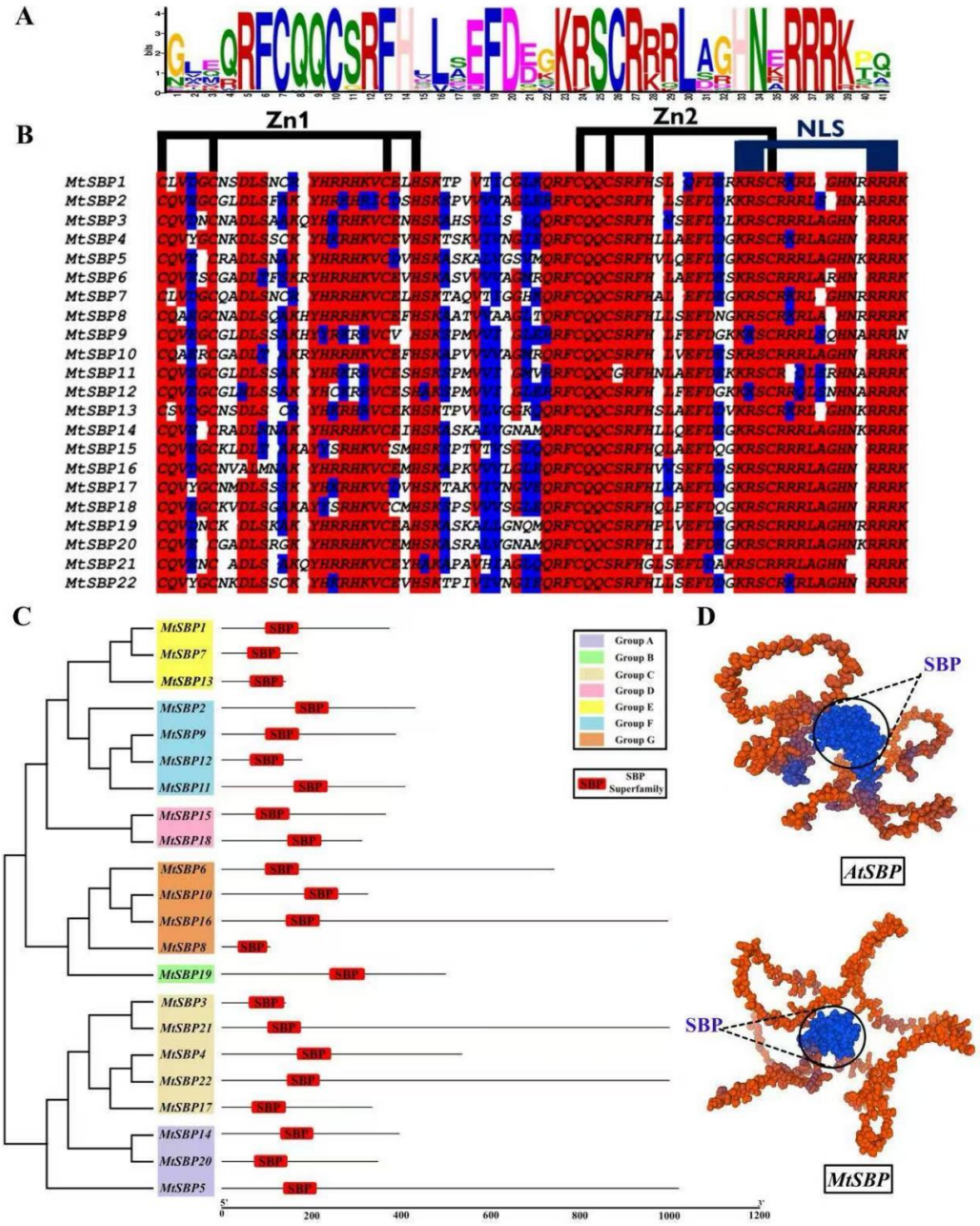

Figure 1. Multiple sequence alignment and structural analysis of SBP domain proteins in *Medicago truncatula*. (A) The Motif sequence logos were based on the alignment of MtSBP domains (B) ClustalW alignment of MtSBP proteins showing conserved two Zn1 and Zn2 zinc finger motifs and a nuclear localization signal (NLS). (C) Presence of the conserved SBP domain across 22 MtSBP genes along phylotree. (D) Predicted 3D structures of AtSBP and MtSBP proteins showing conservation of the SBP domain architecture.

using the conserved domain sequences of all identified MtSBP proteins (Figure 1(A)). The logo analysis revealed several highly conserved amino acid residues indicating strong conservation of the SBP DNA-binding region among family members (Figure 1(A)).

Multiple sequence alignment further confirmed the presence of the highly conserved SBP domain in all 22 MtSBP proteins. This domain consisted of

Table 2. Physicochemical properties of 22 *MtSBP* genes of *Medicago truncatula*.

Gene	A. A	MW (Da)	pI	GRAVY
MtSBP1	376	41,747.45	6.75	−0.686
MtSBP2	434	47,989.13	8.59	−0.833
MtSBP3	143	16,784.41	6.45	−1.305
MtSBP4	539	59,024.06	6.49	−0.598
MtSBP5	1,025	112,971.07	6.38	−0.356
MtSBP6	171	19,628.88	8.64	−1.018
MtSBP7	390	43,350.90	8.49	−0.807
MtSBP8	327	36,196.59	8.75	−0.919
MtSBP9	412	45,977.34	8.18	−0.674
MtSBP10	180	20,625.19	9.20	−1.212
MtSBP11	376	41,864.91	8.95	−0.657
MtSBP12	398	44,941.22	7.56	−0.717
MtSBP13	367	40,921.73	8.46	−0.824
MtSBP14	1,001	110,602.48	5.92	−0.421
MtSBP15	337	36,441.21	9.12	−0.707
MtSBP16	314	34,548.38	8.90	−0.696
MtSBP17	500	56,341.74	7.28	−0.698
MtSBP18	350	38,267.56	9.12	−0.662
MtSBP19	1,003	111,687.71	7.53	−0.541
MtSBP20	1,003	111,555.25	5.80	−0.482
MtSBP21	144	16,950.61	6.66	−1.483
MtSBP22	470	52,323.01	6.16	−0.558

two zinc-finger motifs, designated as Zn1 and Zn2, together with a conserved nuclear localization signal (NLS). The zinc-finger regions contained conserved cysteine and histidine residues responsible for zinc ion coordination, whereas the NLS region was enriched with basic amino acids, particularly arginine and lysine residues, indicating the nuclear localization of MtSBP proteins (Figure 1(B)). Members clustered within the same phylogenetic group generally exhibited similar domain organization and sequence conservation, suggesting functional relatedness and a common evolutionary origin (Figure 1(C)). Furthermore, three-dimensional structural modeling revealed a highly conserved SBP domain between *Medicago truncatula* and *Arabidopsis thaliana* SBP proteins, supporting the conserved DNA-binding function of this transcription factor family (Figure 1(D)). Overall, these results demonstrate the high evolutionary conservation of the SBP domain despite variation in protein size and sequence among MtSBP family members.

3.3 Phylogenetic Analysis of the SBP gene family in *Medicago truncatula*

To study the evolutionary relationships of SBP (SQUAMOSA Promoter Binding Protein) transcription factors, a phylogenetic tree was built using SBP protein sequences from *Medicago truncatula*, *Arabidopsis thaliana*, and *Oryza sativa*. The circular phylogeny produced a clear grouping of SBP genes into seven major clades (A–G), showing a well-conserved but diverse gene family throughout plant lineages (Figure 2). Strong evolutionary conservation of SBP domains among monocots and dicots is suggested by the wide distribution of SBP members from the three species across all clades. While *Oryza sativa* (monocot) sequences formed different but related branches, indicating anticipated lineage-specific divergence following the monocot–dicot split, *Arabidopsis thaliana* and *Medicago truncatula* (dicots) displayed closer grouping within numerous subgroups. Overall, the phylogenetic analysis shows that SBP genes in *Medicago truncatula* are evolutionarily conserved but moderately expanded, with clear orthologous relationships between species and evidence of diversification, which could contribute to functional specialization in developmental and stress-responsive regulatory pathways (Figure 2).

3.4 Gene Structure Analysis of MtSBP genes in *Medicago truncatula*

To determine the evolutionary relationships among MtSBP genes, we examined their gene structure, emphasizing the exon–intron arrangement. Gene structure features, such as the number and order of introns and exons, are critical for gene identification, functional characterization, and evolutionary inference. Accordingly, the intron–exon architecture of MtSBP genes was studied to gain insight into their coding sequence organization. The analysis revealed a consistent distribution pattern of untranslated regions (UTRs), coding sequences (CDSs), and introns throughout the MtSBP gene family (Figure 3). The number of introns among the identified genes ranged from 1 to 9. Similarly, CDS segments varied, ranging from 2 (*MtSBP3*, *MtSBP6*, *MtSBP10*, and *MtSBP21*) to 10 (*MtSBP5*, *MtSBP14*, and *MtSBP20*). Most genes contained UTR sections, with counts ranging from one (*MtSBP1*) to three (*MtSBP2*, *MtSBP7*, and *MtSBP8*). However, no UTRs were found in *MtSBP6*, *MtSBP9*, *MtSBP11*, *MtSBP12*, *MtSBP17*, and *MtSBP18* (Figure 3).

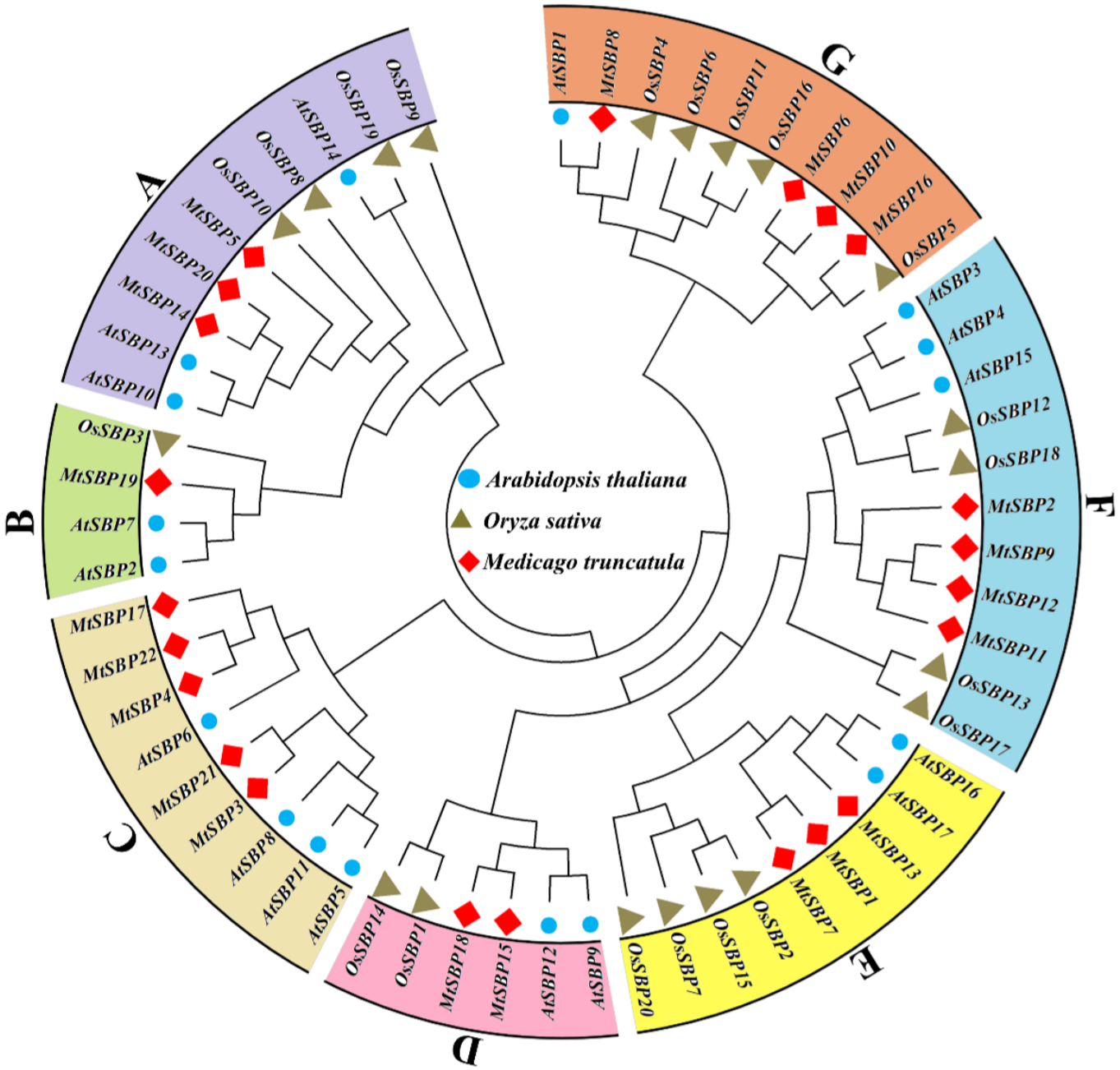

Figure 2. Phylogenetic investigation of *Medicago truncatula* SBP gene family. The evolutionary relationships of SBP proteins from *Medicago truncatula*, *Arabidopsis thaliana*, and *Oryza sativa* are displayed in a circular maximum likelihood phylogenetic tree. The SBP proteins are divided into different clades (A–G). Species are represented by different symbols: *Medicago truncatula* is represented by red diamonds, *Oryza sativa* by green triangles, and *Arabidopsis thaliana* by blue circles. The clustering pattern demonstrates both lineage-specific expansion of SBP gene members in *M. truncatula* and preserved orthologous connections.

3.5 Motif Analysis of MtSBP genes in *Medicago truncatula*

Using the MEME program, motif analysis of MtSBP proteins revealed 10 common motifs throughout the gene family (Figure 4(A, B)). The findings demonstrated that MtSBP proteins with identical motif compositions and configurations typically belong to the same phylogenetic grouping, suggesting a high level of structural conservation within each clade. This implies that within the MtSBP gene family, conserved motif patterns may play a role in subgroup-specific functional activities. Among the identified motifs, motif 1, motif 2, and motif 3 were highly conserved and were present in all 22 MtSBP proteins (Figure 4(A)), highlighting their essential role in maintaining the core structural and

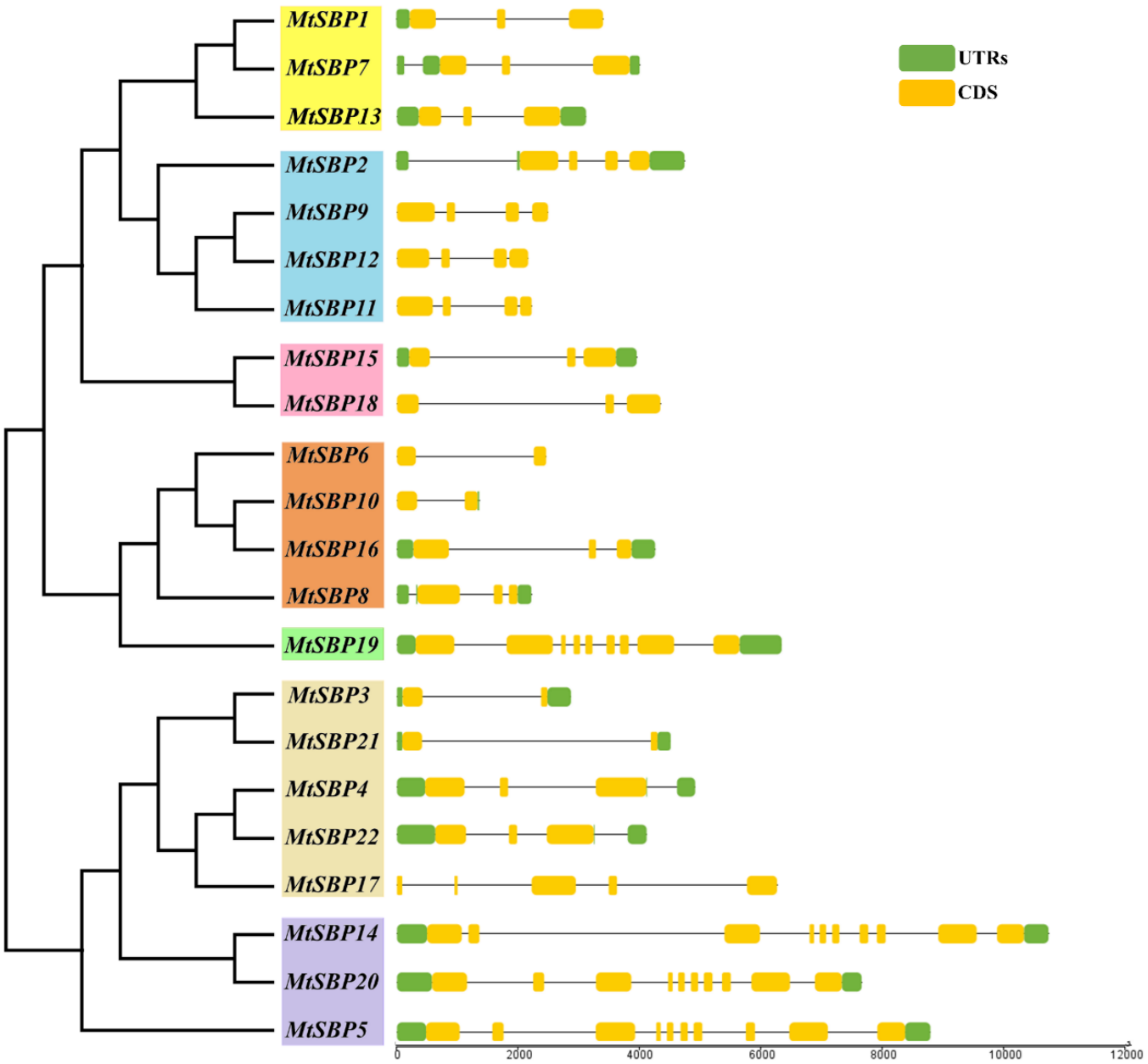

Figure 3. Exon-intron structure of putative *MtSBP* genes. The length of each *MtSBP* gene's exon-intron was shown proportionally alongside the evolutionary tree. Each *MtSBP* gene has black lines, yellow bars, and green bars to denote the number of introns, CDS, and UTR.

functional features of SBP transcription factors. In contrast, the remaining motifs displayed variable distribution across different members, suggesting functional diversification and possible specialization among MtSBP proteins.

3.6 Chromosomal Localization of SBP Genes and Comparative Synteny Analysis in *Medicago truncatula*

The chromosomal distribution of genes is closely linked to their regulatory roles in plant development and gene expression. To investigate the chromosomal distribution of SBP genes, their locations were mapped in *Medicago truncatula* using TBTools and available genome annotation data. As shown in Figure 5(A), the 22 identified MtSBP genes were unevenly distributed across seven chromosomes (Chr1, Chr2, Chr3, Chr4, Chr5, Chr7, and Chr8), while no SBP genes were detected on chromosome 6. Among these, chromosomes 2 and 7 each contained four MtSBP genes, chromosome 8 contained six genes, chromosomes 1 and 5 each contained three genes, and chromosomes 3 and 4 each carried two MtSBP genes (Figure 5(A)).

A comparative synteny analysis of *Medicago truncatula*,

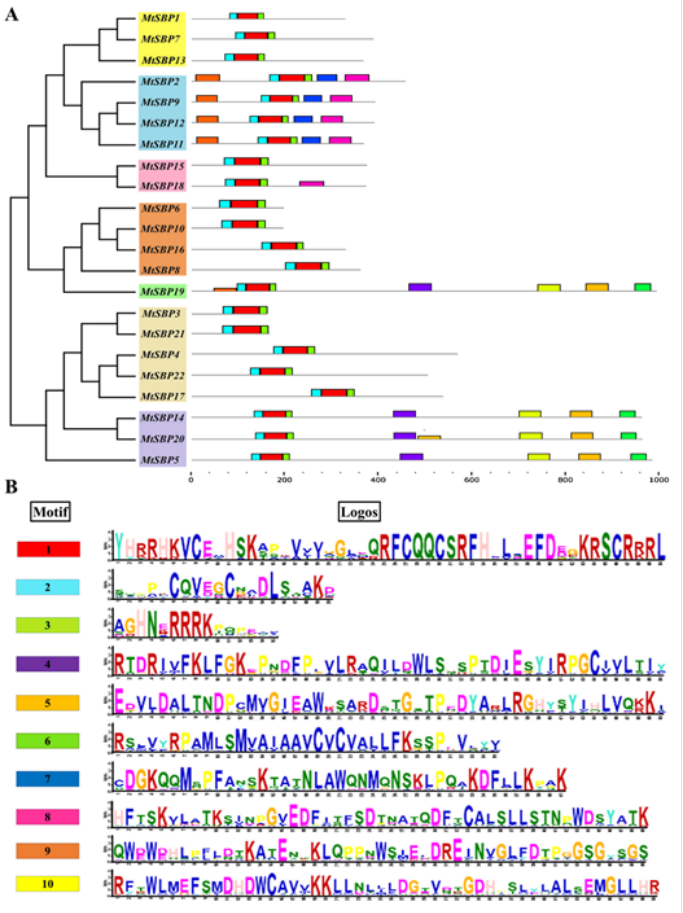

Figure 4. Conserved Motif Analysis of MtSBP Proteins in *Medicago truncatula* (A) Schematic representation of conserved motifs identified in MtSBP proteins using the MEME suite. Each colored box denotes a distinct conserved motif. (B) Sequence logos of the ten identified motifs, illustrating the conserved amino acid patterns and their relative frequencies within each motif.

Arabidopsis thaliana, *Solanum lycopersicum*, and *Brassica rapa* was conducted to better investigate the evolutionary links of SBP genes. The findings demonstrated a significant and highly conserved syntenic connection between *Medicago truncatula* and *Brassica rapa*, demonstrating considerable genomic overlap between the two species' SBP gene families. *Arabidopsis thaliana* had reduced syntenic connections, but *Solanum lycopersicum* had intermediate conservation (Figure 5(B)). The significant synteny observed with *Brassica rapa* shows a tight evolutionary link, and it is possible that SBP genes in both species were maintained through similar ancestral duplication events and functional conservation during evolution. This pattern also suggests that SBP genes may have conserved regulatory roles across related plant lineages, particularly in the *Brassicaceae* and allied species.

3.7 In silico Subcellular localization of *MtSBPs* genes

In silico subcellular localization analysis of the MtSBP gene family suggested a highly diverse but predominantly nucleus-associated distribution pattern across predicted cellular compartments. Three complementary visualization approaches (heatmap, pie chart, and stacked bar representation) collectively demonstrated both the overall localization tendency and gene-specific variability within the family (Figure 6). As demonstrated in the heatmap (Figure 6(A)), the majority of MtSBP proteins showed significant predicted localization signals in the nucleus, showing that nuclear compartmentalization is the predominant and most conserved aspect of this gene family.

The distribution pattern (Figure 6(B)) suggested this tendency, with the nucleus accounting for around 79% of projected localization locations. Chloroplasts accounted for about 4% of the remaining percentage, followed by vacuoles (~3%), Golgi apparatus (~3%), plastids (~2%), cytosol (~1%), extracellular space (~1%), and peroxisomal compartments (~1%). Gene-wise localization profiling (Figure 6(C)) revealed variation among individual MtSBP members. Several genes had exclusive or near-exclusive nuclear localization predictions, while others had multi-compartmental distribution patterns that combined nuclear signals with small contributions from chloroplasts, plastids, and secretory pathway-related organelles. Notably, a few members have more complex localization signals, indicating that the gene family may be functionally diverse. Overall, MtSBPs are likely nuclear proteins, suggesting their potential role as transcription factors, while a few members show additional non-nuclear localization, suggesting possible functional diversification and involvement in other cellular processes. However, these results are based solely on in silico predictions and require experimental validation.

3.8 Gene Ontology (GO) enrichment analysis of MtSBPs genes

Gene Ontology (GO) enrichment analysis of the *MtSBP* gene family revealed that these genes are functionally diversified and significantly associated with multiple biological processes, cellular components, and molecular functions. The enriched GO terms were categorized into three main groups: Biological Processes (BP), Cellular

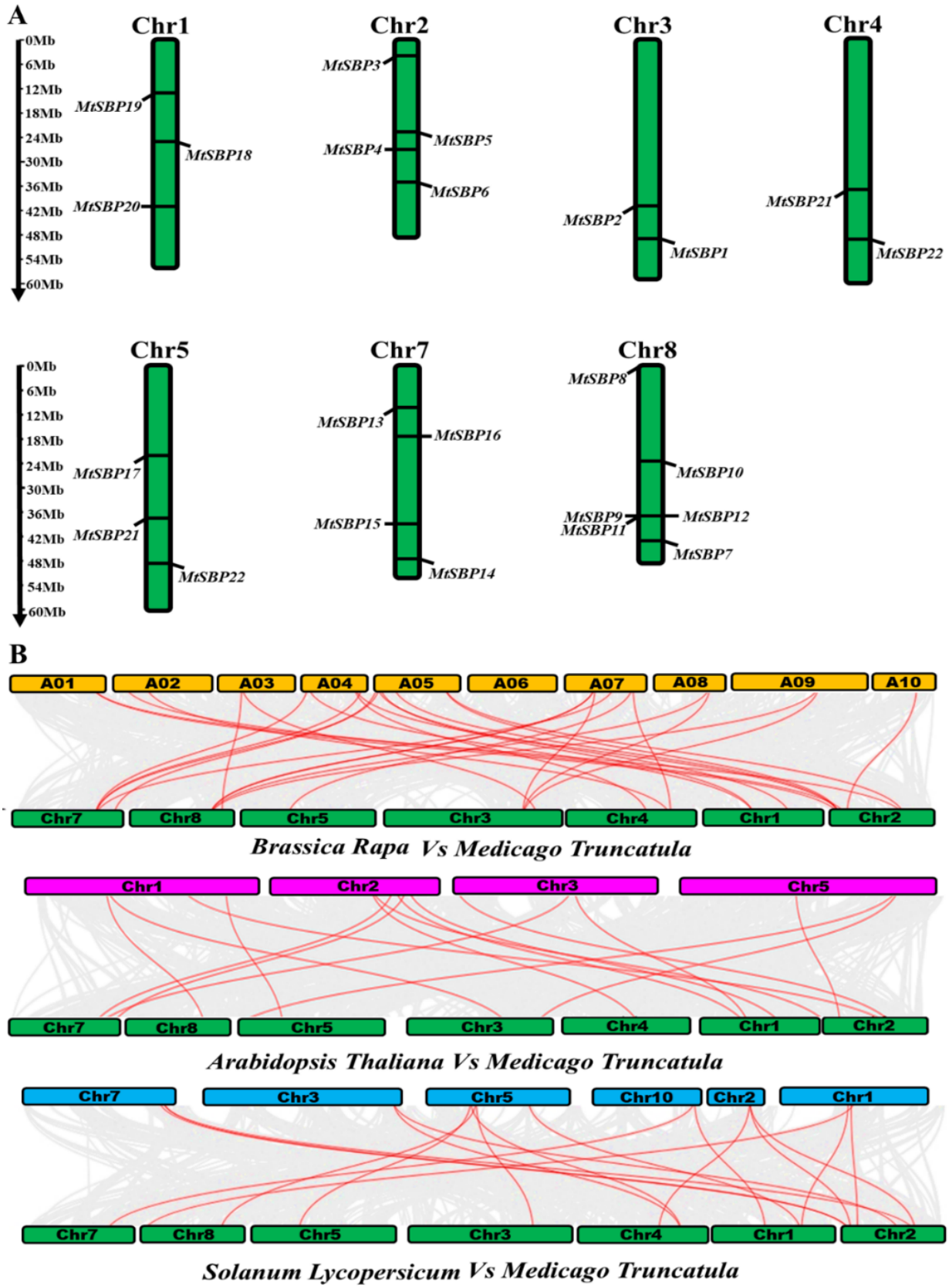

Figure 5. Chromosomal Localization of *MtSBPs* Genes and Comparative Synteny Analysis. (A) Chromosome distribution of genes on 7 different chromosomes. The scale bar on the left shows the length of the chromosome. (B) Synteny analysis of *MtSBPs* genes. Matching red lines indicate segmental duplication.

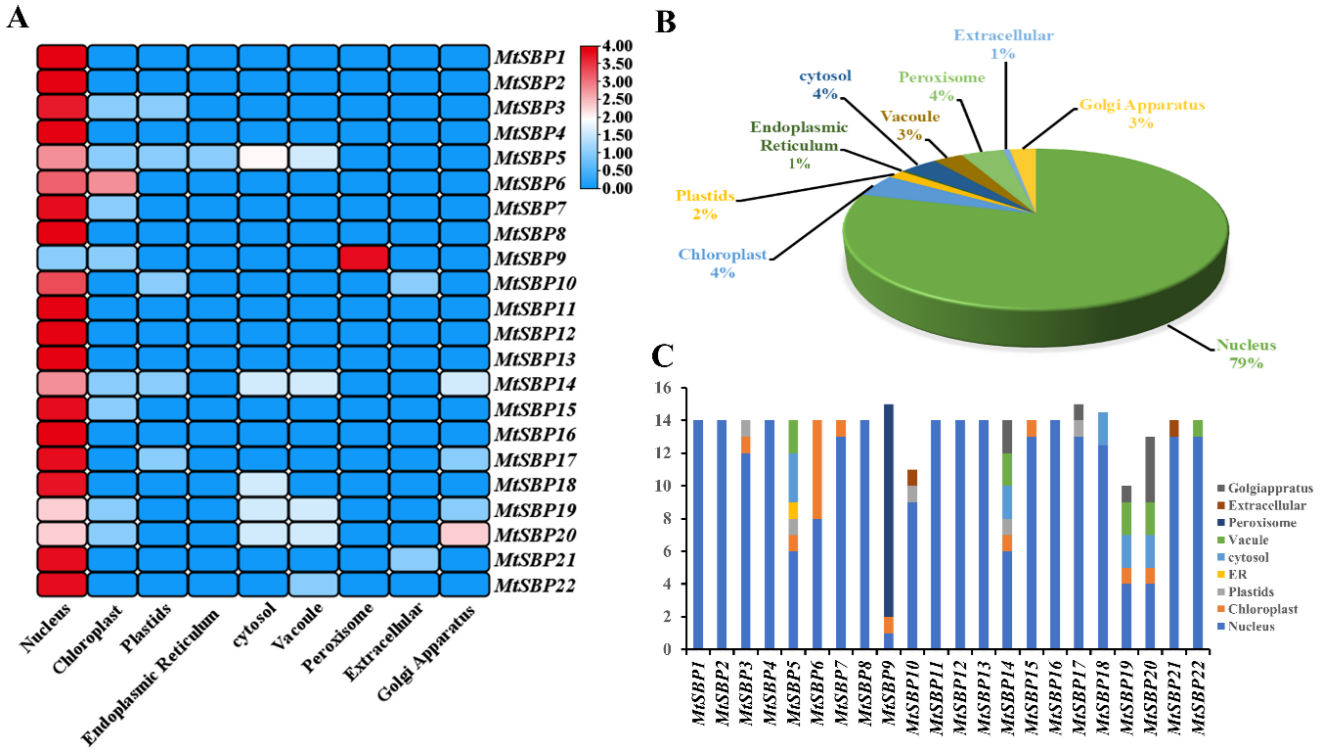

Figure 6. In silico subcellular localization of MtSBPs proteins. Heatmap (A), pie chart (B), and gene-wise distribution (C) showing the predicted subcellular localization of MtSBPs proteins across different cellular compartments.

Components (CC), and Molecular Functions (MF), indicating the broad functional roles of *MtSBP* proteins in plant growth and development (Figure 7). The biological process category was the most highly represented, comprising 37 GO terms. The most abundant functional annotations within this category were associated with "regulation of gene expression," "developmental processes," "response to bacterium," "post-embryonic development," and broader "biological regulation," indicating that *MtSBP* genes are primarily involved in regulating key developmental and stress-responsive pathways.

The cellular component category included seven GO terms that were mainly associated with "nucleus," "membrane-bound organelles," and "cell," suggesting that MtSBP proteins are predominantly localized to nuclear and intracellular compartments, consistent with their roles as transcriptional regulators. The molecular function category was the least represented, containing five GO terms, with "DNA binding" identified as the most prominent, highlighting the sequence-specific regulatory activity of MtSBP transcription factors in gene expression control. Overall, the GO enrichment results indicate that MtSBP genes primarily function as nuclear-localized DNA-binding transcriptional regulators involved in developmental processes, gene expression regulation, and stress-related responses. These findings suggest that MtSBPs act as key components of regulatory networks governing plant growth and development. Detailed annotations of all GO terms are provided in the Supplementary File (Table S1).

3.9 Organ-specific expression profiling of *MtSBPs* genes

Expression data from the Phytozome database suggested considerable variation in MtSBP transcript abundance between leaf and root tissues under different nitrogen sources (ammonium, nitrate, and urea) as well as under symbiotic conditions. This diversity in expression patterns indicates that members of the MtSBP gene family likely perform distinct roles in nitrogen metabolism and symbiotic interactions. *MtSBP14* had the highest overall expression of any gene tested, especially in leaves treated with urea (L.U), followed by symbiotic (L.Sc) and nitrate (L.N) conditions. *MtSBP5* expression was likewise continuously high across all leaf treatments, with maximal levels reported during symbiosis, implying a significant role in leaf-associated nitrogen responses. Furthermore, *MtSBP3*, *MtSBP10*, *MtSBP13*, and *MtSBP16* were highly expressed in leaf tissues, indicating that they are involved in nitrogen-responsive regulatory networks. In root tissues, overall expression levels were comparatively

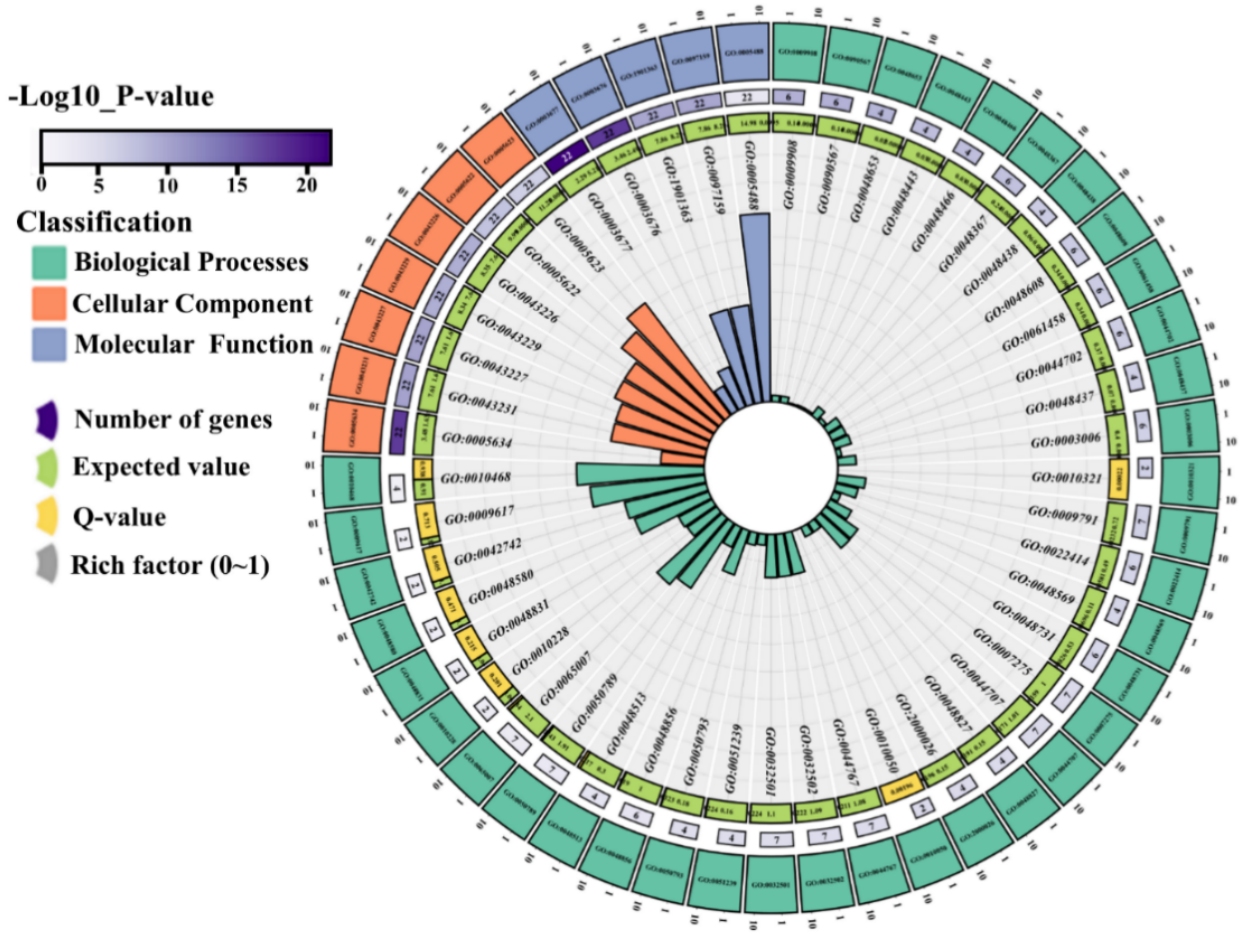

Figure 7. GO enrichment analysis of MtSBP genes. Genes are classified into three categories: Biological Process, Cellular Component, and Molecular Function, shown in different colors. Gene number, expected value, Q-value, and rich factor are represented by color scales, while -log10(p-value) indicates significance levels (white to purple gradient).

lower than those observed in leaves. However, *MtSBP13* and *MtSBP14* remained relatively active under nitrate treatment (R.N). *MtSBP5* also showed elevated expression in roots under ammonium (R.A) and nitrate (R.N) conditions, indicating a potential role in nitrogen uptake and assimilation processes in roots (Figure 8). Notably, *MtSBP17* to *MtSBP22* showed no detectable expression across all tested conditions, suggesting that these genes may be highly tissue-specific, developmentally regulated, or responsive only to untested environmental cues.

4 Discussion

The SQUAMOSA Promoter Binding Protein (SBP)-box gene family, also referred to as SPL (SQUAMOSA Promoter Binding Protein-Like) genes, constitutes one of the most well-characterized plant-specific transcription factor families, with conserved roles across the plant kingdom [5]. All members of this family harbor a highly conserved SBP domain (approximately 75–78 amino acid residues), which contains two typical zinc finger structures (C3H and C2HC) and a nuclear localization signal (NLS), enabling specific binding to the GTAC core sequence of target gene promoters to regulate transcription [2, 5]. With the rapid advancement of genome sequencing technology, SBP-box genes have been systematically identified and characterized in a wide range of plant species, reflecting their fundamental role in regulating multiple aspects of plant growth and development, including vegetative phase transition, floral meristem formation, fruit development, and stress responses [3, 6].

Numerous studies have highlighted the functional diversity and evolutionary conservation of SBP-box genes across different plant lineages. For instance, functional studies in rice demonstrated that *OsSPL10* plays a dual role in trichome formation and salt tolerance, while in tea plant (*Camellia sinensis*), SBP-box genes were shown to respond to abiotic stress and hormone signaling, further confirming the significance of this gene family in regulating plant

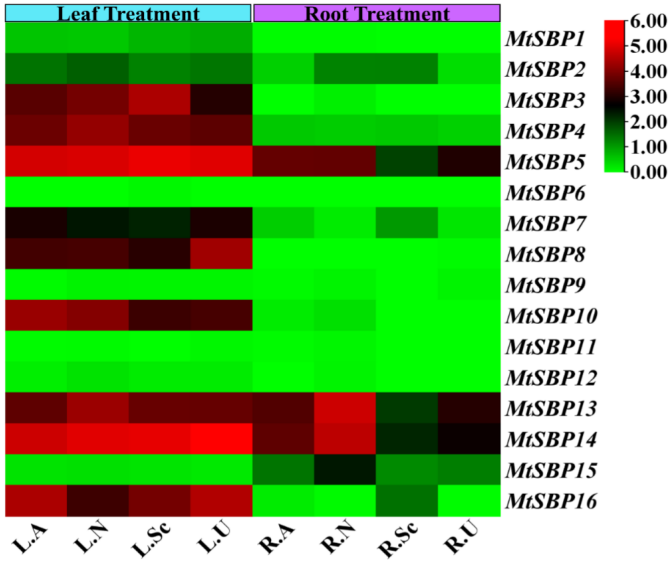

Figure 8. Heatmap showing the MtSBP gene expression profiles in leaf and root tissues under varied nitrogen sources (ammonium, urea, and nitrate) and symbiotic conditions. The color scale represents $\log_{10}(\text{FPKM} + 1)$, where higher values indicate greater transcript abundance.

adaptive growth [12, 28]. In *Citrus clementina*, water deficit during flower induction altered the expression patterns of 15 *CclSBP* genes, establishing a direct correlation between SBP-box gene expression and floral development under environmental stress [29]. In *Petunia*, a model ornamental plant, *PhSPL* genes have been demonstrated to regulate various growth and developmental processes, including vegetative-to-reproductive phase transition, axillary bud development, and inflorescence formation [30]. As a model leguminous plant, *Medicago truncatula* is widely used for functional genomics research due to its small genome size and close evolutionary relationship with other legumes. To fill the gap in the characterization of the SBP-box gene family in *M. truncatula*, we conducted a comprehensive genome-wide retrieval, identification, and assembly using the highly conserved SBP domain as a reference. Our analysis identified 22 *MtSBP* genes in *M. truncatula*, a number higher than that reported in barley (15 genes) [27], and *Arabidopsis thaliana* (17 genes) [6], suggesting a potential expansion of the SBP-box gene family in *M. truncatula* during evolution [25]. This expansion is consistent with previous findings in alfalfa (*Medicago sativa*), a close relative of *M. truncatula*, where 23 *MsSPL* genes were identified, with expansion primarily driven by segmental duplication [26].

To determine the evolutionary relationships among MtSBP genes, we performed a phylogenetic analysis involving SBP-box genes from *M. truncatula*, *A. thaliana*, and *O. sativa*. Our results showed that *M. truncatula*, *A. thaliana*, and *Oryza sativa* contained 22, 17, and 20 SBP-box genes, respectively, which were clustered into seven distinct clades (Figure 2). This clustering pattern reflects a close evolutionary relationship among these species, consistent with the high conservation of the SBP domain across angiosperms [26]. Further domain analysis revealed that all these genes contain highly conserved SBP domains (Figure 1(C)). Synteny analysis further revealed that *M. truncatula* shares more orthologous SBP-box gene pairs with *Brassica* than with *A. thaliana* and *Solanum*. This pattern of syntenic divergence is consistent with previous observations in dicotyledonous plants, where evolutionary divergence among related species leads to variations in orthologous gene distribution [29, 31]. This finding suggests that the SBP-box gene family in *M. truncatula* may have evolved unique adaptive characteristics during the diversification of leguminous plants, which is worthy of further investigation (Figure 5(B)). The predicted nuclear localization of the majority of MtSBP proteins suggests that these genes likely function as transcriptional regulators involved in gene expression control. This finding is consistent with previous studies [26], which also reported nuclear localization of SBP family members, supporting their conserved regulatory role in plant development.

Gene Ontology analysis further indicated that MtSBP genes are predominantly associated with the regulation of biological processes (Figure 7), highlighting their potential involvement in controlling diverse cellular and developmental pathways. This functional annotation aligns with the known role of SBP-box transcription factors in regulating growth, development, and stress-responsive mechanisms [9, 10]. Moreover, the organ-specific expression patterns observed among MtSBP family members suggest functional diversification within this gene family (Figure 8). Such differential expression implies that individual MtSBP genes may have specialized roles in nitrogen utilization and symbiotic interactions, particularly in nodulation and nutrient exchange processes. This functional specialization is consistent with the adaptive evolution of transcription factor families in legumes to support efficient nitrogen metabolism and symbiosis [32].

In summary, using bioinformatics approaches, we systematically elucidated the origin, evolutionary trajectory, and potential functional differentiation of

MtSBP genes in *M. truncatula*. Our findings not only expand the current understanding of the SBP-box gene family in leguminous plants but also establish a theoretical framework for subsequent functional studies of MtSBP genes. Specifically, the identification of gene ontology and gene expression studies provides new insights into the molecular mechanisms underlying *M. truncatula* growth, development, and adaptation. Furthermore, this research serves as a valuable reference for investigating SBP-box genes in closely related leguminous species (e.g., alfalfa) and offers a substantial theoretical foundation for their practical application in plant genetic improvement, such as enhancing stress resistance and optimizing reproductive traits in legume crops [26, 27, 31].

5 Conclusion

In this article, we present the identification and characterization of *MtSBP* genes within the entire genome of *Medicago truncatula* (*barrel medic*). A total of 22 MtSBP genes were classified into seven subgroups based on phylogenetic analysis, which was further supported by conserved domain architecture, motif composition, and gene structure organization. Evolutionary origin analysis, synteny relationships, and gene annotation provided additional evidence for the functional and evolutionary dynamics of this gene family. Organ-specific expression profiling indicated that MtSBP genes are likely involved in nitrogen utilization and symbiotic interactions, particularly in processes associated with nodulation and nutrient exchange. Overall, this study provides a valuable resource for identifying candidate MtSBP genes for future functional validation and highlights their potential roles in legume growth and nitrogen-use efficiency.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

Hadia Hussain served as an Editorial Board Member of the *Plant Innovation Journal* at the time of manuscript submission. To ensure the integrity of the peer-review process, Hadia Hussain was not involved in the editorial handling, peer review, or decision-making process for this manuscript, which was handled independently by another editor. The remaining authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Yu, N., Yang, J. C., Yin, G. T., Li, R. S., & Zou, W. T. (2020). Genome-wide characterization of the SPL gene family involved in the age development of Jatropha curcas. *BMC genomics, 21*(1), 368. [CrossRef]

[2] Peng, X., Wang, Q., Zhao, Y., Li, X., & Ma, Q. (2019). Comparative genome analysis of the SPL gene family reveals novel evolutionary features in maize. *Genetics and Molecular Biology, 42*(2), 380-394. [CrossRef]

[3] Klein, J., Saedler, H., & Huijser, P. (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. [CrossRef]

[4] Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z., Keddie, J., ... & Yu, G. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science, 290*(5499), 2105-2110. [CrossRef]

[5] Preston, J. C., & Hileman, L. C. (2013). Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. *Frontiers in Plant Science*, *4*, 80. [CrossRef]

[6] Yang, Z., Wang, X., Gu, S., Hu, Z., Xu, H., & Xu, C. (2008). Comparative study of SBP-box gene family in Arabidopsis and rice. *Gene, 407*(1-2), 1-11. [CrossRef]

[7] Han, Y. Y., Ma, Y. Q., Li, D. Z., Yao, J. W., & Xu, Z. Q. (2016). Characterization and phylogenetic analysis of fifteen NtabSPL genes in Nicotiana tabacum L. cv. Qinyan95. *Development genes and evolution, 226*(1), 1-14. [CrossRef]

[8] Salinas, M., Xing, S., Höhmann, S., Berndtgen, R., & Huijser, P. (2012). Genomic organization, phylogenetic comparison and differential expression of the SBP-box family of transcription factors in tomato. *Planta, 235*(6), 1171-1184. [CrossRef]

[9] Hou, H., Li, J., Gao, M., Singer, S. D., Wang, H., Mao, L., ... & Wang, X. (2013). Genomic organization, phylogenetic comparison and differential expression of the SBP-box family genes in grape. *PLoS One, 8*(3), e59358. [CrossRef]

[10] Li, J., Hou, H., Li, X., Xiang, J., Yin, X., Gao, H., ... & Wang, X. (2013). Genome-wide identification and

analysis of the SBP-box family genes in apple (Malus × domestica Borkh.). *Plant Physiology and Biochemistry, 70*, 100-114. [CrossRef]

[11] Guo, A. Y., Zhu, Q. H., Gu, X., Ge, S., Yang, J., & Luo, J. (2008). Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family. *Gene, 418*(1-2), 1-8. [CrossRef]

[12] Lan, T., Zheng, Y., Su, Z., Yu, S., Song, H., Zheng, X., ... & Wu, W. (2019). OsSPL10, a SBP-box gene, plays a dual role in salt tolerance and trichome formation in rice (*Oryza sativa L.*). *G3: Genes, Genomes, Genetics, 9*(12), 4107-4114. [CrossRef]

[13] Hanly, A., Karagiannis, J., Lu, Q. S. M., Tian, L., & Hannoufa, A. (2020). Characterization of the role of SPL9 in drought stress tolerance in Medicago sativa. *International journal of molecular sciences, 21*(17), 6003. [CrossRef]

[14] Zhang, H. X., Feng, X. H., Ali, M., Jin, J. H., Wei, A. M., Khattak, A. M., & Gong, Z. H. (2020). Identification of pepper CaSBP08 gene in defense response against Phytophthora capsici infection. *Frontiers in plant science, 11*, 183. [CrossRef]

[15] Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., ... & Fu, X. (2012). Control of grain size, shape and quality by *OsSPL16* in rice. *Nature genetics, 44*(8), 950-954. [CrossRef]

[16] Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., ... & Fu, X. (2015). The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nature Genetics, 47*(8), 949-954. [CrossRef]

[17] Manning, K., Tör, M., Poole, M., Thompson, A. J., King, G. J., Giovannoni, J. J., & Seymour, G. B. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nature genetics, 38*(8), 948-952. [CrossRef]

[18] Yamasaki, H., Hayashi, M., Fukazawa, M., Kobayashi, Y., & Shikanai, T. (2009). *SQUAMOSA* promoter binding protein–like7 is a central regulator for copper homeostasis in *Arabidopsis*. *The Plant Cell, 21*(1), 347-361. [CrossRef]

[19] Xing, S., Salinas, M., Garcia-Molina, A., Höhmann, S., Berndtgen, R., & Huijser, P. (2013). *SPL8* and *miR156*-targeted *SPL* genes redundantly regulate *Arabidopsis* gynoecium differential patterning. *The Plant Journal, 75*(4), 566-577. [CrossRef]

[20] Stone, J. M., Liang, X., Nekl, E. R., & Stiers, J. J. (2005). Arabidopsis AtSPL14, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1. *The Plant Journal, 41*(5), 744-754. [CrossRef]

[21] Limpens, E., Moling, S., Hooiveld, G., Pereira, P.A., Bisseling, T., Becker, J.D., & Küster, H. (2013). Cell- and tissue-specific transcriptome analyses of *Medicago truncatula* root nodules. *PLoS ONE, 8*(5), e64377. [CrossRef]

[22] Wang, H., Lu, Z., Xu, Y., Kong, L., Shi, J., Liu, Y., ... & Han, L. (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*(1), 552. [CrossRef]

[23] Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao, G. (2016). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic acids research*, gkw982. [CrossRef]

[24] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research, 25*(17), 3389-3402. [CrossRef]

[25] Li, J., Gao, X., Zhang, X., & Liu, C. (2020). Dynamic expansion and functional evolutionary profiles of plant conservative gene family SBP-box in twenty two flowering plants and the origin of miR156. *Biomolecules, 10*(5), 757. [CrossRef]

[26] Wang, Y., Ruan, Q., Zhu, X., Wang, B., Wei, B., & Wei, X. (2023). Identification of Alfalfa SPL gene family and expression analysis under biotic and abiotic stresses. *Scientific Reports, 13*(1), 84. [CrossRef]

[27] Tong, T., Fang, Y., Zhang, Z., Zheng, J., Lu, X., Zhang, X., & Xue, D. (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. [CrossRef]

[28] Wang, P., Chen, D., Zheng, Y., Jin, S., Yang, J., & Ye, N. (2018). Identification and expression analyses of SBP-box genes reveal their involvement in abiotic stress and hormone response in tea plant (*Camellia sinensis*). *International journal of molecular sciences, 19*(11), 3404. [CrossRef]

[29] Zeng, R. F., Zhou, J. J., Liu, S. R., Gan, Z. M., Zhang, J. Z., & Hu, C. G. (2019). Genome-wide identification and characterization of squamosa—promoter-binding protein (sbp) genes involved in the flowering development of Citrus clementina. *Biomolecules, 9*(2), 66. [CrossRef]

[30] Zhou, Q., Zhang, S., Chen, F., Liu, B., Wu, L., Li, F., ... & Liu, G. (2018). Genome-wide identification and characterization of the SBP-box gene family in Petunia. *Bmc Genomics, 19*(1), 193. [CrossRef]

[31] He, F., Long, R., Wei, C., Zhang, Y., Li, M., Kang, J., ... & Chen, L. (2022). Genome-wide identification, phylogeny and expression analysis of the SPL gene family and its important role in salt stress in Medicago sativa L. *BMC plant biology, 22*(1), 295. [CrossRef]

[32] Xie, K., Wu, C., & Xiong, L. (2006). Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiology, 142*(1), 280-293. [CrossRef]

Appendix

Table S1. Detailed annotations of all GO terms in MtSBPs genes.

GO.ID	Term	Annotated	Count	Expected	p-value	q-value	Aspect
GO:0009908	Flower development	292	6	0.14	9.80E-10	2.91E-06	P
GO:0090567	Reproductive shoot system development	302	6	0.14	1.20E-09	2.91E-06	P
GO:0048653	Anther development	41	4	0.02	2.30E-09	3.71E-06	P
GO:0048443	Stamen development	62	4	0.03	1.30E-08	1.26E-05	P
GO:0048466	Androecium development	62	4	0.03	1.30E-08	1.26E-05	P
GO:0048367	Shoot system development	509	6	0.24	2.70E-08	2.18E-05	P
GO:0048438	Floral whorl development	123	4	0.06	2.00E-07	1.24E-04	P
GO:0048608	Reproductive structure development	729	6	0.34	2.30E-07	1.24E-04	P
GO:0061458	Reproductive system development	729	6	0.34	2.30E-07	1.24E-04	P
GO:0044702	Single organism reproductive process	782	6	0.37	3.50E-07	1.70E-04	P
GO:0048437	Floral organ development	154	4	0.07	5.00E-07	2.20E-04	P
GO:0003006	Developmental process involved in reproduction	854	6	0.4	5.90E-07	2.20E-04	P
GO:0010321	Regulation of vegetative phase change	3	2	0	5.90E-07	2.20E-04	P
GO:0009791	Post-embryonic development	1536	7	0.72	6.70E-07	2.32E-04	P
GO:0022414	Reproductive process	1032	6	0.49	1.80E-06	5.81E-04	P
GO:0048569	Post-embryonic organ development	225	4	0.11	2.30E-06	6.96E-04	P
GO:0048731	System development	1124	6	0.53	2.90E-06	8.26E-04	P
GO:0007275	Multicellular organismal development	2116	7	1	5.90E-06	1.59E-03	P
GO:0044707	Single-multicellular organism process	2152	7	1.01	6.70E-06	1.71E-03	P
GO:0048827	Phyllome development	309	4	0.15	7.90E-06	1.91E-03	P
GO:2000026	Regulation of multicellular organismal development	316	4	0.15	8.70E-06	1.96E-03	P
GO:0010050	Vegetative phase change	10	2	0	8.90E-06	1.96E-03	P
GO:0044767	Single-organism developmental process	2284	7	1.08	1.00E-05	2.11E-03	P
GO:0032502	Developmental process	2319	7	1.09	1.10E-05	2.22E-03	P
GO:0032501	Multicellular organismal process	2337	7	1.1	1.20E-05	2.24E-03	P
GO:0051239	Regulation of multicellular organismal process	342	4	0.16	1.20E-05	2.24E-03	P
GO:0050793	Regulation of developmental process	380	4	0.18	1.80E-05	3.23E-03	P
GO:0048856	Anatomical structure development	2115	6	1	0.00011	1.90E-02	P
GO:0048513	Organ development	637	4	0.3	0.00013	2.17E-02	P
GO:0050789	Regulation of biological process	4048	7	1.91	0.00046	7.43E-02	P
GO:0065007	Biological regulation	4461	7	2.1	0.00086	1.34E-01	P
GO:0010228	Vegetative to reproductive phase transition of meristem	120	2	0.06	0.00137	2.01E-01	P
GO:0048831	Regulation of shoot system development	126	2	0.06	0.00151	2.15E-01	P
GO:0048580	Regulation of post-embryonic development	204	2	0.1	0.00389	4.71E-01	P
GO:0042742	Defense response to bacterium	217	2	0.1	0.00438	5.05E-01	P
GO:0009617	Response to bacterium	262	2	0.12	0.00633	7.13E-01	P
GO:0010468	Regulation of gene expression	1924	4	0.91	0.00852	9.38E-01	P
GO:0005634	Nucleus	2849	22	3.48	2.30E-18	1.63E-15	C
GO:0043231	Intracellular membrane-bounded organelle	6223	22	7.61	7.00E-11	1.67E-08	C
GO:0043227	Membrane-bounded organelle	6226	22	7.61	7.10E-11	1.67E-08	C
GO:0043229	Intracellular organelle	6821	22	8.34	5.30E-10	7.64E-08	C
GO:0043226	Organelle	6829	22	8.35	5.40E-10	7.64E-08	C
GO:0005622	Intracellular	8171	22	9.99	2.80E-08	2.83E-06	C
GO:0005623	Cell	9224	22	11.28	4.10E-07	3.22E-05	C
GO:0003677	DNA binding	2562	22	2.29	2.20E-22	5.21E-19	F
GO:0003676	Nucleic acid binding	3877	22	3.46	2.10E-18	2.49E-15	F
GO:1901363	Heterocyclic compound binding	8795	22	7.86	1.40E-10	8.29E-08	F
GO:0097159	Organic cyclic compound binding	8796	22	7.86	1.40E-10	8.29E-08	F
GO:0005488	Binding	16770	22	14.98	0.00021	9.95E-02	F

Detailed amino acid sequences of all identified MtSBP proteins

>MtSBP1

MEWNLKAPSWDLGGIEEATLPNIETMEESNRFGVYKMKGEFSVDLKLGQVGNSATDQSPLPLSNDAVVVSKIATPTSSSGSSKRARAMNNNTTLTVSCL
VDGCNSDLSNCRDYHRRHKVCELHSKTPEVTICGLKQRFCQQCSRFHSLEQFDERKRSCRKRLDGHNRRRRKPQPEPITRPAGSFLSNYQGTQLLPFSS
STTMVNSTWSSGLISSCESGRLHINNQHQQVHVVDKQDHFLGSTATTYEGKQLQFLHNDNNNPSLHNETAPLLRTSSKMFCDSLATSVHESPCALSLLS
SSQTRIPDNGLNQMVQQPHSMSHMQPLGLSLHGNNSFESMEGVLVPNGSESDHCSSLYNMGSDGSQGNDAPQLFPYQWE

>MtSBP2

MEWNVKSPGQWDWENLFFLNSKAAETHRLQSTDWSMEEDREINVGMLIPSGGSGYSVSKLMHASSSRSSKSASNNSSSNEDSKTSMLTQEGSPDNSTGK
KESSKGDPIETSPAAEPLLTLKLGKRLYFEDVSTGSHSKKASSSAVPLLCGKKGKSSSQNMLNPSCQVEGCGLDLSFAKDYHRKHRICDSHSKSPVVVV
AGLERRFCQQCSRFHDLSEFDDKKRSCRRRLSDHNARRRKPQPEAVKLNPSALSSSPYDGRQAMGPFAFPKNTSNLAWPDMPNSKLPQTKDFMLKTPKN
FNKIVTMLSDDSSGHFISKGKGTNIAVPGLEDPNTLSDPSATQDVNRALSLLSTNSWGAYDTKPPSFVHSNRTTGTPQYATAQRSPFSSPEYWHTDQHQ
ASSSACISFSGYDNSNRYQDFQLFSEPYESSFPCNQLD

>MtSBP3
MDGSWGEGKRIYEYREENEYEVEVEIEEEEDVSYGDDEKRKRVVTDHLYNKKGSKAGGSVTPSCQVDNCNADLSAAKQYHKRHKVCENHSKAHSVLISE
LQQRFCQQCSRFHEVSEFDDLKRSCRRRLAGHNERRRKSASEYH

>MtSBP4
MESWSFDSRDKGYVSNDTLGSRTKGSILGWDLKTPSSFLPQQNIENDNNNHGFEELGFHGMLGKQLSNVVDDDDDVVVGSKIVTSNSFVMATPNAFSER
EQQHFNSKHSNSIGDTNGSNSLIDLKLGRFGDHRDGIGTPFSKGTTILSSSGSSTPSKRVRSSAIHSQIAYCQVYGCNKDLSSCKDYHKRHKVCEVHSK
TSKVIVNGIEQRFCQQCSRFHLLAEFDDGKRSCRKRLAGHNERRRKPQVGIHSGRAGRLLQSYGDSRFQGTMLTSASFICQDILPGEVFSSEKCGNSNW
WRPIKAEGGTGFRPLSSIPITNAHPQSRTLFPSYNNKQFPFLQENGATSATGSIFCENNSQYPPIQGGSNSGSQSLFQDTSLGSEDFNVFETESAVQGL
SGVSDGGCALSLLSSQSQNSSSQSTGIPMPGHLVIPSSHSHYSMSHVSESGVSDRFPSELNHTDGSHLSPILISNNNEIAHFELVDGVFQGSDFVNVKD
RLSCEDGTTIDLLQLSSQLQRVEHQRRALQVKRENDSSCSLQIT

>MtSBP5
MDADFEGKNQHLYEAVVPEMKGAVGRGSKDWDLNDWRWDGDLFTAKQLNSVPTDCRNRQFLPEIRENVDVSNNLISGEGSRELEKRRRGFGGEGLEMND
DFGSLNLNLGGQVYPIMDGEEKSGKKTKIAPVPTSNRAVCQVEDCRADLSNAKDYHRRHKVCDVHSKASKALVGSVMQRFCQQCSRFHVLQEFDEGKRS
CRRRLAGHNKRRRKTHPDTAVVNGGSPNEERGSSYLLMSLLRILSNMHSNGPDHTRNLDGLSHLIGNLTSLAGTFNGRNIASLLEGPQELVKAGTSGAA
QNVPNSNPNGAEPSRPDSSIEMTNGLIHQDPPESRLQCATVPANHLTQKCIPSSSVGVGCLKPPLIPQSSNLVPSRGSLPPRPVATETTVGRNRLCNID
LNNVYDDGQDYVENPENSNPPLALGVESRDHSSFVQYESLKSSPPQTSRNSDSTSTQSPSSSSGEGQSRTDRIVFKLFGKDPNDIPHVLRSQVLSWLSN
SPTEIESYIRPGCIILTICLRLENSAWDELCYNLGPSLRKLLAASNDSLWRTGWIYTRVQHSVAILYNGQLVLDVPSRLGSPQNCQILCIKPLAVSANE
DVKFTVKGLSLFLSSARLLCALEGKYLVEDMCYDLIDGADAAIGHHELQTLSFSCHIPNMTGRGFIEVEDNSLSSCSFPFIVAEQEICSEICSLETIIE
AAETADDIQIKAKLMEEKTRAMNFVQEMGWLLHRIRIKFRLGPMTPVQDRFHLNRYTWLVGFSMDHDWCAVMKKLLDTIFEGEVDTGEHISAELALLNM
GLLHKAVKRNCRPMVELLLNFVPVKASDGGDSKEMQVNKVPDGFLFRPDTVGPAGLTPLHVAASMNGYETVLDALTDDPGMVGIEAWKSAKDNTGLTPN
DYASLRGHYSYIQLVQRKTSKSSQTQHVLDIPGTLVDGNTMKQSDGHKSSKVLSLHTEKIATTAIPNHCGICQQKLAYGSVGGMRRALVYRPAMLSMVA
IAAVCVCVALLFKSSPRVSYVFQPFSWESLDYGSI

>MtSBP6
MEGKQTSVEKVWKDIIFEEIEVEEEEEGGVEDDEKKKGVCVSGRKEPSRGGRFSSPICQVESCGADLTFSKRYHRRHKVCEVHSKASVVVVAGMRQRFC
QQCSRFHELAEFDESKRSCRRRLARHNERRRKSTAGTCTEGSSTGQKNNDGDWRQIHMNITSSSGHDFLNFR

>MtSBP7
MEWSNLKAPSWDLTEMDQGTLNNIETMDGSIRRFGDYRTKGDFSVDLKLGHVGNSSTESGLTKSKDVGGFTKMTSSSSSTSGSSKRARAINNGTQIVSC
LVDGCQADLSNCRDYHRRHKVCELHSKTAQVTIGGHKQRFCQQCSRFHALEEFDEGKRSCRKRLDGHNRRRRKPQPEPITRSSSFLSNYQGTQLLPFSS
SHVYPSTAMMNPTWVGGDVRLHSHNQHQQMHHLVDKQDLFLGTTSSTSSYKEGKQVSFIQNNQPTAAATMNNQPLPASSMSRTFSRTNPFSERYKMFCD
NNTSCALSLLSSPVPQTHDHPENGLNQMVNTHSSFMQPLGLSLHDHSLGSVDSVLGSNDQSGHCSSMYNMGSSESHSQGNNNEASPLFPFQWD

>MtSBP8
MLDYEWGNPSNIMLTGNEDNSGAAATDQAHRQIFDHYASQAMLSDNYLNGPGGGPTIDINSGDFTHHHNQFSPHNQQHHNIHQSFFDPRAFHGGSSTAS
SYPPPQPQPPPSMLSLDPLPGHGHSPGFLLVPKSEDVNRPIDFVGSRLGLNLGGRTYFSSEDDFVTRLYRRSRPPEPGSTGSSNSPRCQAEGCNADLSQ
AKHYHRRHKVCEFHSKAATVVAAGLTQRFCQQCSRFHLLSEFDNGKRSCRKRLADHNRRRRKTQHPNTQDIHKSHNTLDSSATRSPPESGTQSTSSVTV
AVSPPDYFRQRSYQTPSPSTTSSSMFFSTG

>MtSBP9
MDLNAKSPSSQWEWDHLPLPDTKATENKKLQPPNWSIKLDRENNVWLFDTPTGSGYSGSELIPGSTSGSSKSDSISSSPTRDISKTCKFAFESSHDDSN
GKIELPKEGPIEASNAPMLSCVSSEPLLNLKLGKRMKFEDVSSLVFGRKCKSNDQNLQCPPLCQVEGCGLDLSSAKHYYRKRRVCVDHSKSPMVVIDGL
ERRFCQQCSRFHDLFEFDGKKKSCRRRLSQHNARRRNHSRKTAQSSQSALSSSPCDGKQQMNPFAYSKTAINLALQNIQNSEFPRIKDFPLKYAKVNIG
TPSFVTMLSDDSDVHFTSKVLETKSINAGVEDFISFSDTDATEDFTCASSLLSTSPWDSYATKSVSLVHSDRTTSTAQAITHTTSQRIPLTSSEYWHGD

LDQHVDSSMWISNSRL

>MtSBP10
MEARSIERKRTLMMEKMRRNNMVEEEMENEVEEEEGDVSLTEEEKKKGVGGGRRGGGGGVSPPCCQAERCGADLTDAKRYHRRHKVCEFHSKAPVVVVA
GMRQRFCQQCSRFHDLVEFDESKRSCRRRLAGHNERRRKTNPETANEGSSHSKGQHQPKETQCRIQMNLPGSSGYKSFNIR

>MtSBP11
MDLNAISPSSQWDWEHQSFFNAKATEYPKLQPPSWSIEQDREINVGLFDTPGGSGIFGSASRSSKSASNSSSLNRNNSKTCIFACGNSHTPELSSVSGE
PLLTLLGKRVYIEDVCPKSDSKNLSFSRDLASSLSIGKKCKFNSQNLQFPRCQVEGCGLDLSSAKDYHRKRRVCESHSKSPMVVIDGMVRRFCQQCGRF
HNLAEFDEKKRSCREQLSRHNARRRKHQPEALQSSQSVLSSRCDGKQQMSPFANSKTATNLAWQNMQNSKLPQAKDFLLKPAKDNANTQGTVTMFSDDS
IVHFTSKVLATKSIDRGVEDFITVSDTNDEQDFICALSLLSTNPWDSYATKSMSLEHSNRPTSAVQAVTHPMSQRMPLA

>MtSBP12
MDLNAISPSSQWEWDHLPLLDTKATENQKLQPPNWTMELDPEINFGLYDTPVGSGSSGSDLIHGSAFNAPKFSCVSGEPLLNPRLGKRMYFEEVCPESD
SKNLSFSRDLMSSLILEKKCKSNGQNLQCPPHCQVEGCGLNLSSAKDYHCKRRVCESHAKSPMVVIDGLERRFCQQCSRFHDLFEFDGKKKSCRRQLSN
HNARRRKYHRQAVQSSQSALSYSRSDGKQQMSPFTNSKTATNLAWQNMHNSKLPQAKDFLLKPSKDNIDTPSTVTMLSHDSNFHFTSKILATKSVNPGV
EDFMNFSDTNATQDFTCALSLLSTNQWDSYATKSISQEHSNRPTSPFQATTHAMSQCIPLASSEPCCGHHDQYVNSNIWISNSADSNHFQKFQLFREPY
EL

>MtSBP13
MDWDWKEFAWDPSGFDEELKNEGDSMDLRLGDQASDSLEKSVHDSGKEDSKAVSSSLSPSGSLKRSRLQNGSQSMICSVDGCNSDLSDCREYHKRHRVC
EKHSKTPVVLVGGKQQRFCQQCSRFHSLAEFDDVKRSCRKRLDGHNKRRRKPQPPSLFMAAEKFMYNYKGPRILHFGSPQTYVNPIMRNIWPATSITEA
ESGYDHQRLLYRIDKHKQDKGHPHWQEILPKSGDVNKAAPGTPISHPIRGAVGSSAGEKGGRKLSSDGKPGSFDSGCALYLLSTLQTQSSELSLMQSSI
NSPTQSSSGTLHFDATKEYSYSGKVKDKPNGPVFVLDGNTTNLHCNGMLQMGPNGLIENDNSMTLPFFWE

>MtSBP14
MEAFQLYGNGGGSSDLRAMGKNSKEWDLNNWKWDSHLFIATSKLTPVPEHRQFLPIPVGGGGGGGGSNSNSSSSCSEQLDLGICQVKEGERKRRVIVVE
DELGLGLNKEGGNLSLNLGGGVATWEGNNGKKSRVAGGGSSSRAFCQVEDCRADLNNAKDYHRRHKVCEIHSKASKALVGNAMQRFCQQCSRFHLLQEF
DEGKRSCRRRLAGHNKRRRKTNQDAVPNGSSPNDDQTSSYLLISLLKILSNMQPDRSNQTADQDLLTHLLRSLANQNGEQGGRNLSNLLREPENLLKEG
SLSGKSEMVSTLVTNGSQGSPTVTVQNQTVSISEIQHQVMHSHDARVADQQTTFSAKPGVSNSPPAYSEARDSTAGQTKMNDFDLNDIYIDSDDGIEDI
ERLPVTTNLGASSLDYPWMQQDSHQSSPPQTSGNSDSASAQSPSSSTGETQNRTDRIVFKLFGKGPGDFPLVLKAQILDWLSHSPTDIEGYIRPGCVVL
TIYLRQAEVVWEELCFDLTSSLNRLLGVSDDDFWRTGWVHIRVQHQMAFIFNGQIVIDTPLPFRSNNYGKILSVSPIAIPSSKTAQFSVKGINLTRPAT
RLLCALEGNYLDCEDTDEPMDQCSKDLDELQCIQFSCSVPAMNGRGFIEIEDQGLSSSFFPFIVVEEDVCSEICVLEPLLESSDTYPDNEGAGKIQAKN
QAMDFIHEMGWLLHRRQIKSSVRLNSSMDLFPLDRFKWLMEFSVDHDWCAVVKKLLNLMLDGTVSTGDHTSLYLALSELGLLHRAVRRNSRQLVELLLR
FVPQNISDKLGPEDKALVNGENQNFLFRPDAVGPAGLTPLHIAAGKDGSEDVLDALTNDPCMVGIEAWNSARDSTGSTPEDYARLRGHYTYIHLVQKKI
NKSQGGAHVVVDIPSIPTKFDTSQKKDESCTTFQIGNAEVKKVRKDCKLCDHKLSCRTAVRKSFVYRPAMLSMVAIAAVCVCVALLFKSSPEVLYIFRP
FRWESLDYGTS

>MtSBP15
MDSGGNSSSEESSLNGLKFGQRIYFEDTALTAASAAAASTTIAAGSPSSSGSKKGRGGSVQHSQPPRCQVEGCKLDLTDAKAYYSRHKVCSMHSKSPTV
TVSGLQQRFCQQCSRFHQLAEFDQGKRSCRRRLAGHNERRRKPPPSSLLTSRFARLSSSVFGNSDRGGSFLMEFASNPKHSLRNSPGNQTTAIGWPWPG
NTESPSSNLFLQGSVGGTSFPGARHPPEETYTGVTDSNCALSLLSNQTWGSQNTEPSPGLNNMLNFNGTPMTQLGTSSHGVAMHQIPNNYEVVPDLGRG
HILHPLGSQHSGELDLLQQGRRHYMDVEHSRAYESSQWSL

>MtSBP16
MDFGGNMFCLNNRDQSNTNNNGNSNGFTWCSSGSTTTGTAWNMNTFNVNEVNNSVSAANRTEAGLANALMYLPQNEGGGRQYQSAYGGGVSRGHMMPDP
HLTCLKLGKRHYFEDVSGGGGVMGEKKGKGGYCGGGKTAAVGGYTAVTRCQVDGCNVALMNAKEYHRRHKVCEMHSKAPKVVVLGLEQRFCQQCSRFHV
VSEFDDSKRSCRRRLAGHNERRRKSSHDSVGRNSSQGGCALSLLSSRTDSWLSPADLSVRCSAALSELIAENRASIMARQYVSDRDWHLQHHAVEDYKE
IQSESNYFPQHMFPQTH

>MtSBP17
MAALQQVKRKVEKGKVYRGGRSSGVVRTLGDAAAPADQYVAPSQGTRFLHYFNISVTSIQFVVMESWSYIPEERSYLFSDEMDFSLDAFMRSRKPLVEW

ENKSSCNFERDGFNSDREVVKSMEFVDLGFPMESSRTVRGSVQTSSCELDSNNSSKRENSSVHVIAFDSSFAEEDSEAKHLSSRVESKNHDSSLIDLKL
GRLVDCRGGGSDRDGKTFTSESMPIHQNVLAKRSRTSSLPALAPVCQVYGCNMDLSSSKDYHKRHKVCDVHSKTAKVIVNGVEQRFCQQCSRFHLVAEF
DDGKRSCRRRLAGHNERRRKPQFDYMTSKQHKILQSYQGTNNFSFQDIFQSGILFPEKHDQISQSGHIKLEEDPICSPQLEAPATHGHELSSHALSLLS
AQSQNPLHQAAGNPSATSSTFFRDIRTQDRGDQVSETPLMTSSMNKHERNESFPCGINSKENIKSEHGGTVDLFQLSSNLQRVEKQRNSVLVKWENEDC
CFSNV

>MtSBP18
MSSGFVPSPSSPPNSSTEFLDGLKFGQKIYFEDATSTVVVPASQTKPNAIGVGGSSSSSSSSGMKRGNNHPPRCQVEGCKVDLSGAKAYYSRHKVCCMH
SKSPSVVVSGLEQRFCQQCSRFHQLPEFDQGKRSCRRRLAGHNERRRKPQPSSFLTSRSPKLSSPTFGNNGKGNNDFLMECASYQKISMGNALSTPRSA
KPVPGNQATQFTWQNNSSTTSDFFLQGTNFHAPPTRHLPIMDNYNEVTDSSCALSLLSNQTWCSRNTTPPSVEMNNLLNFNGSLTTQSSQVVPDYQLQT
NASSWFLKGVDSRNCLSPEAVPDLGLGQNSQAIHSDLHGELDVSQGKRHYWPL

>MtSBP19
MEKVAPPLLPLHPPMLSSHQFYDSSNTKKRDLLSSYDVVHIPNDNWNPKEWNWDSIRFMTAKSTTVEPQQVEESLNLNLGSTGLVRPNKRIRSGSPTSA
SYPMCQVDNCKEDLSKAKDYHRRHKVCEAHSKASKALLGNQMQRFCQQCSRFHPLVEFDEGKRSCRRRLAGHNRRRRKTQPDEVAVGGSPPLNQVAANL
EIFNLLTAIADGSQGKFEERRSQVPDKEQLVQILNRIPLPADLTAKLLDVGNNLNAKNDNVQMETSPSYHHRDDQLNNAPPAPLTKDFLAVLSTTPSTP
ARNGGNGSTSSADHMRERSSGSSQSPNDDSDCQEDVRVKLPLQLFGSSPENDSPSKLPSSRKYFSSESSNPVDERTPSSSPPVVEMNFGLQGGIRGFNS
NCISTGFGGNANKETSQSHSCTTIPLDLFKGSKSNNMIQQSSSVQSVPFKAGYASSGSDYSPPSLNSDTQDRTGRIMFKLFDKHPSHFPGTLRTQIYNW
LSTRPSDLESYIRPGCVVLSIYASMSSAAWVQLEENFLQRVDSLIHNSDSDFWRNGRFLVYSGSQLASHKDGRIRMCKPWGTWRSPELISVSPLAIVGG
QETSISLKGRNLSAPGTKIHCTGADCYTSSEVIGSGDPGMVYDEIKLSGFEVQNTSPSVLGRCFIEVENGFKGNSFPVIIANASICKELRPLESEFDEE
EKMCDAISEEHEHHFGRPKSRDEALHFLNELGWLFQRERFSNVHEVPDYSLDRFKFVLTFSVERNCCMLVKTLLDMLVDKHFEGEGLSTGSVEMLKAIQ
LLNRAVKRKCTSMVDLLINYSITSKNDTSKKYVFPPNLEGPGGITPLHLAASTTDSEGVIDSLTNDPQEIGLKCWETLADENGQTPHAYAMMRNNHSYN
MLVARKCSDRQRSEVSVRIDNEIEHPSLGIELMQKRINQVKRVGDSCSKCAIAEVRAKRRFSGSRSWLHGPFIHSMLAVAAVCVCVCVLFRGTPYVGSV
SPFRWENLNYGTM

>MtSBP20
MGERLGAENYHFYGVGGSSDLSGMGKRSREWNLNDWRWDGDLFIASRVNQVQAESLRVGQQFFPLGSGIPVVGGSSNTSSSCSEEGDLEKGNKEGEKKR
RVIVLEDDGLNDKAGALSLNLAGHVSPVVERDGKKSRGAGGTSNRAVCQVEDCGADLSRGKDYHRRHKVCEMHSKASRALVGNAMQRFCQQCSRFHILE
EFDEGKRSCRRRLAGHNKRRRKTNQEAVPNGSPTNDDQTSSYLLISLLKILSNMHSDRSDQPTDQDLLTHLLRSLASQNDEQGSKNLSNLLREQENLLR
EGGSSRNSGMVSALFSNGSQGSPTVITQHQPVSMNQMQQEMVHTHDVRTSDHQLISSIKPSISNSPPAYSETRDSSGQTKMNNFDLNDIYVDSDDGTED
LERLPVSTNLATSSVDYPWTQQDSHQSSPAQTSGNSDSASAQSPSSSSGEAQSRTDRIVFKLFGKEPNEFPLVLRAQILDWLSQSPTDIESYIRPGCIV
LTIYLRQAEAVWEELCCDLTSSLIKLLDVSDDTFWKTGWVHIRVQHQMAFIFNGQVVIDTSLPFRSNNYSKIWTVSPIAVPASKRAQFSVKGVNLMRPA
TRLMCALEGKYLVCEDAHESTDQYSEELDELQCIQFSCSVPVSNGRGFIEIEDQGLSSSFFPFIVAEEDVCTEIRVLEPLLESSETDPDIEGTGKIKAK
SQAMDFIHEMGWLLHRSQLKYRMVNLNSGVDLFPLQRFTWLMEFSMDHDWCAVVKKLLNLLLDETVNKGDHPTLYQALSEMGLLHRAVRRNSKQLVELL
LRYVPDNTSDELGPEDKALVGGKNHSYLFRPDAVGPAGLTPLHIAAGKDGSEDVLDALTNDPCMVGIEAWKNARDSTGSTPEDYARLRGHYTYIHLVQK
KINKTQGAAHVVVEIPSNMTESNKNPKQNESFTSLEIGKAEVRRSQGNCKLCDTKISCRTAVGRSMVYRPAMLSMVAIAAVCVCVALLFKSSPEVLYMF
RPFRWESLDFGTS

>MtSBP21
METRRSEGKRSLKYKEDHEEEEEEEEDTDFEEEEDGRRKRVVTDLYSKRSSKKAGSSNIPPCCQVENCDADLSEAKQYHRRHKVCEYHAKAPAVHIAGL
QQRFCQQCSRFHGLSEFDDAKRSCRRRLAGHNERRRKSAIDYQGE

>MtSBP22
MESWSYISEEKGSEELGFCEMFGKQYSDDFNFIGNVPSKKLDGKFDDPNEFSVKGDTNFKLSNSVVESNGWDSLIDLKLGRFDDHGNASIDVAFSKGAA
PILSSCESSTPPKRVRVHSMTAYCQVYGCNKDLSSCKDYHKRHKVCEVHSKTPIVIVNGIEQRFCQQCSRFHLLSEFDDGKRSCRKRLAGHNERRRKPQ
GIHSANSGRLFQPFGDIGFQGTKPPAASFLCPEVFQSGFWRPMKAEHEAGFRHLSSVPVTNGHLQSRSPFPSYNEKQFPFLHENVATSTTGSKFSENNS
HYAHAIGYASLGNEDFNVFHTAPSIQGLSGISDSCALSLLSSQSNNTSSQSSEIPFDNPLVIPSSHSHVYSIRSQATSSSRVSDRFLSKLNPADGSHLS
PILVSDNNDIVNFEMEDGIFHDSDFVNARDCLSCEDDATIDLLQLSSQLQRV