

Article

# Dynamic Expansion and Functional Evolutionary Profiles of Plant Conservative Gene Family SBP-Box in Twenty Two Flowering Plants and the Origin of miR156

Jing Li <sup>1,2</sup>, Xiaoyang Gao <sup>1</sup>, Xuan Zhang <sup>1,2</sup> and Changning Liu <sup>1,3,4,\*</sup>

- <sup>1</sup> CAS Key Laboratory of Tropical Plant Resources and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla 666303, Yunnan, China; lijing2@xtbg.ac.cn (J.L.); gaoxiaoyang@xtbg.ac.cn (X.G.); zhangxuan@xtbg.ac.cn (X.Z.)
- <sup>2</sup> College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
- <sup>3</sup> Center of Economic Botany, Core Botanical Gardens, Chinese Academy of Sciences, Menglun, Mengla 666303, Yunnan, China
- <sup>4</sup> The Innovative Academy of Seed Design, Chinese Academy of Sciences, Menglun, Mengla 666303, Yunnan, China
- \* Correspondence: liuchangning@xtbg.ac.cn; Tel.: +86-691-8713009

Received: 2 March 2020; Accepted: 4 May 2020; Published: 13 May 2020



**Abstract:** Conservative gene families in plants, which are closely related to innovations in flowering plants, have long and complex evolutionary histories. Here, we used the *SQUAMOSA* promoter-binding protein (SBP-box) gene family as an example to study conservative gene families in flowering plants. In total, 11 groups, including nine angiosperm-conservative groups and two monocot- and eudicot-specific groups, were identified. Among the nine angiosperm-conservative groups, four are conserved in all land plants and the remaining five are angiosperm-specific. The five angiosperm-specific groups exhibit structural and functional diversity and evolved together, along with the evolution of flowering plants. The expansion of *SBP* genes was affected by miR156, and the miR156-regulated *SBP* genes tend to retain more copies. Our results reflect the dynamic evolutionary process of the different groups, with the identification of two genetic lines via synteny analyses. In addition, miR156 showed a close evolutionary relationship with *SBP* genes, suggesting that it may originate from face-to-face tandem duplication of *SBP* genes. *SBP* genes without an miR156 binding locus are usually functionally conservative or housekeeping like, belonging to the terrestrial-conservative group. In contrast, *SBP* genes with miR156 binding sites are selected by angiosperms to regulate more complex physiological processes.

**Keywords:** gene family; flowering plants; SBP; synteny; duplication; miR156; coevolution

## 1. Introduction

Angiosperms (flowering plants), with at least 300,000 species in multiple families, are the most diverse and latest-diverging group in the plant kingdom [1–4]. A gene family is a group of genes with the same characteristics, which usually contain a specific domain, motif, or special structural composition. Conservative gene families consist of genes whose sequences are conserved over time and are widely observed in many species. There are many conservative gene families in flowering plants. They appear in a large range of taxa and often play critical roles. Multiple types have also been found. For instance, some gene families are the main catalytic carriers of important biochemical processes in plants. The P450 super family has a large number of plant members and is the key enzyme

for chemical synthesis [5,6]; some conservative families are important transcription factors (TFs), which, in plants, are structurally diverse and involved in many development processes and responses to abiotic and biotic stresses. For instance, the WRKY superfamily is involved in pathogen defense [7], senescence, and trichome development [8,9], and the MADS-box gene family is strongly correlated with the origin and evolution of reproductive structures such as flowers and ovules [10].

*SQUAMOSA* promoter-binding protein (SBP-box) (*SBP*) genes encode plant-specific TFs (SBP or SPL) that were first identified in *Antirrhinum majus* [11]. They were named for their ability to regulate the expression of *SQUAMOSA* (a kind of MADS-box) genes. Several characteristics make SBP-box a representative conservative gene family in plants. First, group members have a long evolutionary history. *SBP* genes have been identified in green algae, bryophytes, and angiosperms [12–14]. In addition, the SBP-box gene family has evolved many groups [12–15]. In a previous study, we showed that the *SBP* genes of Euphorbiaceae can be divided into 10 groups [16]. The SBP composition of different plant taxa is usually different [12–14]. In addition, *SBP* genes can be regulated by miR156 [17,18], and the evolution of SBP may be influenced by it. Finally, the SBP-box gene family is functionally diverse. Although it is highly conserved in plants, *SBP* genes in different groups may be specific to certain plant taxa [12–14]. These characteristics suggest that the SBP-box gene family is a very typical conservative gene family. Therefore, it would be helpful to study this family to understand other conservative gene families in angiosperms.

Over the last two decades, tremendous efforts have been made to explore whole-genome duplication (WGD) events. WGDs accelerate the birth and death of plant genes and have long been recognized as an important source of biochemical, genetic, and evolutionary novelty [19–21]. The moves from aquatic to terrestrial, from sporogenesis to seed reproduction, from gymnosperms to angiosperms, and from monocots to eudicots have typically been accompanied by WGDs. For example, all seed plants experienced WGD event  $\zeta$  [22]; all angiosperms experienced WGD event  $\epsilon$  [22]; and two important clades in angiosperms, namely, monocots and eudicots, both experienced WGD events ( $\gamma$  for eudicots and  $\tau$  for monocots) [23–26]. However, it seems that flowering plants appeared abruptly and were soon the most successful organisms on the planet [1,3,19]. Harsh environmental conditions are an external factor that promotes plant evolution. In the Cretaceous, the Earth was dry and hot, with high levels of oxygen. However, around the Cretaceous–Paleogene (K–Pg) boundary, global cooling and darkness were the two main stresses. The following question remains: what role did conservative gene families play in the success of flowering plants? WGD events promote gene evolution and provide us with the clues we need to understand plant gene families.

The evolution of a gene family is a dynamic process. As changes occur in the physiological processes and phenotypes of plants, the structures and functions of a gene family may also be undergoing constant change. Typically, ancient WGDs in plants have been accompanied by key innovations [20], including the formation of vessels and flowering mechanisms, and the formation and evolution of many flower organs [21]. In a specific gene family, both the composition and function may change after a WGD [21,27–30]. The influence of different WGDs in different periods on a certain gene family may differ. At least 10% of all floral regulators in *Arabidopsis thaliana* can be traced back to the common ancestors of the angiosperm [31]. The TF gene families that are involved in responses to biotic and abiotic stresses tend to have retained large numbers of copies during the three waves of angiosperm WGDs [32]. The evolutionary trajectories of the different taxa of angiosperms are diverse. All eudicots experienced  $\gamma$  [23,24,26], while all monocotyledons experienced  $\tau$  [25]. After  $\tau$ , Gramineae experienced  $\sigma$  and  $\rho$  [25], unlike other monocots. Angiosperms also feature diverse physiological and reproductive processes and phenotypes among different taxa. Some plants are androgynous and others are diclinous; some plants blossom each year and others blossom only once in their lifetime. The gene families associated with these processes or phenotypes may experience a distinct evolutionary trajectory for different taxa. Therefore, the evolutionary process of a specific conservative gene family may be in a synchronous relationship with the physiological and phenotypic evolution of flowering

plants. Hence, the evolutionary patterns of a specific conservative gene family during typical periods of its evolution need to be studied.

In this study, we comprehensively analyzed the *SBP* genes in 22 angiosperms and investigated their dynamic evolution to better understand the SBP-box gene family of flowering plants. First, 22 sequenced plant genomes, which represent major lineages of flowering plants, were used to predict *SBP* genes, and phylogenetic analyses were performed to obtain a comprehensive SBP-box gene family classification. Then, an evolutionary profile of *SBP* genes was created in phylogenetic analyses of angiosperms and other typical control species, providing a map of the evolution of the *SBP* genes. Moreover, the dynamic evolution of different SBP groups and the correlations between SBP and miR156 were assessed based on synteny relationships, and a possible model of the emergence of miR156 was developed. Finally, we performed a preliminary functional exploration of different groups of *SBP* genes by analyzing their upstream regulatory sequences, along with their tissue-specific and stress-treatment expression profiles. We performed comprehensive analyses of the angiosperm SBP-box gene family and its dynamic evolution to provide a reference for the study of other conservative gene families in plants.

## 2. Materials and Methods

### 2.1. Data Sources

The genomic and proteomic sequences of *Brachypodium distachyon*, *Ananas comosus*, *Sorghum bicolor*, *Panicum hallii*, *Oryza sativa*, *Musa acuminata*, *Setaria italica*, *Elaeis guineensis*, *Zea mays*, *Brassica rapa*, *Camelina sativa*, *Phaseolus vulgaris*, *Vigna angularis*, *Cajanus cajan*, *Medicago truncatula*, *Solanum lycopersicum*, *Capsicum annuum*, *Nicotiana tabacum*, *Prunus persica*, *Malus domestica*, and *Prunus mume* were taken from the NCBI database [33]. The genomic and proteomic sequences for *A. thaliana* were downloaded from the TAIR database [34]. The SBP protein sequences for *A. thaliana*, *O. sativa*, *Physcomitrella patens*, *Selaginella moellendorffii*, *Picea abies*, *Amborella trichopoda*, and *Chlamydomonas reinhardtii* were obtained from the PlantTFDB database [35]. The pre-miR156 sequences and mature-miR156 sequences for *A. thaliana* and *O. sativa* were obtained from the miRBase database [36].

### 2.2. SBP Gene Identification and Characterization

Both a genome-wide Hidden Markov Model (HMM) search and local Blastp were performed to identify possible *SBP* genes [37,38]. The HMM model of PF03110 was downloaded from Pfam [39]. A local version of the HMMER v3.2.1 program was used to perform the hmmsearch, and the E-value was set to the default value. The known sequences of *A. thaliana* SBP proteins were used as queries in the Blastp process, and the E-value was set to  $1 \times 10^{-10}$ . The results obtained by the two methods were combined and left for later analyses after removing redundancy. To ensure the authenticity of the SBP domain, all retrieved sequences were tested using the SMART v8.0 tool and the NCBI Conserved Domain Database (CDD) [40,41]. Sequences with incomplete SBP domain structures were completely discarded. Some basic physicochemical parameters of SBP proteins, such as the protein length, molecular weight (Mw), and isoelectric point (pI), were predicted using the ExPaSy Proteomics Server [42]. Additional conserved motifs outside the SBP domain of all SBP proteins were identified by MEME v5.1.0 [43]. The settings for the minimum width, maximum width, maximum number of motifs, and minimum sides were specified as 6, 150, 15, and 2, respectively.

### 2.3. Phylogenetic Analyses and Detection Orthologous Genes

The SBP protein sequences of these 22 species were aligned using MUSCLE v2.0 [44], and the alignments were used for phylogenetic analyses. Both the Bayesian method in MrBayes v3.2.7 and the maximum likelihood method in PAUP\* v4.0 [45,46] were implemented. The orthologs among these species were detected using OrthoFinder v1.0.6 [47].

#### 2.4. MIR156 Prediction

MiRNAs were predicted using a homologous search based on mature miRNA sequence conservation in plant species. The mature miR156 sequences were downloaded from miRBase [36]. A total of 470 mature sequences were deposited in miRBase (Release 22.1). After all redundancies were removed, the remaining unique sequences were kept as miRNA reference sequences, and they were subjected to Blastn against the genome [38]. Output sequences > 18 nt in length and with less than three mismatches compared to the miRNA sequences were collected for further analyses. The predicted miRNAs of mature sequences, along with sequences 200 nt upstream and downstream, were used as assumed miRNA precursor sequences. If either the upstream or downstream sequence was less than 200 nt, the entire available sequence was used as an miRNA precursor sequence. Sequences in protein-coding genes were removed, and the remaining sequences were subjected to RNA secondary-structure prediction using MFold v3.6 [48]. All sequences met the criteria given in a previous study and could be used for subsequent analyses [49]. The sequence logo was created using WebLogo v2.8.2 [50].

#### 2.5. MiR156 Target Prediction

The target sequences of *SBP* genes were predicted using psRNATarget [51]. All well-identified mature sequences of miR156 were downloaded from the miRBase database [36]. All miR156 mature sequences were put into psRNATarget as query sequences, and they were then made into searches with default parameters. The sequence logo was created using WebLogo v2.8.2 [50].

#### 2.6. Gene Duplication and Synteny Analyses and Estimations of Synonymous ( $K_s$ ) and Nonsynonymous ( $K_a$ ) Substitutions per Site and Their Ratio

Tandem duplications are defined as two paralogs that are adjacent or separated by one non-homologous gene in the same chromosome, while segmental duplications are homologous and distributed in two blocks of the same or different chromosomes. Where each gene was analyzed as the anchor point, both 30 coding genes upstream and 30 downstream of the anchor points were used to form micro-fragments. Blastp searches were conducted between these micro-fragments and the orthologous fragments were those containing no less than 10 gene pairs from the same gene family with an E-value  $\leq 1 \times 10^{-10}$  [38]. The same method was used for synteny searches and segment duplication searches, although synteny relationships describe those between species.  $K_a$ ,  $K_s$ , and their ratio in all of the duplicated *SBP* gene pairs were estimated using the KaKs\_calculator v2.0 [52,53], following the MA model (which combines multiple computational models and averages them) [54]. It should be noted that only the *SBP* genes located on chromosomes were used for the exploration of duplications and synteny relationships. Considering that some genes show synteny with multiple other genes (one-to-more situation), we took the average of these  $K_s$  values. If a one-to-more situation came from different species, they were not averaged. In addition, according to the synteny relationships, a heatmap cluster was developed. Each synteny was counted once for the two genes making up that synteny. The R package Pheatmap v1.0.12 was used to cluster different groups and genes [55].

#### 2.7. Conservative Analysis

All *SBP* genes from 22 angiosperms were regarded as the reference group, and *SBP* genes from the five species of *C. reinhardtii*, *P. patens*, *S. moellendorffii*, *P. abies*, and *A. trichopoda* were used as control groups. We adopted both OrthoFinder v1.0.6 search and phylogenetic relationships to find the sequences sharing a common ancestor for reference and control species [47]. A phylogeny analysis followed the above-mentioned algorithms. In addition, the HMM profiles of each group were created by HMMER and were used to find the homologous gene of each group [37].

## 2.8. Upstream Regulatory Element Analyses

At the transcription initiation site upstream, −100 to 2000 nt were collected. All of these upstream sequences were uploaded to Binding Site Prediction (PlantTFDB) to predict their possible TF-binding sites [35]. Then, the binding sites identified were transformed into the *A. thaliana* TFs with the homologous binding site. Finally, Gene Ontology (GO) analyses were conducted using the R package clusterProfiler v3.12 [56,57].

## 2.9. Tissue Expression and Stress Response

The normalized expression data for *AtSPL* genes in different tissues and under different treatments were retrieved from AtGenExpress Visualization Tool. To visualize the tissue-expression data, the normalized data were log transformed and presented using Pheatmap v1.0.12 (an R software package) [55]. To show the response-expression profiles of *SBP* genes under different stresses, we counted the number of experiments in which the value of the expression changed significantly (the expression value was double the control group or less than half) under each stress condition.

## 3. Results

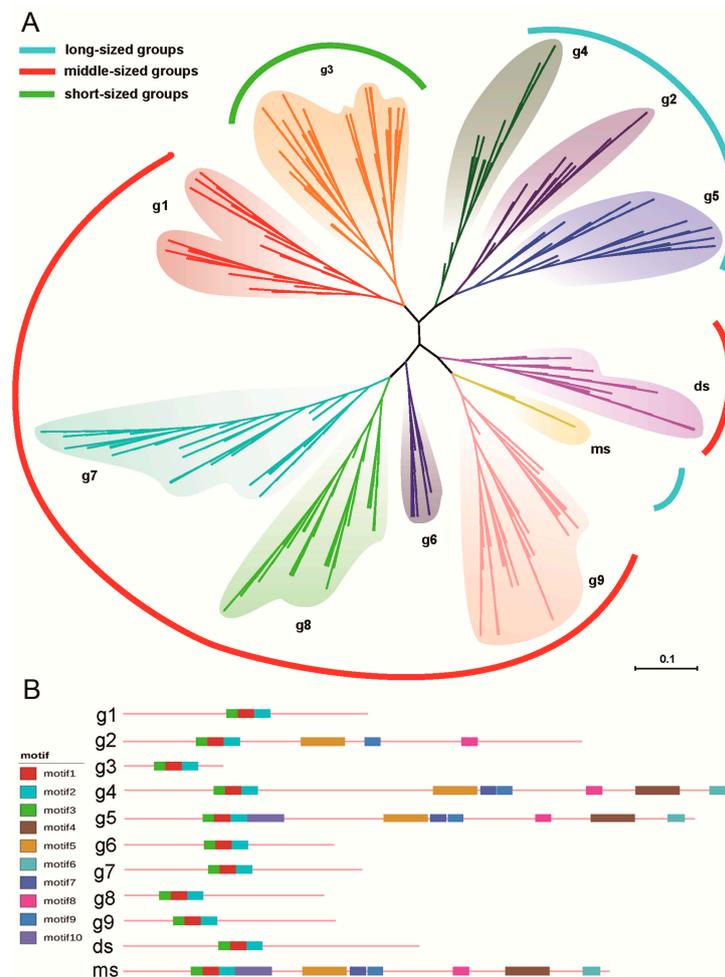
### 3.1. The SBP-Box Gene Family Is Diverse in Angiosperms

To fully understand the SBP-box gene family in flowering plants, we selected nine species from four families of monocots and 13 species from four families of eudicots. The chosen monocots were *B. distachyon*, *A. comosus*, *S. bicolor*, *P. hallii*, *O. sativa*, *M. acuminata*, *S. italica*, *E. guineensis*, and *Z. mays*, and the eudicots were *B. rapa*, *A. thaliana*, *C. sativa*, *P. vulgaris*, *V. angularis*, *C. cajan*, *M. truncatula*, *S. lycopersicum*, *C. annuum*, *N. tabacum*, *P. persica*, *M. domestica*, and *P. mume*. Because *O. sativa* and *A. thaliana* are well-studied model plants, we took their SBP information from a public database. We identified the *SBP* genes of the other species to obtain more reliable SBP information (Table 1). A total of 522 *SBP* genes were obtained from 22 species, of which 310 were eudicots and 212 were monocots (Table 1 and Table S1).

**Table 1.** Source and number of *SBP* genes of each species.

Organism	Source	Number	Nomenclature
<i>Arabidopsis thaliana</i>	PlantTFDB	16	<i>AtSPL</i>
<i>Brassica rapa</i>	This research	29	<i>BrSBP</i>
<i>Camelina sativa</i>	This research	45	<i>CsSBP</i>
<i>Phaseolus vulgaris</i>	This research	22	<i>PvSBP</i>
<i>Vigna angularis</i>	This research	20	<i>VaSBP</i>
<i>Cajanus cajan</i>	This research	23	<i>CcSBP</i>
<i>Medicago truncatula</i>	This research	17	<i>MtSBP</i>
<i>Solanum lycopersicum</i>	This research	16	<i>SISBP</i>
<i>Capsicum annuum</i>	This research	15	<i>CaSBP</i>
<i>Nicotiana tabacum</i>	This research	40	<i>NtSBP</i>
<i>Prunus persica</i>	This research	16	<i>PpSBP</i>
<i>Malus domestica</i>	This research	34	<i>MdSBP</i>
<i>Prunus mume</i>	This research	17	<i>PmSBP</i>
<i>Brachypodium distachyon</i>	This research	17	<i>BdSBP</i>
<i>Sorghum bicolor</i>	This research	19	<i>SbSBP</i>
<i>Panicum hallii</i>	This research	19	<i>PhSBP</i>
<i>Oryza sativa</i>	PlantTFDB	19	<i>OsSPL</i>
<i>Setaria italica</i>	This research	11	<i>SiSBP</i>
<i>Zea mays</i>	This research	32	<i>ZmSBP</i>
<i>Ananas comosus</i>	This research	17	<i>AcSBP</i>
<i>Elaeis guineensis</i>	This research	24	<i>EgSBP</i>
<i>Musa acuminata</i>	This research	54	<i>MaSBP</i>

We constructed phylogenetic trees including all 522 genes. To obtain more accurate results, the trees were constructed using both the maximum likelihood (ML) and Bayesian methods. The criterion for reliable results was that each clustered group contained the same genes, regardless of what method was used. Indeed, we obtained reliable results, with the outcomes of both methods matching; our trees also had relatively high bootstrap values (Figures S1 and S2). An unrooted tree (Figure 1A) was used to show the grouping scheme, according to the results of the ML process. In addition, the OrthoFinder v1.0.6 results suggested that the clustered genes were orthologous. These analyses resulted in 11 groups. The SBPs were concentrated into three length ranges, namely, the short group (g3), the moderate-length groups (g1, g6, g7, g8, g9, and ds), and the long groups (g2, g4, g5, and ms) (Figure S3). Most of the SBP genes were conserved in nine groups (g1, g2, g3, g4, g5, g6, g7, g8, and g9) of angiosperms, containing both monocot and eudicot SBPs. The monocots and eudicots also exhibited differences: one group (ds) was eudicot-unique, and one group (ms) was Gramineae-unique. In general, the SBPs of conserved groups from monocots and eudicots had similar physical and chemical properties (Table 2). Likewise, the SBP protein structure in each group was highly similar, based on conservative motif analyses (Figure S4). Due to the high similarities within each group, we display the most representative structure for each in Figure 1B.



**Figure 1.** Phylogenetic grouping of SBP genes in flowering plants and the conservative motif structure in each group. (A) The phylogenetic tree of SBP genes in flowering plants. Eleven groups were obtained, namely, g1, g2, g3, g4, g5, g6, g7, g8, g9, ds, and ms. All these groups were divided into three length ranges (including the long groups, moderate-length groups, and short groups), and the groups that belonged to the same length range were marked in the same color in the outmost area. (B) The conserved motif of each group. The motif1, motif2, and motif3 together refer to the SBP domain.

Groups g1 to g9 were conserved in angiosperms, but their abundance varied between eudicots and monocots. We counted the average *SBP* gene number per species in each group (Table 2). Some groups had a lower abundance in both monocots and eudicots; for example, the average gene number for g2 and g4 was less than 2. Three moderate-length groups (g7, g8, and g9) were relatively abundant in both monocots and eudicots. Some groups exhibited significant differences between the monocots and eudicots. Group g1 had less than two *SBP* genes per species in eudicots, but nearly four genes per species in monocots. In addition, g3 contained more than five *SBP* genes per species in eudicots, but less than two per species in monocots. Most groups of *SBP* genes were conserved in all or nearly all species (Table 2), but g6 covered only 54% of eudicots and ms was only present in Gramineae.

**Table 2.** The protein physical and chemical properties and gene number of each group.

Group	Species Number	Sequence Number	Average Number	Average Length (AA)	Average Molecular Weight	Average Isoelectric Point
g1-eudicot	13	22	1.69	306.9091	34,068.8	8.913333
g1-monocot	9	35	3.89	405.4857	43,528.37	8.200286
g2-eudicot	12	18	1.5	722.4444	80,763.75	6.218889
g2-monocot	9	13	1.44	859.6923	95,320.8	5.746154
g3-eudicot	13	70	5.38	167.0857	19,179.53	8.482571
g3-monocot	9	15	1.67	203.2	21,441.35	9.813333
g4-eudicot	11	20	1.82	1031.25	114,075.8	8.3985
g4-monocot	8	12	1.5	1029.75	112,982.2	7.616667
g5-eudicot	13	35	2.69	987.7429	109,648.9	6.420571
g5-monocot	8	15	1.875	983	108,311.9	5.984
g6-eudicot	7	12	1.71	352.5	39,079.3	8.4025
g6-monocot	8	10	1.25	436.9	46,449.51	8.838
g7-eudicot	13	36	2.77	422.0833	46,785.88	8.303889
g7-monocot	9	34	3.78	426.6765	46,448.93	8.961471
g8-eudicot	13	29	2.23	363.1034	39,381.31	8.907143
g8-monocot	9	30	3.33	382.1	40,155.08	8.927667
g9-eudicot	13	32	2.46	368.1563	40,526.01	8.3625
g9-monocot	8	42	5.25	396.4286	42,237.75	7.825714
ds	13	36	2.77	496.0278	54,836.9	7.570833
ms	6	7	1.17	879.2857	98,183.74	6.792857

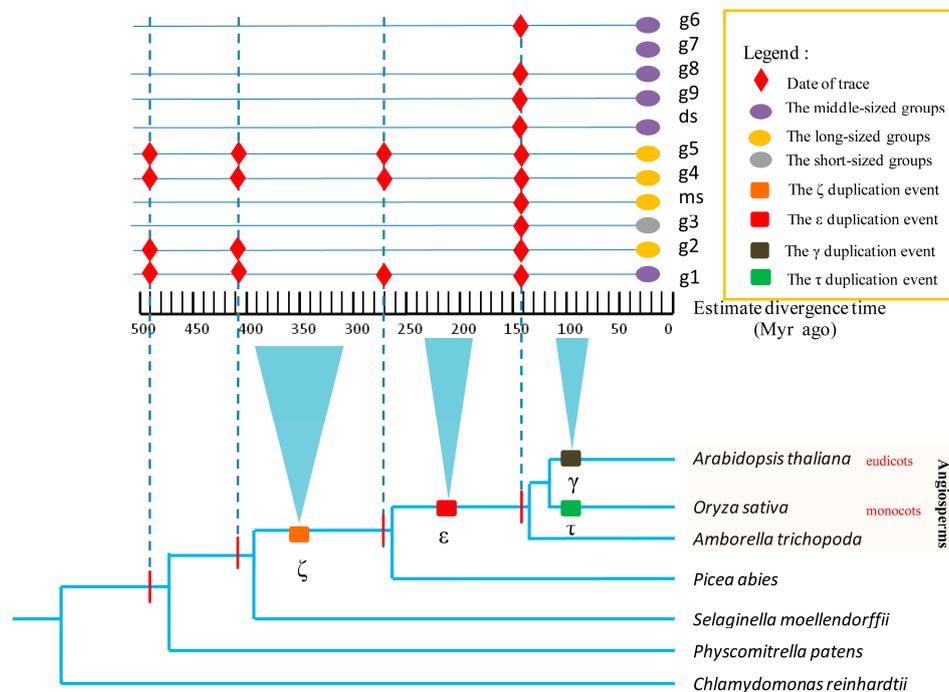
### 3.2. Short and Moderate-Length SBPs Are Typical Targets of miR156

Because some *SBP* genes can be regulated by miR156, it is a key link for exploring the regulatory relationship between miR156 and *SBP*. We used mature sequences of miR156/157 (due to the high similarity between the two, they are both referred to as miR156 here) to search against the *SBP* mRNA sequences to determine *SBP* genes with miR156 binding sites. A total of 346 *SBP* genes containing binding sites for miR156 were detected. Moreover, we found that the binding sites in the *SBP* genes are not random. All long groups (g2, g4, ms, and g5) contained no target sites, while most of the other groups contained binding sites (except for g1). Genome-wide predictions of miR156, including all pre-sequences and mature sequences, were performed (Tables S2 and S3). All details of the miR156s are given in Supplementary Table S4. A corresponding prediction of the secondary structure was also performed (Figure S5). The length of the mature sequence for miR156 and its target site were both generally 20 to 21 nt. The target sequences showed a high consistency at each site, while the miR156 sequences had less consistency at sites 8–13 (Figure S6).

### 3.3. Evolutionary Profiles of *SBP* Genes in Typical Ancient WGD Events

After obtaining the grouping scheme for the angiosperm *SBP*-box gene family, we investigated the phylogenetic relationships of this family in green plants to explore their evolutionary profiles. We carried out phylogenetic analyses of *SBP* genes between angiosperms (we chose *A. thaliana* and *O. sativa* as representative angiosperms) and control plants (Figure 2). All groups were far from aquatic organisms, because no homologous *SBP* gene was found in *C. reinhardtii*. According to the typical

WGDs and the evolution of the main characteristics of angiosperms, three typical stages (the early evolution of angiosperms [ $\epsilon$  to  $\gamma$  and  $\tau$ ], the differentiation of monocots and eudicots [ $\gamma$ ,  $\tau$ , and later events to the K–Pg boundary], and the recent evolution of angiosperms [after K–Pg]) were analyzed. The *SBP* genes of angiosperms showed different evolutionary profiles in these three typical periods.



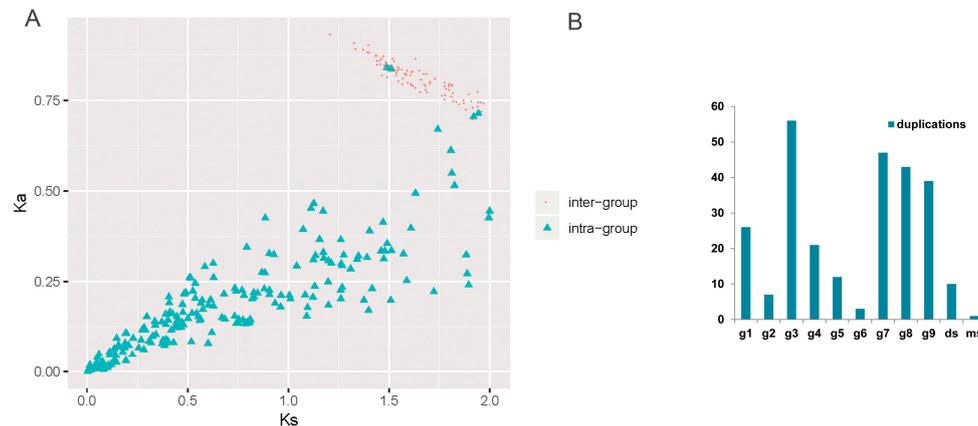
**Figure 2.** The evolutionary profiles of flowering plant *SBP* genes compared with other taxa of plants. The timeline is in the middle, and the approximate time range of each whole genome duplication event is shown in an inverted triangle linked to the timeline. The dotted line refers to the furthest traceable time of each species. The red rhombus on the cross of a dotted line and a solid line represents homologous genes of a group that were found in a species.

In the early evolution of flowering plants, the ancestors of four groups ( $g1$ ,  $g2$ ,  $g4$ , and  $g5$ ) were formed. All of these were homologous to *P. patens*, allowing them to be traced back to about 500 million years ago [58]. Some conserved functions could be assumed in these groups [59,60]. Many angiosperm-specific *SBP* genes emerged during the transformation from gymnosperm plants to flowering plants (after  $\epsilon$ , and before  $\gamma$  and  $\tau$ ). Five groups ( $g3$ ,  $g6$ ,  $g7$ ,  $g8$ , and  $g9$ ) were angiosperm-specific. Like the angiosperms themselves, these groups appeared suddenly in the long history of plant evolution. Moreover, during the differentiation of monocots and eudicots,  $ds$  appeared as an independent eudicot group, while  $ms$  became a specific Gramineae group, distinguished from other monocots. Although the *SBP* phylogenetic composition of monocots and eudicots was not significantly differentiated during this period, some eudicot and monocot groups showed obvious differences in their copy numbers (Table 2). However, monocot and eudicot groups also displayed similar retention tendencies. For example,  $g7$ ,  $g8$ , and  $g9$  exhibited high copy retention in both eudicots and monocots (Table 2). In angiosperms, some WGDs occurred in individual families or species. Many copies were formed during these WGDs, resulting in a large degree of gene redundancy.

### 3.4. Expansion Profile Describing the Dynamic Evolutionary Histories

The inheritance and innovation of plant gene families is a complex and dynamic process. A deeper understanding of the dynamic process of evolution will help reveal the underlying mechanism of that evolution in plants. To this end, we explored the expansion of *SBP* by duplication. In total, 339 *SBP* genes were found in duplicate relationships. These duplicate genes consisted of many *SBP* genes (71%)

(Figure S7A). Segment duplications played the most significant role in SBP expansion (Figure S7B and Table S5). Interestingly, both intra- and inter-group duplications were found (Table S5), which implies that some groups may have had genetic and evolutionary relationships. Selection pressure analyses showed that the intra- and inter-group duplication pairs were all subject to purification selection (Figure 3A). The  $K_s$  values of all inter-group duplications were greater than 1.2 (mostly earlier than  $\gamma$  and  $\tau$ ). In contrast to the inter-group duplications, 81.2% of intra-group duplications had  $K_s$  values of less than 1.2 (mostly later than  $\gamma$  and  $\tau$ ). Of all intra-group duplications, some miR-regulated groups had the largest degree of duplicated retention (Figure 3B).

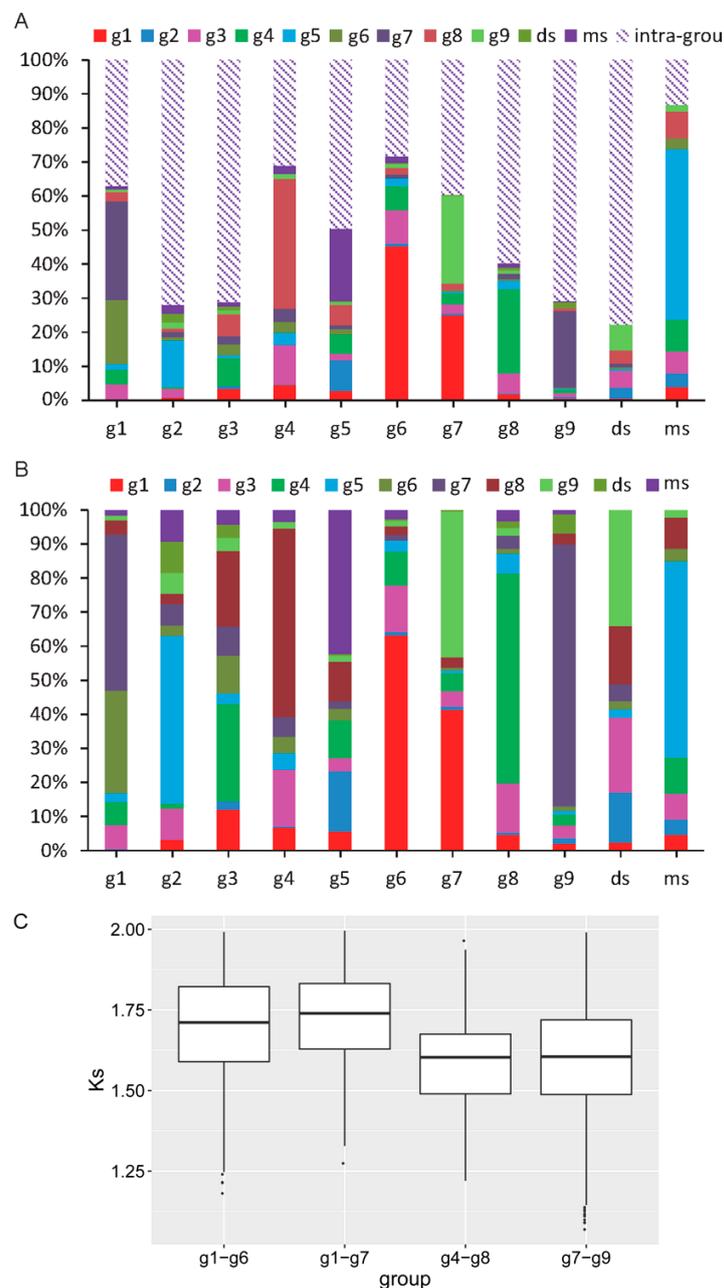


**Figure 3.** The selection pressure analysis of intra- and inter-duplicated *SBP* genes and the number of intra-group duplications in each group. (A) The X axis refers to the  $K_s$  value, and the Y axis refers to the  $K_a$  value. The inter-group duplications were marked with red dots, and the intra-group duplications were marked with green triangles. (B) A bar chart of the intra-group duplications in each group.

These analyses provide genetic correlations between groups, but a higher-resolution correlation must be obtained to reveal dynamic evolutionary histories. Synteny mining among groups revealed clues that could be used to explore such correlations. All *SBP* genes which have been identified as being located on chromosomes were used for synteny analyses. As with the duplication results, there were both intra-group and inter-group synteny relationships. A total of 442 *SBP* genes had synteny relationships, accounting for 98.2% of all *SBP* genes (Figure S8A). There were significantly more intra-group synteny relationships (2999) than inter-group ones (1342) (Table S6.1,6.2). Any given group may have synteny relationships with multiple other groups, but the most relationships often appear among any given group's own members (Figure 4A); this was the case for nine groups. Interestingly, the inter-group relationships were not arbitrary. These are presented as a stacked bar chart in Figure 4B, organized into eight groups in two clusters. Four with-target groups were directly or indirectly associated with g1. Group g1 had 30% synteny with g6 and 45% synteny with g7. In addition, g9 was closely related to g7, having 76% synteny with it. In addition, ds had more than 34% synteny with g7. In addition, two other with-target groups (g8 and g3) had strong synteny relationships with g4. Because the synteny relationships involved in each group are not all one-to-one correlations, that is, any given gene may have synteny relationships with multiple other genes, we conducted cluster analyses for all synteny relationships. The results support our previous analyses to some extent (Figure S9).

The  $K_s$  values for g4–g8, g1–g6, g1–g7, and g7–g9 were calculated, for they all produced more than 100 pairs of synteny relationships and provide greater statistical significance (Figure 4C). All of these  $K_s$  values mainly fell between 1.2 and 2, consistent with the duplication results (Figure 3A). For the four groups (g1, g6, g7, and g9) identified in the same cluster, the  $K_s$  values for the with-target groups and g1 (g1–g6 and g1–g7) were larger than those for the two with-target groups (g7–g9), suggesting that duplications between the angiosperm-specific group and the plant-conservative group mainly occurred before the expansion of angiosperm-specific groups. After the differentiation of

the with-target group from the without-target group, the with-target group differentiated into other angiosperm-specific groups.

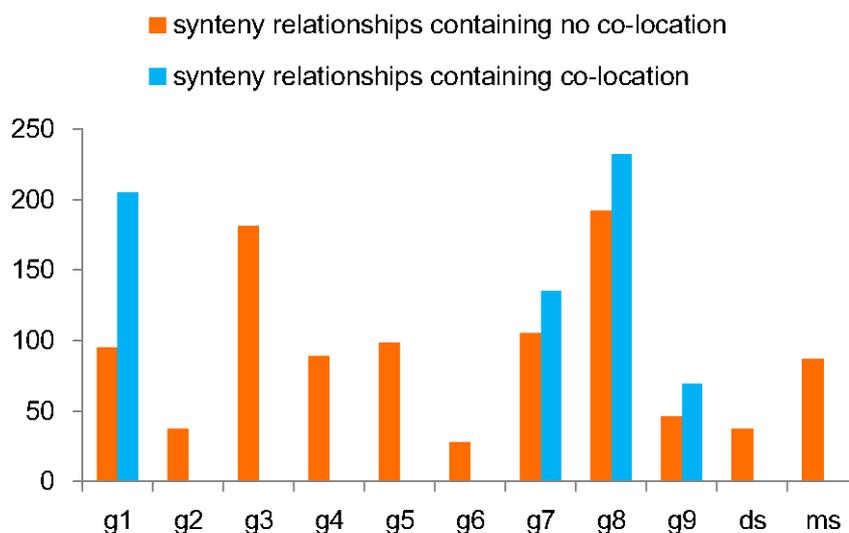


**Figure 4.** The synteny relations between groups and the *Ks* value of three pairs of groups. (A) A percentage stacked bar chart of all inter-group and intra- group synteny relationships. (B) A percentage stacked bar chart of all inter-group synteny relationships. (C) The *Ks* value of four pairs of groups with a relatively high number of synteny relationships.

### 3.5. Coevolutionary Profile of miR156 and SBP Showing the Origin of miR156

There may be evolutionary correlations between miR156 and SBP. We investigated this possibility and found that there were many duplications (Figure S7B and Table S7) and synteny relationships among miR156s (Figure S8B and Table S8). This makes us suspect that there are synteny relationships between miR156 and SBP. If so, it may be that their coevolution could be derived through these relationships. First, because of the possibility of the co-location of miR156 and SBP, we examined

coincidence of genes in the micro-fragments formed by SBP and miR156. Co-location was defined as the co-occurrence of miR156 and SBP in a micro-fragment. After excluding all co-locations, we identified 161 miR156 genes (accounting for 73% of miR156 genes) with synteny relationships with 214 SBP genes (accounting for 48% of SBP genes) (Figure S8C,D and Table S9). Synteny relationships between miR156 and SBP were found in all groups (Figure 5). Some groups displayed a significant correlation with miR156 due to their relatively large number of synteny relationships. For seven groups, more than 50 synteny pairs were found (Figure 5). Co-location with miR156 was found for four groups (g1, g7, g8, and g9) (Table S10). These results further confirm their evolutionary correlations and produce an adjusted synteny number for these groups (Figure 5 and Table S9).



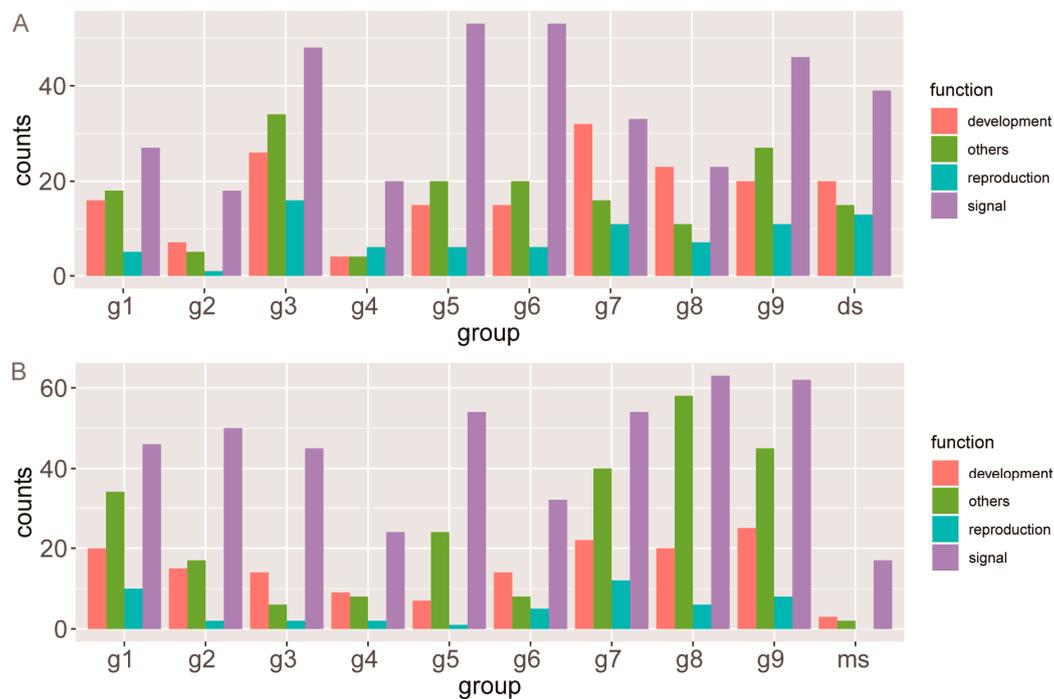
**Figure 5.** The number of synteny relationships between the SBP group and miR156. The orange bar refers to the number of synteny relationships without co-locations, while the blue bar refers to the number of synteny relationships with co-locations.

Next, we considered the formation of stem-loop structures. We found that the location of the genes of g1 and g6 provides natural conditions for the formation of miR156, such as AcSBP3 (in g6) and AcSBP4 (in g1), which are separated by less than 8000 nucleotides, and their 3' ends are in the face-to-face form (Table S11). It is very interesting that up to 13 of the 21 g6 SBP genes had a gene from g1 located close to them (Table S11). In a previous study, we revealed that the distance between g1 and g6 in Euphorbiaceae is consistent with tandem duplication [16]. The data suggest that the same is true in this study for g1 and g6. In addition, as many as 10 of the 13 pairs of SBP genes are in the reverse direction. Furthermore, nine of the 10 reverse SBP gene pairs are in face-to-face relationships, which provide a potential structure for the formation of an miRNA stem-loop frame.

### 3.6. Functional Evolutionary Profile Indicating the Functional Differentiation of Each Group

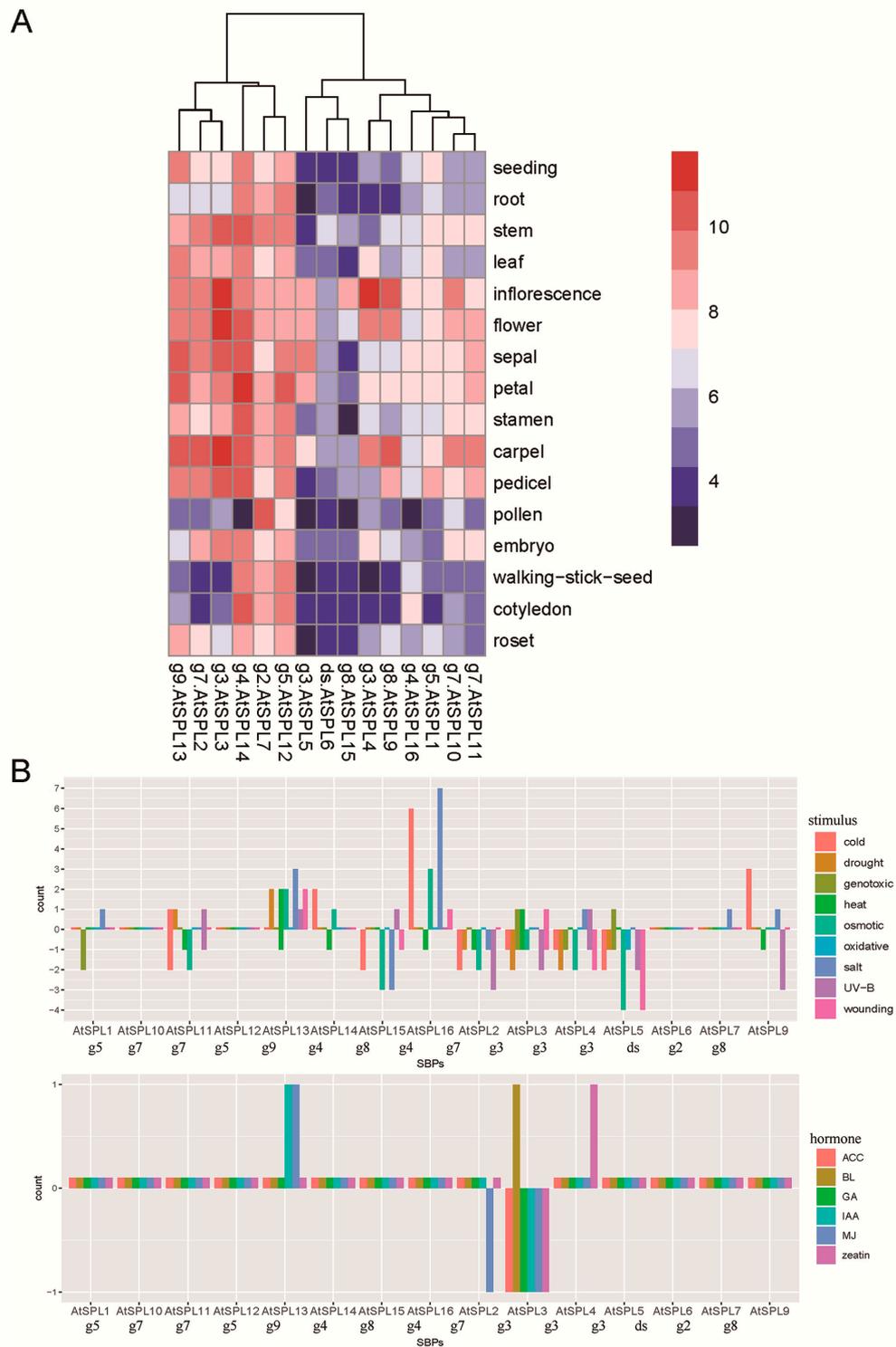
Gene transcription initiation is closely related to the upstream regulatory region. To predict the functions of each group, we predicted the possible binding TFs for each and performed functional enrichment analyses of the predicted TFs. To improve the reliability, we deleted annotations that shared fewer than three species in a group. Following the annotation for each Gene Ontology (GO) item, we divided all enriched biological processes into four categories (development, reproduction, signal, and other) (Figure 6 and Tables S12 and S13). Regarding dicots, almost all miR-regulated SBP genes were associated with relatively high numbers of GO-enrichment items in development and reproduction (Figure 6A). Similarly, most groups of miR-regulated SBP genes from monocots were also enriched in these categories (Figure 6B). However, the g3 monocot group (g3-monocot) was significantly different from the g3 dicot group (g3-dicot), which was enriched in many fewer items

related to development and reproduction (Figure 6). In addition, the long groups, which are not miR-regulated groups, had relatively fewer enrichment items related to development (Figure 6).



**Figure 6.** Gene Ontology (GO) enrichment of the predicted binding transcription factors (TFs) upstream in each group. (A) GO enrichment of dicots. (B) GO enrichment of monocots.

To validate these results, we analyzed *AtSPL* gene expression in various tissues (Figure 7A) and under various conditions (Figure 7B). Some miR-regulated *SBP* genes indeed showed a high expression during certain growth and reproduction processes (Figure 7A). *AtSPL3/2/13* (belonging to g3, g7, and g9) were highly expressed in multiple tissues, and they exhibited strong correlations with growth and reproduction. *AtSPL12/7/14* were highly expressed in nearly all tissues and displayed a housekeeping expression pattern.



**Figure 7.** The expression profile of AtSPLs. (A) The tissue expression of AtSPLs. (B) The number of differential expressions of AtSPL genes in each stress treatment.

## 4. Discussion

### 4.1. A Consistent Classification Standard Is Helpful for Exploring Conservative Gene Families in Plants

With reliable classification standards as a reference, it is simple to determine similarities and differences in the analyses of gene families. As genome-sequencing data for plants increase and

bioinformatics tools are developed, the in-silico identification of gene families will grow. However, previous studies have often been limited to only a few species, lacking representativeness of the broad range of angiosperms. Some species may lack individual groups of a gene family. In a phylogenetic study of *A. thaliana* and *O. sativa*, all *SBP* genes were divided into nine groups, but *A. thaliana* was absent in one [14]. Different studies may give different classifications. In analyses of *A. thaliana*, *O. sativa*, and *P. patens*, all of the AtSPLs were divided into six groups [13]. However, in another study of *A. thaliana* and *O. sativa*, all AtSPLs were divided into eight groups [14]. The differences in the species and evaluation criteria used may have led to great differences in the classification results. We obtained a reliable angiosperm *SBP* classification through large-scale phylogenetic analyses and described the evolutionary characteristics for each group, providing a reference for further research.

#### 4.2. Promotion of the Evolution of Angiosperms by the Differential Expansion of the *SBP*-Box Gene Family during Typical Ancient WGDs

A certain conservative gene family may have a similar evolutionary profile to a given plant. The origin of angiosperms can be traced back to at least the Cretaceous [1–4], and the  $\epsilon$  event can be traced to about 200 million years ago [22]. During the Cretaceous, the Earth was dry and hot, with high oxygen levels. Some typical characteristics of angiosperms were formed during this period, such as the evolution of the carpel, the emergence of double fertilization, and the origin of the flower [61]. Likewise, the main *SBP* branches of angiosperms were formed during this period (the grouping scheme including g1, g2, g3, g4, g5, g6, g7, g8, and g9), and the angiosperm-specific *SBP* genes (g3, g6, g7, g8, and g9) all appeared in this period. The *SBP* gene promotes the adaptation of angiosperms to the environment and the evolution of angiosperms' specificity. It is involved in the regulation of floral organ formation, transformation of the plant from the vegetative phase to the reproductive phase, the formation of plant morphology, and signal response and stress resistance [17,18,62].

WGDs provide impetus for differentiation and innovation in plants. During the process of the differentiation of monocots and eudicots, multiple WGDs occurred. Monocots and eudicots showed different expansion profiles, and the expansions of different types of *SBP* genes were also different. The number of *SBP* copies retained by monocots after  $\tau$  was much greater than that after  $\gamma$  (which occurred at a similar time as  $\tau$ ) in eudicots [32]. MiR-regulated groups are involved in multiple aspects of plant reproduction, growth, and the stress response [17,18,63–66] and tend to retain more copies of *SBP* genes. After the  $\gamma$  and  $\tau$  events, many TF gene families related to development, morphogenesis, and the response to abiotic and biotic stresses retained large numbers of copies [32,67].

In angiosperms, the innovation of the *SBP*-box gene family was not shared by a large range of plants, but it increased the functional complexity through the high retention of copies. The genes of some multiple-copy groups perform similar functions redundantly [17,18]. Some genes from different groups may have similar functions [68], and even copies of a group may be functionally differentiated [69].

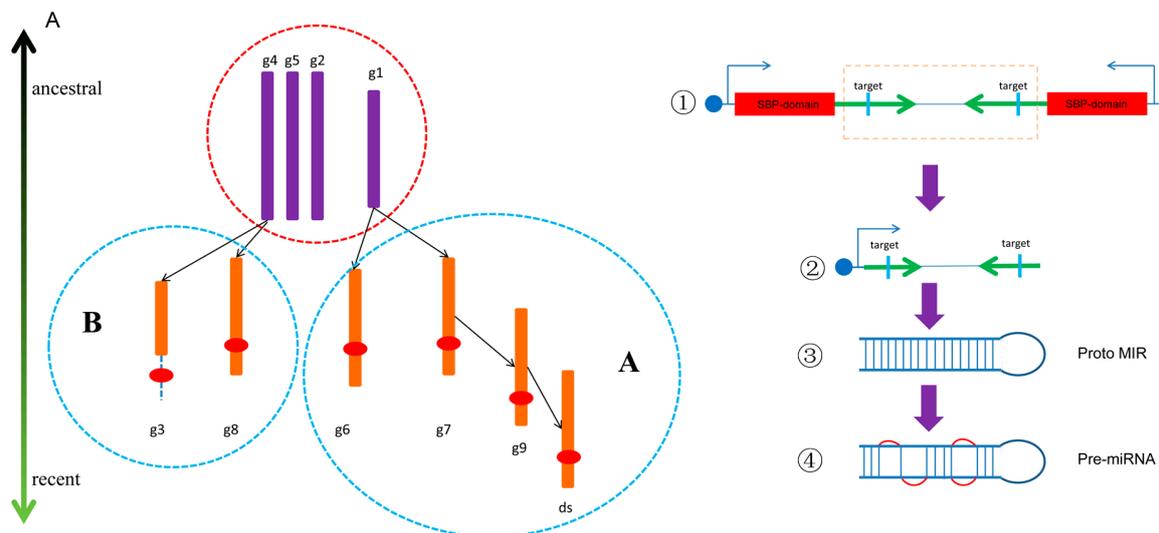
#### 4.3. Short and Moderate-Length *SBPs* as Important Regulators of Growth and Reproduction in Angiosperms

There are many conservative gene families in plants, but those with many functions and differentiations are more attractive to study. They react to changes in the environment and respond quickly. Some *SBP* genes, regulated by miR156, can be quite variable. Although they appeared much earlier than angiosperms [70], they are also continuing to evolve as angiosperms do. It appears that regulation by miR156 determines the fate of *SBP* genes. Groups without a target site are all conservative groups of terrestrial plants. Those with a target are all angiosperm-specific groups. The miR156-*SBP* module allowed *SBP* genes to adapt to more complex and variable regulatory processes and participate in more complex regulatory networks [65]. G1 appears to connect groups with and without target groups. The *SBP* genes in g1 were significantly different in length from other terrestrial-conservative *SBPs*; although their length is similar to that of angiosperm-specific groups, they do not have target sites.

Moreover, instead of the functional housekeeping of other terrestrial-conservative groups, they are involved in angiosperm-specific physiological processes [68,71–73].

#### 4.4. Dynamic Evolutionary Process Analyses Reveal the Expansion Process of Different Groups

The emergence of a new branch of a conservative gene family in plants may be accompanied by plant innovation [21]. Gene families related to important physiological processes have been used to determine plant classifications [74]. However, it is difficult to understand the evolutionary genetic relationships between branches. The duplication of one branch by another is likely to leave synteny clues. The lack of exploration of these relationships has led to an overall poor description of inter-branch evolution in previous gene family analyses in multiple species. Using the inter-group synteny relationships, we made an inferential map (Figure 8A) of the dynamic evolution of the SBP-box gene family. Different groups of SBP genes do not originate independently, and with-target groups may form without-target groups. Although moderate-length groups have similar sequence lengths, they may originate from two lines: line A and B. Line A contains five groups, four of which are with-target groups and one of which is a no-target group. G6 and g7 may come directly from g1, and g9/ds may have been produced successively after g7. Line B contains three groups, and its two with-target groups (g3 and g8) may originate from the without-target group g4.



**Figure 8.** A derivation graph of the dynamic evolutionary process of SBP genes in flowering plants and the presumed model of the emergence of miR156. (A) A derivation graph of the dynamic evolutionary process of SBP genes in flowering plants. The terrestrial plant conservative groups are circled with a red dotted line. Angiosperm-specific groups associated with a terrestrial group relate to solid arrows, and the direction of the arrow represents the direction of evolution. Line A and Line B are circled by a blue dotted line, respectively. The red ovals refer to miR156 target sequences and the 3'UTR region is marked with a dotted line. (B) The presumed model of the emergence of miR156. ① The emergence of the face-to-face tandem duplication of two SBP genes. ② The formation of genes containing the ancestor sequence of MIR156 and MIR156\*. ③ The formation of proto MIR156. ④ The formation of pre-miR156.

#### 4.5. MiR156 May Originate From Face-to-Face Tandem Duplication of SBP Genes

The regulation module between a conservative gene family and its regulator miRNA is often traced to the origin of terrestrial plants. Very long periods of independent evolution make it difficult to clearly explain the relationships between an miRNA and its target gene. However, three prevailing models account for the possible emergence of miRNAs, namely, the inverted duplication of target genes [75], intermittent evolution [76], and inverted-repeat transferable elements [77,78]. Our research shows that there is another possibility for the generation of miR156. Based on our results, the formation

of miR156 may be related to tandem duplications between g1 and g6, with a face-to-face relationship leading to the formation of a stem-loop structure (Figure 8B). Then, this stem-loop structure may have evolved further to obtain the precursor structure of miR156. In protein structure analyses, we found that moderate-length and short SBPs seem to be formed through the truncation of long SBPs. The ancestors of g1 may also have the characteristics of long sequences, but the formation of miR156 led to sequence loss in g1. There are still many unknowns about the formation of miRNAs, and much work remains to be done to reveal their generation and evolution.

## 5. Conclusions

Our results provide a comprehensive phylogenetic classification of the angiosperm SBP-box gene family. Moreover, we studied the evolution of angiosperm SBP genes, accompanied by several typical ancient WGDs of flowering plants, and presented the different evolutionary profiles of these genes corresponding to the WGDs. In addition, synteny relationships between different SBP groups and between SBP and miR156 were found, and the correlation between SBP genes and miR156 was elucidated. The dynamic genetic and evolutionary development of different SBP groups suggests two obvious genetic lines, and a possible path of emergence was proposed for miR156. Furthermore, an evolutionary map of SBP functional evolution was presented in a preliminary exploration of the angiosperm SBP-box gene family.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2218-273X/10/5/757/s1>: Figure S1: The phylogenetic tree reconstructed by the maximum likelihood (ML) method; Figure S2: The phylogenetic tree reconstructed by the Bayesian method; Figure S3: The three length ranges of SBP proteins; Figure S4: The conservative motifs for each SBP protein. Figure S5: The RNA folding structure of each predicted miR156; Figure S6: The conservatism of the miR156 mature sequence and target sequence; Figure S7: The number and type of duplicated genes; Figure S8: The number of genes in synteny relationships; Figure S9: The heatmap clustering of synteny SBP genes and groups; Table S1: The basic information of SBP genes from 22 species; Table S2: The mature miR156 sequence of 22 species from 5' to 3'; Table S3: The pre-sequences of miR156s; Table S4: The basic information of miR156s; Table S5: Duplications of intra-groups and inter-groups; Table S6: The synteny relationships of intra-groups and inter-groups; Table S7: Duplications among miR156s; Table S8: The synteny pairs of miR156; Table S9: The synteny pairs between SBP and miR156. The red ones show the synteny pairs with co-locations; Table S10: The co-location relationships between SBP and miR156; Table S11: The co-location of genes from g1 and g6, and their directions; Table S12: GO enrichment of eudicots; Table S13: GO enrichment of monocots.

**Author Contributions:** C.L. conceived and supervised this study. C.L. and J.L. designed the experiments. J.L. carried out the experiments and analyzed and interpreted the data. X.G. and X.Z. participated in the discussion and provided valuable advice and practical contributions. C.L., J.L., and X.G. wrote the manuscript. All authors reviewed, edited, and approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (No. 31970609), Start-up Fund from Xishuangbanna Tropical Botanical Garden, 'Top Talents Program in Science and Technology' from Yunnan Province.

**Acknowledgments:** We thank all colleagues in our laboratory for helpful discussions and technical assistance.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Crane, P.R.; Friis, E.M.; Pedersen, K.R. The origin and early diversification of angiosperms. *Nature* **1995**, *374*, 27–33. [[CrossRef](#)]
2. Magallón, S.; Castillo, A. Angiosperm diversification through time. *Am. J. Bot.* **2009**, *96*, 349–365. [[CrossRef](#)]
3. Soltis, P.S.; Soltis, D.E. The origin and diversification of angiosperms. *Am. J. Bot.* **2004**, *91*, 1614–1626. [[CrossRef](#)]
4. The Angiosperm Phylogeny Group; Chase, M.W.; Christenhusz, M.J.M.; Fay, M.F.; Byng, J.W.; Judd, W.S.; Soltis, D.E.; Mabberley, D.J.; Sennikov, A.N.; Soltis, P.S.; et al. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **2016**, *181*, 1–20.

5. Mizutani, M.; Ohta, D. Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant Biol.* **2010**, *61*, 291–315. [[CrossRef](#)]
6. Morant, M.; Bak, S.; Møller, B.L.; Werck-Reichhart, D. Plant cytochromes p450: Tools for pharmacology, plant protection and phytoremediation. *Curr. Opin. Biotech.* **2003**, *14*, 151–162. [[CrossRef](#)]
7. Oh, S.-K.; Baek, K.-H.; Park, J.M.; Yi, S.Y.; Yu, S.H.; Kamoun, S.; Choi, D. *Capsicum annuum* WRKY protein CaWRKY1 is a negative regulator of pathogen defense. *New Phytol.* **2008**, *177*, 977–989. [[CrossRef](#)]
8. Eulgem, T.; Rushton, P.J.; Robatzek, S.; Somssich, I.E. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **2000**, *5*, 199–206. [[CrossRef](#)]
9. Rinerson, C.I.; Scully, E.D.; Palmer, N.A.; Donze-Reiner, T.; Rabara, R.C.; Tripathi, P.; Shen, Q.J.; Sattler, S.E.; Rohila, J.S.; Sarath, G.; et al. The WRKY transcription factor family and senescence in switchgrass. *BMC Genom.* **2015**, *16*, 912. [[CrossRef](#)]
10. Becker, A.; Theißen, G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* **2004**, *29*, 464–489. [[CrossRef](#)]
11. Klein, J.; Saedler, H.; Huijser, P. A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity genes *QUAMOSA*. *Mol. Gen. Genet.* **1996**, *250*, 7–16.
12. Guo, A.-Y.; Zhu, Q.-H.; Gu, X.; Ge, S.; Yang, J.; Luo, J. Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family. *Gene* **2008**, *418*, 1–8. [[CrossRef](#)]
13. Riese, M.; Höhmann, S.; Saedler, H.; Münster, T.; Huijser, P. Comparative analysis of the SBP-box gene families in *P. patens* and seed plants. *Gene* **2007**, *401*, 28–37. [[CrossRef](#)]
14. Yang, Z.; Wang, X.; Gu, S.; Hu, Z.; Xu, H.; Xu, C. Comparative study of SBP-box gene family in *Arabidopsis* and rice. *Gene* **2008**, *407*, 1–11. [[CrossRef](#)]
15. Cardon, G.; Höhmann, S.; Klein, J.; Nettesheim, K.; Saedler, H.; Huijser, P. Molecular characterisation of the *Arabidopsis* SBP-box genes. *Gene* **1999**, *237*, 91–104. [[CrossRef](#)]
16. Li, J.; Gao, X.; Sang, S.; Liu, C. Genome-wide identification, phylogeny, and expression analysis of the SBP-box gene family in Euphorbiaceae. *BMC Genom.* **2019**, *20*, 912. [[CrossRef](#)]
17. Jung, J.-H.; Lee, H.-J.; Ryu, J.Y.; Park, C.-M. SPL3/4/5 integrate developmental aging and photoperiodic signals into the FT-FD module in *Arabidopsis* flowering. *Mol. Plant* **2016**, *9*, 1647–1659. [[CrossRef](#)]
18. Shikata, M.; Koyama, T.; Mitsuda, N.; Ohme-Takagi, M. *Arabidopsis* SBP-box genes *SPL10*, *SPL11* and *SPL2* control morphological change in association with shoot maturation in the reproductive phase. *Plant Cell Physiol.* **2009**, *50*, 2133–2145. [[CrossRef](#)]
19. De Bodt, S.; Maere, S.; Van de Peer, Y. Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* **2005**, *20*, 591–597. [[CrossRef](#)]
20. Guo, H.; Lee, T.-H.; Wang, X.; Paterson, A.H. Function relaxation followed by diversifying selection after whole-genome duplication in flowering plants. *Plant Physiol.* **2013**, *162*, 769. [[CrossRef](#)]
21. Soltis, P.S.; Soltis, D.E. Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* **2016**, *30*, 159–165. [[CrossRef](#)]
22. Jiao, Y.; Wickett, N.J.; Ayyampalayam, S.; Chanderbali, A.S.; Landherr, L.; Ralph, P.E.; Tomsho, L.P.; Hu, Y.; Liang, H.; Soltis, P.S.; et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* **2011**, *473*, 97–100. [[CrossRef](#)]
23. Jaillon, O.; Aury, J.-M.; Noel, B.; Policriti, A.; Clepet, C.; Casagrande, A.; Choisne, N.; Aubourg, S.; Vitulo, N.; Jubin, C.; et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **2007**, *449*, 463–467.
24. Jiao, Y.; Leebens-Mack, J.; Ayyampalayam, S.; Bowers, J.E.; McKain, M.R.; McNeal, J.; Rolf, M.; Ruzicka, D.R.; Wafula, E.; Wickett, N.J.; et al. A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* **2012**, *13*, R3. [[CrossRef](#)]
25. Tang, H.; Bowers, J.E.; Wang, X.; Paterson, A.H. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 472. [[CrossRef](#)]
26. Vekemans, D.; Proost, S.; Vanneste, K.; Coenen, H.; Viaene, T.; Ruelens, P.; Maere, S.; Van de Peer, Y.; Geuten, K. Gamma paleohexaploidy in the stem lineage of core eudicots: Significance for MADS-box gene and species diversification. *Mol. Biol. Evol.* **2012**, *29*, 3793–3806. [[CrossRef](#)]

27. Jiang, W.-K.; Liu, Y.-L.; Xia, E.-H.; Gao, L.-Z. Prevalent role of gene features in determining evolutionary fates of whole-genome duplication duplicated genes in flowering plants. *Plant Physiol.* **2013**, *161*, 1844–1861. [[CrossRef](#)]
28. Kassahn, K.S.; Dang, V.T.; Wilkins, S.J.; Perkins, A.C.; Ragan, M.A. Evolution of gene function and regulatory control after whole-genome duplication: Comparative analyses in vertebrates. *Genome Res.* **2009**, *19*, 1404–1418. [[CrossRef](#)]
29. Rensing, S.A. Gene duplication as a driver of plant morphogenetic evolution. *Curr. Opin. Plant Biol.* **2014**, *17*, 43–48. [[CrossRef](#)]
30. Wang, Y.; Tan, X.; Paterson, A.H. Different patterns of gene structure divergence following gene duplication in *Arabidopsis*. *BMC Genom.* **2013**, *14*, 652. [[CrossRef](#)]
31. Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. *Science* **2013**, *342*, 1241089. [[CrossRef](#)]
32. Wu, S.; Han, B.; Jiao, Y. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant* **2019**, *13*, 59–71. [[CrossRef](#)]
33. Ncbi Resource Coordinators. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2015**, *43*, D6–D17. [[CrossRef](#)]
34. Lamesch, P.; Berardini, T.Z.; Li, D.; Swarbreck, D.; Wilks, C.; Sasidharan, R.; Muller, R.; Dreher, K.; Alexander, D.L.; Garcia-Hernandez, M.; et al. The *Arabidopsis* information resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **2011**, *40*, D1202–D1210. [[CrossRef](#)]
35. Jin, J.; Tian, F.; Yang, D.-C.; Meng, Y.-Q.; Kong, L.; Luo, J.; Gao, G. PlantTFdb 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **2016**, *45*, D1040–D1045. [[CrossRef](#)]
36. Griffiths-Jones, S.; Grocock, R.J.; van Dongen, S.; Bateman, A.; Enright, A.J. miRbase: MicroRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **2006**, *34*, D140–D144. [[CrossRef](#)]
37. Ghahramani, Z. An introduction to hidden markov models and bayesian networks. *Hidden Markov Models World Sci.* **2001**, *45*, 9–41.
38. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. Blast+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
39. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2018**, *47*, D427–D432. [[CrossRef](#)]
40. Letunic, I.; Bork, P. 20 years of the smart protein domain annotation resource. *Nucleic Acids Res.* **2017**, *46*, D493–D496. [[CrossRef](#)]
41. Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I.; et al. CDD: NCBI’s conserved domain database. *Nucleic Acids Res.* **2014**, *43*, D222–D226. [[CrossRef](#)] [[PubMed](#)]
42. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [[CrossRef](#)] [[PubMed](#)]
43. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. Meme suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [[CrossRef](#)] [[PubMed](#)]
44. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
45. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MRBAYES 3.2: Efficient Bayesian phylogenetic inference and model selection across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)]
46. Swofford, D.L. *Paup\*: Phylogenetic analysis using parsimony (and other methods)*, version 4.0.B10; Sinauer Associates: Sunderland, MA, USA, 2002.
47. Emms, D.M.; Kelly, S. Orthofinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157. [[CrossRef](#)]
48. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415. [[CrossRef](#)]

49. Patanun, O.; Lertpanyasampatha, M.; Sojikul, P.; Viboonjun, U.; Narangajavana, J. Computational identification of microRNAs and their targets in cassava (*Manihot esculenta crantz.*). *Mol. Biotechnol.* **2013**, *53*, 257–269. [[CrossRef](#)]
50. Crooks, G.E.; Hon, G.; Chandonia, J.-M.; Brenner, S.E. Weblogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)]
51. Dai, X.; Zhuang, Z.; Zhao, P.X. psRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **2018**, *46*, W49–W54. [[CrossRef](#)]
52. Hurst, L.D. The ka/ks ratio: Diagnosing the form of sequence evolution. *Trends Genet.* **2002**, *18*, 486–487. [[CrossRef](#)]
53. Wang, D.; Zhang, Y.; Zhang, Z.; Zhu, J.; Yu, J. Kaks\_calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinf.* **2010**, *8*, 77–80. [[CrossRef](#)]
54. Zhang, Z.; Li, J.; Zhao, X.-Q.; Wang, J.; Wong, G.K.-S.; Yu, J. Kaks\_calculator: Calculating *ka* and *ks* through model selection and model averaging. *Genom. Proteom. Bioinf.* **2006**, *4*, 259–263. [[CrossRef](#)]
55. Kolde, R. *Pheatmap: Pretty Heatmaps*, version 1.0.12; Raivo Kolde, 2019.
56. Yu, G. ClusterProfiler: Universal enrichment tool for functional and comparative study. *BioRxiv* **2018**, 256784. [[CrossRef](#)]
57. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **2012**, *16*, 284–287. [[CrossRef](#)]
58. Schaefer, D.G.; Zrýd, J.-P. The moss *Physcomitrella patens*, now and then. *Plant Physiol.* **2001**, *127*, 1430–1438. [[CrossRef](#)]
59. Castruita, M.; Casero, D.; Karpowicz, S.J.; Kropat, J.; Vieler, A.; Hsieh, S.I.; Yan, W.; Cokus, S.; Loo, J.A.; Benning, C.; et al. Systems biology approach in *Chlamydomonas*; reveals connections between copper nutrition and multiple metabolic steps. *Plant Cell* **2011**, *23*, 1273. [[CrossRef](#)]
60. Kropat, J.; Tottey, S.; Birkenbihl, R.P.; Depège, N.; Huijser, P.; Merchant, S. A regulator of nutritional copper signaling in *Chlamydomonas*; is an SBP domain protein that recognizes the GTAC core of copper response element. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18730. [[CrossRef](#)]
61. Stuessy, T.F. A transitional-combinational theory for the origin of angiosperms. *Taxon* **2004**, *53*, 3–16. [[CrossRef](#)]
62. Chao, L.-M.; Liu, Y.-Q.; Chen, D.-Y.; Xue, X.-Y.; Mao, Y.-B.; Chen, X.-Y. *Arabidopsis* transcription factors SPL1 and SPL12 confer plant thermotolerance at reproductive stage. *Mol. Plant* **2017**, *10*, 735–748. [[CrossRef](#)]
63. Wang, J.-W.; Czech, B.; Weigel, D. MiR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* **2009**, *138*, 738–749. [[CrossRef](#)]
64. Schwarz, S.; Grande, A.V.; Bujdoso, N.; Saedler, H.; Huijser, P. The microRNA regulated SBP-box genes *SPL9* and *SPL15* control shoot maturation in *Arabidopsis*. *Plant Mol. Biol.* **2008**, *67*, 183–195. [[CrossRef](#)]
65. Wang, H.; Wang, H. The miR156/SPL module, a regulatory hub and versatile toolbox, gears up crops for enhanced agronomic traits. *Mol. Plant* **2015**, *8*, 677–688. [[CrossRef](#)]
66. Wang, J.-W.; Schwab, R.; Czech, B.; Mica, E.; Weigel, D. Dual Effects of miR156-Targeted *SPL* Genes and CYP78A5/KLUH on Plastochron Length and Organ Size in *Arabidopsis thaliana*. *Plant Cell* **2008**, *20*, 1231. [[CrossRef](#)]
67. Khan, S.-A.; Li, M.-Z.; Wang, S.-M.; Yin, H.-J. Revisiting the role of plant transcription factors in the battle against abiotic stress. *Int. J. Mol. Sci.* **2018**, *19*, 1634. [[CrossRef](#)]
68. Xing, S.; Salinas, M.; Garcia-Molina, A.; Höhmann, S.; Berndtgen, R.; Huijser, P. SPL8 and miR156-targeted *SPL* genes redundantly regulate *Arabidopsis gynoecium* differential patterning. *Plant J.* **2013**, *75*, 566–577. [[CrossRef](#)]
69. Mena, M.; Ambrose, B.A.; Meeley, R.B.; Briggs, S.P.; Yanofsky, M.F.; Schmidt, R.J. Diversification of c-function activity in maize flower development. *Science* **1996**, *274*, 1537–1540. [[CrossRef](#)]
70. Cho, S.H.; Coruh, C.; Axtell, M.J. miR156 and miR390 Regulate tasiRNA Accumulation and Developmental Timing in *Physcomitrella patens*. *Plant Cell* **2012**, *24*, 4837–4849. [[CrossRef](#)]
71. Unte, U.S.; Sorensen, A.-M.; Pesaresi, P.; Gandikota, M.; Leister, D.; Saedler, H.; Huijser, P. *SPL8*, an SBP-box gene that affects pollen sac development in *Arabidopsis*. *Plant Cell* **2003**, *15*, 1009. [[CrossRef](#)]
72. Xing, S.; Quodt, V.; Chandler, J.; Höhmann, S.; Berndtgen, R.; Huijser, P. SPL8 acts together with the brassinosteroid-signaling component bim1 in controlling *Arabidopsis thaliana* male fertility. *Plants* **2013**, *2*, 416–428. [[CrossRef](#)]

73. Zhang, Y.; Schwarz, S.; Saedler, H.; Huijser, P. SPL8, a local regulator in a subset of gibberellin-mediated developmental processes in *Arabidopsis*. *Plant Mol. Biol.* **2007**, *63*, 429–439. [[CrossRef](#)] [[PubMed](#)]
74. Kim, S.; Soltis, D.E.; Soltis, P.S.; Zanis, M.J.; Suh, Y. Phylogenetic relationships among early-diverging eudicots based on four genes: Were the eudicots ancestrally woody? *Mol. Phylogenet. Evol.* **2004**, *31*, 16–30. [[CrossRef](#)]
75. Allen, E.; Xie, Z.; Gustafson, A.M.; Sung, G.-H.; Spatafora, J.W.; Carrington, J.C. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* **2004**, *36*, 1282–1290. [[CrossRef](#)] [[PubMed](#)]
76. Fenselau de Felippes, F.; Schneeberger, K.; Dezulian, T.; Huson, D.H.; Weigel, D. Evolution of *Arabidopsis thaliana* microRNAs from random sequences. *RNA* **2008**, *14*, 2455–2459. [[CrossRef](#)]
77. Piriyaongsa, J.; Jordan, I.K. Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* **2008**, *14*, 814–821. [[CrossRef](#)] [[PubMed](#)]
78. Li, Y.; Li, C.; Xia, J.; Jin, Y. Domestication of transposable elements into microRNA genes in plants. *PLoS ONE* **2011**, *6*, e19212. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).