Molecular Evolution, Genome-Wide Characterization and Codon Bias Analysis of the SBP-Box Family in Cucumis sativus

CURRENT STATUS: UNDER REVIEW

BMC Genetics  BMC Series

Yue You  Zhejiang Academy of Agricultural Sciences
Yuanting Zheng  Zhejiang Academy of Agricultural Sciences
Jian Wang  Zhejiang Academy of Agricultural Sciences
Sujuan Li  Zhejiang Academy of Agricultural Sciences
Jianfeng Shao  Zhejiang Academy of Agricultural Sciences
Guoan Qi  Zhejiang University
Fei Xu  Zhejiang Academy of Agricultural Sciences
Gangjun Wang  Zhejiang Academy of Agricultural Sciences
Zhonghua Chen  Western Sydney University - Penrith Campus
Shengchun Xu  shengchun2001@163.com  Zhejiang Academy of Agricultural Sciences
Corresponding Author
ORCiD: 0000-0001-8135-0990
| DOI:        | 10.21203/rs.2.17638/v1 |
|------------|-----------------------|
| SUBJECT AREAS | Population Genetics  |
| KEYWORDS   | Codon usage bias, Cucumis sativas, flowering, powdery mildew, SBP-box gene family |
Abstract

Background: SQUAMOSA promoter binding protein (SBP)-box genes encode a group of transcription factors which extensively play essential roles in plant development and stress responses. However, the SBP-box gene family has not been well characterized in cucumber (Cucumis sativas).

Results: In present study, 15 putative SBP-box genes were identified distributing on 4 chromosomes of cucumber. Evolutionary analysis showed that the green plant SBP family originated from a common ancestor. Phylogenetic analysis divided CuSBPs into 6 groups similar to those of Arabidopsis and rice. Intron-exon and motif structure within each group shared common features according to evolutionary study. Expression pattern analysis of transcriptional data about flowering and resistance to powdery mildew demonstrated conserved SBP-box genes function in vegetative-to-reproductive transition and potential roles in other regulatory pathways. Moreover, codon bias analysis explained the mutation and selection pressure exerted on genes.

Conclusions: This study comprehensively characterized cucumber CuSBP gene families, which would provide a foundation to explore the functions of CuSBPs for improving yield, quality and stress tolerance of cucumber in the future.

1 background

Transcription factors (TFs) are proteins binding to DNA sequence in specific manner and activate or inhibit transcription of target gene to specifically express in different tissues and environments. TFs have major roles in regulating plant defense and stress resistance[1]. SBP (SQUAMOSA promoter binding protein)-box genes are
plant-specific TFs that regulate plant development [2, 3] and stress responses [4, 5] by targeting specific genes. Since the first identification in *Antirrhinum majus* [6], SBP-box genes have been identified and functionally characterized in various plant species, including angiosperms such as *Arabidopsis thaliana* [7], rice (*Oryza sativa*) [8], maize (*Zea mays*) [9], tomato (*Lycopersicon esculentum* Mill) [10] and silver birch (*Betula pendula*) [11], as well as early divergent land plants such as moss (*Physcomitrella patens*) [12]. Typically, SBP-box proteins tend to have a highly conserved SBP domain with 76 amino acids containing two zinc finger-like motifs and a nuclear location signal sequence [13].

SBP-box family has been reported to play an important role at various developmental stages of plants, including development of shoot [14], leaf [15], flower [16], the ripening of fruit [17], cellular homeostasis [18] and signal transduction [19]. A major function of the SBP-box family is to participate in the regulation of plant growth and development. In *Arabidopsis*, *SBP-LIKE* (*SPL*) genes, including *AtSPL3*, *AtSPL4* and *AtSPL5*, have been found to be involved in long floral induction [20], and promoting the floral meristem identity transition [3]. *AtSPL8* showed relationship with the pollen sac development [21]. Several *SPLs* were targeted by miR156 to regulate both the juvenile-to-adult vegetative transition and the vegetative-to-reproductive transition, including *AtSPL2*, *AtSPL9*, *AtSPL10*, *AtSPL11*, *AtSPL13* and *AtSPL15* [3]. In maize, it was demonstrated that SBP-box family might be involved in plant growth and development, morphogenesis, adversity response, the development of flower organs and photosynthesis [22]. In addition, the SBP-box family is also involved in response to biotic and abiotic stresses. In *Arabidopsis thaliana*, *AtSPL14* functioned to the presence of a programmed cell death (PCD)-inducing fungal toxin FB1 [23]. *AtSPL6* could positively
regulate N TIR-NB-LRR receptor-mediated plant innate immunity[24]. In birch (Betula platyphylla Suk), BpSPL9 improved the scaveng of ROS produced by salt and drought stress [25]. Moreover, a set of work of SBP-box gene families in other species indicated the diversity of the functions of SBP-box family. In rice, OsSPL14 encoded Ideal Plant Architecture 1 (IPA1), which is a target gene for miR156. A point mutation in OsSPL14 perturbing OsmiR156-directed regulation, increased the expression level of OsSPL14, which would lead to reduced tiller number, increased lodging resistance, and enhanced grain yield [26]. In tomato, silencing the LeSPL-CNR gene could inhibit fruit ripening [17]. In pepper (Capsicum annuum L), CaSBP genes tended to be involved in pathogen resistance pathway against phytophthora capsici, which caused blight and fruit rot [27]. In sorghum (Sorghum bicolor L), SbSBPs tended to be involved in the ABA signaling pathway for abiotic stress response [28].

Cucumber is one of the most popular vegetables in the world. The whole genome sequencing of cucumber has been completed a decade ago [29] and refined in recent years (http://cucurbitgenomics.org/), which provided excellent resources to identify important gene families. Several important TF families, including LOX [30], WRKY [31] and ERF [32], have already been identified and analyzed in cucumber. However, research about SBP-box family has not been reported. In the present study, we conducted genome-wide identification and characterization and molecular evolution analysis of the SBP-box gene family in cucumber. Phylogeny, gene structure and the codon bias were systematically analyzed. Moreover, expression patterns comparison showed that the SBP-box genes potentially participated in flowering and contributed to powdery mildew resistance.
2 results

2.1 Identification of SBP-box genes in cucumber and their chromosome location

A total of 15 \textit{CuSBP} genes were identified in \textit{cucumber}. They were named according to NCBI database (Table 1). The length of genomic sequences of these genes ranged from 2 kb to 12 kb, and the amino acid residues of the predicted proteins ranged from 141 to 1031. The results were quite similar to previous studies\cite{27, 33}, in which the lengths of genes and proteins showed significant variation. \textit{CuSBPs} located on 4 out of the 7 chromosomes, including chr1, chr3, chr4 and chr6 of cucumber.

2.2 Conserved domain and phylogenic analysis

The common feature of SBP-box gene members is that they all have SBP domain, which consists of two Zinc finger-like structures (C3H and C2HC), and a highly conserved nuclear localization signal peptide. In present research, conservative SBP domains were used to do the multiple alignment. The sequences found in \textit{Cucumis sativus} shared all of these three features (Table S1), which proved that these SBP domains were highly conserved in \textit{Cucumis sativus} (Figure 1A). The sequence logo was drawn to show the distribution of amino acid with residue at each site (Figure 1B). Several highly conserved sequences were found at some positions within SBP domains, such as CQQC sequences and KRSCR sequences. Notably, the fourth His in Zn-1 of \textit{CuSBP7a} was different from the others, which was replaced by Cys.

To further investigate the evolutionary relationship of SBP-box genes in \textit{Arabidopsis}, rice and cucumber, a phylogenic tree was constructed using 51 SBP sequences from three species (Figure 2). The sequences were clustered into 6 groups (Group I-
The number of group members varies greatly. The Group IV and VI had the largest number of members, both of which were 14. However, there were only three members in the Group I. The CuSBP family gene was distributed in all six groups and closely related to the Arabidopsis AtSPL family genes and had a relatively distant relationship with rice. Among them, CuSBP7a, CuSBP8, CuSBP9, CuSBP14, CuSBP1a, CuSBP6, CuSBP13a and CuSBP13b were distributed in the same sub-branches with Arabidopsis homologous genes.

2.3 Structural organization and conserved motif analysis of CuSBP genes

A phylogenetic tree for cucumber was constructed to further understand the structures of CuSBP (Figure S1). The exon-intron structures were generated based on their genome and coding sequences. The results revealed high variation in the number of introns, ranged from 2 to 10. Group I proteins contain 10 introns, Group II contains 2, Group III contains 2–3, Group IV contains 3, Group V contains 10, and Group VI contains 3 introns, which was in agreement with previous studies [27, 34]. Thirty conserved motifs were characterized in CuSBPs of cucumber by MEME (Figure S2). The number of motif types varied from 4 (CuSBP1b, CuSBP1c, CuSBP3) to 13 (CuSBP1a, CuSBP14). Generally, Motif1, 2 and 3 constitutes the SBP domain. Group VI had more diversity of motifs because it contained largest number of members in SBP gene family. CuSBP1a and CuSBP14 both belonged to Group V, and shared exactly same motifs, in which motif 8 and 30 were specially presented. Motif 4, 17, 12, and 9 were shared by Group I and Group V. Motif 28 was the only repeated motifs, which repeated twice in CuSBP3.

2.4 Cis-element analysis of CuSBP genes

The analysis of cis component in promoter sequence is helpful to understand gene
regulation patterns. In present study, cis-acting regulatory elements’ function in promoter regions of CuSBP genes were classified into two types (Table S2). Type I involved in stress responses, such as ABRE (involved in the abscisic acid responsiveness), MBS (MYB binding site involved in drought-inducibility), ARE (essential for the anaerobic induction). Type II included elements involved in development process, like CAT-box (related to meristem expression), GCN4_motif (involved in endosperm expression), and P-box (gibberellin-responsive element). The distribution of these elements varied a lot among CuSBPs, even if belonging to the same group.

2.5 Expression patterns of CuSBP genes in different periods of flowering

The transcriptomes dataset of different flowering periods in cucumber were downloaded from NCBI to investigate the expression patterns of CuSBP genes during flowering. Usually, flowers enter the stage of green bud at day 1, green yellow bud stage at day 3, yellow bud stage at day 4, and flowering stage at day 5. CuSBP genes with fragments per kilobase million (RPKM) value >5.0 were regarded as expressed genes (Figure 3). Most of gene members shared overlapping expression pattern during flowering period, only some of them had difference at specific stages. In general, we divided them into 3 types based on their expression. The first type expressed at a low level or had no expression, including CuSBP1b, CuSBP6, CuSBP7b, CuSBP8, CuSBP13a and CuSBP13b. CuSBP1b and CuSBP8 both belonged to Group III. CuSBP1b had low expression at day1, while CuSBP8 had no expression at day 5. The second type includes CuSBP3, CuSBP7a, CuSBP9, CuSBP12 and CuSBP16. They had high expression level at the early flowering stage, and the expression level decreased along the timeline. The third type expressed at a high level and
especially high at a specific development stage. *CuSBP1a* and *CuSBP14* were members in Group V. They were both up-regulated at day 3 and day 5 while at day 1 and day 4 they were down-regulated to same degree. Taking promoter analysis into consideration, type three genes’ promoter regions all had light responsive elements (Table S2).

2.6 Expression profile of Cucumber SBP-box genes against powdery mildew

Powdery mildew is one of severe diseases in cucumber, which could decline the production dramatically. To study the role of SBP-box gene family against powdery mildew, we used the leaf transcriptional data of powdery mildew resistant segment substitution line SSL508-28 and the parent D8 with inoculation of 48h (PM) and no inoculation (CT) to explore the role of SBP gene families in powdery mildew resistance (Figure 4). *CuSBP* genes with FPKM value >5.0 are regarded as expressed genes.

Ten members of *CuSBPs* were clustered according to their expression level without *CuSBP3*, *CuSBP6*, *CuSBP8*, *CuSBP9* and *CuSBP13b*(Figure 4). *CuSBP14* and *CuSBP1a* showed high expression level both in PM treatment and the control of the two germplasms. *CuSBP1c*, *CuSBP7a*, *CuSBP12* and *CuSBP13c* all showed down-regulated expression and *CuSBP7b*, *CuSBP16* and *CuSBP13a* showed opposite trends after PM treatment in two germplasms. Notably, *CuSBP13a* showed significantly higher expression in PM treatment of SSL508-28 than the control (about 20 folds). The expression of *CuSBP1b* were opposite in different germplasms.

2.7 Codon usage analysis of Cucumber SBP-box genes

Codon degeneracy and codon usage are interesting and challenging problems to be
investigated, which are related to gene function or sequence properties. To study the codon usage in CuSBPs, we calculated the GC content on the 1st, 2nd and 3rd site of a synonymous codon (GC1, 2, 3s content), relative synonymous codon usage, codon adaptation index and effective number of codons. The calculated values are shown in Table S3. Firstly, the codon usage pattern was studied. The GC content is 46.32±3.49% with GC1 51.09±2.94%, GC2 45.66±3.70%, GC3 42.20±7.31%. The GC content is relatively lower at the 3rd site, while with relatively big variance. The neutrality plot (GC12 vs GC3s) was drawn (Figure 5a), in which the correlation of the points is not significant (P>0.05) and the slope of the regression line is close to 0. Association between ENC and GC3s were learned with a standard curve shown in (Figure 5b) under the hypothesis that there is no selection.

3 Discussion

3.1 Evolution of cucumber SBP family genes

The SBP gene family plays an important role in various stages of plant development, such as stem growth[22], leaf development[35], flowering[14], fruit ripening[36] and signal transduction[21]. Phylogenetic analysis of multiple plants species focusing on SBP-box gene family has been reported, such as Arabidopsis(16 members), rice(19 members), Betula luminifera(18 members)[25], common bean(23 members) [37], grape(19 members)[38], Ziziphus jujube(16 members)[39], Petunia(21 members)[40]. In this study, we identified 15 SBP family genes, all of which contain a conserved SBP domain. To investigate the homology of SBP family genes in cucumber, Arabidopsis and rice, phylogenetic tree analysis showed that the SBP genes were clustered into 6 groups (Group I-Group VI) (Figure 2). Preston divided the SBP family genes of 9 species into eight clades[13]. In present study, dividing
Group IV and Group VI from the first node respectively into Group IV -1, Group IV -2, Group VI -1, Group VI -2 generates 8 same clades as previous studies. However, there is no CuSBP in Group IV -2.

A recent study on SBP-box genes elaborated that land plants’ SBP-box genes were generated through duplication events from one common ancestor. Based on the timeline of duplications, SBP-box genes are divided into group 1 and group 2-1, group 2-2, among which group 2-2 plays the major role of the expansion of SBP-box genes.[13] Group I in present study corresponds to group1, which is the first formed lineage. Group V in present study corresponds to group 2-1, which is the subgroup in the second group retaining similar evolutionary features to group 1 compared with group 2-2. We further constructed a phylogenetic tree of cucumber SBP genes and analyzed their conserved motifs and introns (Figure S1, S2). The number of introns in the CuSBP genes ranged from 2 to 10, motifs from 4 to 13. From the evolutionary point of view, the genes assigned under the same sub-group have similar numbers, types and distributions of motifs and introns, such as CuSBP14 and CuSBP1a have identical conservative motifs in Group V, which further support the grouping of the phylogenetic tree. Some genes have different isoforms, which may be caused by paralogous genes of different chromosomes or by alternative splicing, such as CuSBP13a, CuSBP13b and CuSBP13c, which are assigned to the same sub-group. However, CuSBP1a, CuSBP1b and CuSBP1c are assigned to different sub-groups.

3.2 Expression pattern and potential function analysis of CuSBP genes

SBP-box genes are plant-specific transcription factors, which play important roles in plant growth and development processes. They usually exhibit high expression level
in panicles, flower buds, shoot apices in studied species[41-43]. At the same time, some SBP family genes are also involved in some biotic and abiotic stress responses.[4, 25] Currently, there are few functional data available describing SBP-box genes in cucumber.

To explore potential function of CuSBPs in cucumber, two sets of transcriptional data (flowering and PM) are used. Flowering serves as transition from vegetative phase to reproductive phase, which makes it crucial in reproductive plants. Therefore, it is important to study the expression pattern of CuSBP family genes in cucumber flowering process, discover its potential functions in flowering in order to provide potential resources for future cucumber breeding to improve cucumber yield. From the expression patterns of CuSBP family genes in different flower development stages, they could be divided into two groups, one group had lower expression levels during flowering (CuSBP8/6/1b/13a/7b/13b) The other group has a higher expression level (CuSBP3/12/16/7a/9/1c/14/13c/1a). It is speculated that genes with low expression during flowering may be involved in the vegetative growth process of plants. CuSBP8, CuSBP1b and OsSBP10 are members of Group III, and their expression patterns are similar. OsSBP10 was shown to be highly expressed in rice seedlings and young spikelets[44]. CuSBP13 homologous gene AtSPL13 expresses highly in hypocotyl, shoot apical meristem, leaf primordia and developing inflorescence. Disruption of AtSPL13 regulation delays post-germinative transition from the cotyledon to vegetative-leaf stage.[35] In opposite, CuSBP14 and CuSBP1a have high expression levels throughout the flowering process and are distributed in the Group V, which may play a key role in the flowering process. Its mutant fbr6, a homologous gene of AtSBP14 in Arabidopsis, showed a transition to flowering later than wild type[23]. Group II contains CuSBP1c, AtSPL3/4/5 and
OsSPL13. In Arabidopsis, SPL3 3’ UTR recognized by miR156/157 prevents early flowering[16]. SOC1 (Suppressor of overexpression of constans1) integrates photoperiod and GA signals to promote flowering via the SPL3/4/5[19]. CuSBP1c homologous gene AmSBP1 also acts in the initiating flower development[45]. OsSPL13 was found associated with spikelet size in cultivated rice by a GWAS analysis[46]. CuSBP1c shows decreased expression level as the flower buds develop, indicating similar roles with AtSPL3/4/5. In GroupV, CuSBP9 expresses less as the flowering time goes. AtSPL2 have also been found to regulate ASYMMETRIC LEAVES 2 (AS2) to control floral organ development[36]. OsSPL14 controls shoot branching in the vegetative stage with the regulation of miRNA[26], pointing out that CuSBP9 potential role in juvenile-to-adult vegetative transition. 

Powdery mildew is one of the world most damaging diseases to cucumber, which is caused mainly by Podosphaera fusca [47]. It seriously affects photosynthesis and disturb metabolism, resulting in premature aging and declined production. Therefore, it is important to study the resistance functions of CuSBP family genes involved in the regulation of powdery mildew. In GroupV, CuSBP14 and CuSBP1a maintained high expression levels in both cultivars. After treatment with PM, the expression levels of the two genes in SSL508-28 were lower than those in D8. CuSBP14, a homologous gene of CuSBP14 in Arabidopsis, is more sensitive to programmed cell death (PCD)-inducing fungal toxin FB1. This indicates that CuSBP14 and CuSBP1c may be involved in the negative regulation of biotic stress. CuSBP7a has a lower expression level in SSL508-28 compared to D8, and AtSBP7, OsSBP9 and CuSBP7a belong to the Group 1. AtSPL7 and OsSPL9 are considered functional genes in copper regulation pathways and Loss-of-function mutations in SPL9 resulting in enhanced plant resistance to rice stripe virus[35, 48].
CuSBP1b and CuSBP13a were highly up-regulated after treatment with powdery mildew in SSL508-28. This indicates that these two genes may play a positive regulatory role in the regulation of powdery mildew resistance. CuSBP12 was downregulated after treatment of two varieties of PM, but CuSBP6 with the highest homology was not detected in this database. AtSBP6 is a homologous gene of CuSBP6/12, which can actively regulate the defense genes and regulate Plant innate immune system. NbSPL6 is essential for the N-mediated resistance to Tobacco mosaic virus.[24] This also indicates that homologous genes of different species may differ in functional evolution, but the specific function remains to be studied.

3.3 SBP-box genes had high codon usage bias in cucumber

Codon usage bias reflects the genetic information hidden in the RNA sequence and accessed us to the evolutionary process of genes in organism. In present work, we find high codon usage bias in cucumber SBP-box genes.

ENC is a parameter acknowledged to be used to measure the codon bias degree. Previous studies demonstrate that gene expression is negatively correlated with ENC value[49, 50], in other words, important highly expressed genes tend to have low ENC values. The ENC value of CuSBPs varies from 49 to 61 with the mean value 53.2, which indicates high codon bias. Lowest ENC values among them is 49, indicating severe selection exerted on them and they are well functionalized, which is identical to previous discussion. Group III members have top 3 ENC values, suggesting their roles may not be irreplaceable. In ENC-plot, notably, CuSBPs have low ENC value compared to the ENC expected value, rejecting the null hypothesis that there is no selection. The presented gene points are lying below the curve expect one and the genes were not narrowly distributed in the plot, which
demonstrated that both mutation pressure and selection affected the codon usage pattern, reinforcing the theory of equilibrium between mutation and selection[51]. In the neutrality plot, the correlation of the points is not significant (\(P>0.05\)) and the slope of the regression line is close to 0. The weak correlation between two GC12 and GC3s suggests that there is high mutation bias or low conservation of GC content levels, which means the mutation pressure rather than translational selection plays the major role.

RSCU is commonly used to analysis the synonymous codon usage (Table S4). Additionally, most abundantly used codons are A/T ended, as the result of compositional constrains (i.e., A and T)[52], which is part of the mutation pressure. For subsequent researches, if transgenic SBP-box genes are required to be expressed in cucumber, the insertion sequence can be modified by preferred codon patterns in this study.

4 conclusion

In summary, the present provided a comprehensive understanding of SBP-box genes in cucumber. Genome-wide identification and characterization of SBP-box genes have been done in cucumber including phylogenic, gene structure and promoter analysis. Expression pattern analysis of flowering and powdery mildew resistance indicated that various potential roles of SBP-box gene in cucumber need to be explored. Moreover, codon usage analysis of CuSBPs could provide us the image of molecular evolution and further transformation work.

5 materials and methods

5.1 Identification of SBP-box genes in cucumber
The whole genome information of cucumber was downloaded from NCBI genome database (https://www.ncbi.nlm.nih.gov). The protein sequences of SBP-box genes from both *Arabidopsis thaliana* and *Oryza sativa* were downloaded from Uniprot as ‘input files’ (https://www.uniprot.org/). Then, two methods were used to genome-widely identify candidate SBP-box genes in *Cucumis sativus*: (1) the ‘input files’ were blasted against cucumber genome with E-value < 1e-5 in order to obtain putative SBP-box family genes, (2) HMMER3 were used to get the HMM model and results [53]. These two methods led to the same result. After that, we conformed the sequences by Pfam (http://pfam.xfam.org/) and batch cd-search in NCBI.

### 5.2 Sequence alignment and phylogenetic analysis

Multiple amino acid sequence alignment was performed using DNAMAN software (Lynnon Biosoft, CA, USA). The sequence logo was obtained using the online platform Weblogo4 for conserved sequences. Phylogenetic trees were constructed using MEGA 7.0 (https://www.megasoftware.net/index.php) with the neighbor-joining method and 1000 bootstrap replicates for SBP-box family genes in *Arabidopsis thaliana*, *Oryza sativa* and *Cucumis sativus*.

### 5.3 Intron-exon structure analysis

Intron/Exon structures were determined by aligning coding sequences to their corresponding genomic sequences. A diagram of intron/exon structures was obtained using TBtools [54].

### 5.4 Motif analysis

Multiple EM for Motif Elicitation (MEME) software (http://meme.nbcr.net/meme/) was used to search for motifs in all 15 SBP-box genes. The number of motifs was set to 30.
5.5 Promoter analysis

Plantcare (http://bioinformatics.psb.ugent.be/webtools/plantcare/) was used to analyze the cis-acting regulatory elements of CuSBPs.

5.6 Expression analysis of CuSBPs

Two sets of transcriptome data of Cucumis sativus were obtained from NCBI database to study the expression profiles of CuSBP genes. Data (GSE76358) including samples of different flowering periods were downloaded from the database reported previously by Sun et al. [55]. The normal ovary blooms at 4-5 days after labeling (when the ovary is visible), with the majority blooms at day 5. Corolla of day1, day3, day4, and day5 after labeling were collected for RNA sequencing. Log$_2$ RPKM values were used to draw the heatmap. Another data set was genes expression pattern against powdery mildew with GEO accession: GSE81234. To investigate the candidate genes governing Pm5.1 and their effects on powdery resistance, the RNA-sequencing based transcriptomes of the powdery mildew resistant segment substitution line SSL508-28 and recurrent parent D8 were compared 48h after inoculation with the PM pathogen. Log$_2$ FPKM values were used to draw a heatmap.

5.7 CuSBPs codon usage analysis

To study the codon usage in CuSBPs, we calculated the GC content on the 1st, 2nd and 3rd site of a synonymous codon (GC1, 2, 3s content), relative synonymous codon usage (RSCU), codon adaptation index (CAI)[56], as well as effective number of codons (ENC) using online website Optimizer (http://genomes.urv.es/OPTIMIZER/) [57]. The correlation between exon length, CAI, GC, GC3, GC2, GC1, and ENC were
calculated by R software.

To further analyze codon usage bias, plot of ENC vs GC3s (ENC plot) and plot of GC12 and GC3s (Neutrality Plot) were generated according to previous study. In ENC plot, the standard cure was calculated by:

$$\text{ENC}_{\text{exp}} = 2 + S + \left(\frac{29}{S^2 + (1 - S^2)}\right)$$

Where, S is the frequency of G + C (i.e., GC3s). And in Neutrality Plot, GC12 is the frequency of GC at 1st and 2nd sites.

6 declarations

6.1 Ethics approval and consent to participate

Not applicable.

6.2 Consent for publication

Not applicable.

6.3 Availability of data and materials

The datasets analysed during the current study are available in the NCBI genome database (https://www.ncbi.nlm.nih.gov). The protein sequences of SBP-box genes from both Arabidopsis thaliana and Oryza sativa were downloaded from Uniprot as ‘input files’ (https://www.uniprot.org/). Two sets of transcriptomes data of Cucumis sativus were obtained from NCBI database to study the expression profiles of CuSBP genes. Data (GSE76358) including samples of different flowering periods were downloaded from the database reported previously by Sun et al. [55]
6.4 Competing interests

The authors declare that they have no competing interests.

6.5 Funding

The research was supported by National Key R&D Program of China (2018YFD0100901). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

6.6 Authors’ contributions

Conceived and designed the experiments: SX, YY, YZ, JW. Performed the experiments: YY, YZ, SL, JS. Analyzed the data: YY, YZ, GQ, FX, GW. Contributed reagents/materials/analysis tools: YY, YZ, ZC. Wrote the paper: YY, YZ, ZC, SX.

6.7 Acknowledgements

Not applicable

7 references

1. Singh K, Foley RC, Onate-Sanchez L: Transcription factors in plant defense and stress responses. Curr Opin Plant Biol 2002, 5(5):430-436.

2. Ferreira e Silva GF, Silva EM, Azevedo Mda S, Guivin MA, Ramiro DA, Figueiredo CR, Carrer H, Peres LE, Nogueira FT: microRNA156-targeted SPL/SBP box transcription factors regulate tomato ovary and fruit development. Plant J 2014, 78(4):604-618.
3. Xu M, Hu T, Zhao J, Park MY, Earley KW, Wu G, Yang L, Poethig RS:

Developmental Functions of miR156-Regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Genes in Arabidopsis thaliana. *PLoS Genet* 2016, 12(8):e1006263.

4. Singh RP, Jha PN: The PGPR Stenotrophomonas maltophilia SBP-9 Augments Resistance against Biotic and Abiotic Stress in Wheat Plants. *Front Microbiol* 2017, 8:1945.

5. Singh RP, Runthala A, Khan S, Jha PN: Quantitative proteomics analysis reveals the tolerance of wheat to salt stress in response to *Enterobacter cloacae* SBP-8. *PLoS One* 2017, 12(9):e0183513.

6. Klein J, Saedler H, Huijser P: A new family of DNA binding proteins includes putative transcriptional regulators of the Antirrhinum majus floral meristem identity gene SQUAMOSA. *Mol Gen Genet* 1996, 250(1):7-16.

7. Cardon GH, Hohmann S, Nettesheim K, Saedler H, Huijser P: Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in the floral transition. *Plant J* 1997, 12(2):367-377.

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

9. Zhang W, Li B, Yu B: Genome-wide identification, phylogeny and expression analysis of the SBP-box gene family in maize (*Zea mays*). *J Integr Agr* 2016, 15(1):29-41.

10. Salinas M, Xing SP, Hohmann S, Berndtgen R, Huijser P: Genomic organization, phylogenetic comparison and differential expression of the SBP-box family of transcription factors in tomato. *Planta* 2012,
11. Lannenpaa M, Janonen I, Holtta-Vuori M, Gardemeister M, Porali I, Sopanen T: A
new SBP-box gene BpSPL1 in silver birch (Betula pendula). *Physiol Plantarum* 2004, 120(3):491-500.

12. Riese M, Hohmann S, Saedler H, Munster T, Huijser P: *Comparative analysis of the SBP-box gene families in P-patens and seed plants*. *Gene* 2007, 401(1-2):28-37.

13. Zhang SD, Ling LZ, Yi TS: *Evolution and divergence of SBP-box genes in land plants*. *BMC Genomics* 2015, 16:787.

14. Wu G, Poethig RS: *Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3*. *Development* 2006, 133(18):3539-3547.

15. Usami T, Horiguchi G, Yano S, Tsukaya H: *The more and smaller cells mutants of Arabidopsis thaliana identify novel roles for SQUAMOSA PROMOTER BINDING PROTEIN-LIKE genes in the control of heteroblasty*. *Development* 2009, 136(6):955-964.

16. Gandikota M, Birkenbihl RP, Hohmann S, Cardon GH, Saedler H, Huijser P: *The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings*. *Plant Journal* 2007, 49(4):683-693.

17. Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB: *A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening*. *Nat Genet* 2006, 38(8):948-952.

18. Yamasaki H, Hayashi M, Fukazawa M, Kobayashi Y, Shikanai T: *SQUAMOSA*
Promoter Binding Protein-Like7 Is a Central Regulator for Copper Homeostasis in Arabidopsis. Plant Cell 2009, 21(1):347-361.

19. Jung JH, Ju Y, Seo PJ, Lee JH, Park CM: The SOC1-SPL module integrates photoperiod and gibberellic acid signals to control flowering time in Arabidopsis. Plant J 2012, 69(4):577-588.

20. Jung JH, Lee HJ, Ryu JY, Park CM: SPL3/4/5 Integrate Developmental Aging and Photoperiodic Signals into the FT-FD Module in Arabidopsis Flowering. Mol Plant 2016, 9(12):1647-1659.

21. Xing S, Quodt V, Chandler J, Hohmann S, Berndtgen R, Huijser P: SPL8 Acts Together with the Brassinosteroid-Signaling Component BIM1 in Controlling Arabidopsis thaliana Male Fertility. Plants (Basel) 2013, 2(3):416-428.

22. Chuck GS, Brown PJ, Meeley R, Hake S: Maize SBP-box transcription factors unbranched2 and unbranched3 affect yield traits by regulating the rate of lateral primordia initiation. Proc Natl Acad Sci U S A 2014, 111(52):18775-18780.

23. Stone JM, Liang X, Nekl ER, Stiers JJ: Arabidopsis AtSPL14, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1. Plant J 2005, 41(5):744-754.

24. Padmanabhan MS, Ma S, Burch-Smith TM, Czymmek K, Huijser P, Dinesh-Kumar SP: Novel positive regulatory role for the SPL6 transcription factor in the N TIR-NB-LRR receptor-mediated plant innate immunity. PLoS Pathog 2013, 9(3):e1003235.

25. Ning K, Chen S, Huang HJ, Jiang J, Yuan HM, Li HY: Molecular characterization and expression analysis of the SPL gene family with
BpSPL9 transgenic lines found to confer tolerance to abiotic stress in *Betula platyphylla Suk*. *Plant Cell Tiss Org* 2017, **130**(3):469-481.

26. Jiao YQ, Wang YH, Xue DW, Wang J, Yan MX, Liu GF, Dong GJ, Zeng DL, Lu ZF, Zhu XD *et al*: *Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice*. *Nat Genet* 2010, **42**(6):541-U536.

27. Zhang HX, Jin JH, He YM, Lu BY, Li DW, Chai WG, Khan A, Gong ZH: *Genome-Wide Identification and Analysis of the SBP-Box Family Genes under Phytophthora capsici Stress in Pepper (Capsicum annuum L.).* *Front Plant Sci* 2016, **7**:504.

28. Chang JZ, Yan FX, Qiao LY, Zheng J, Zhang FY, Liu QS: *Genome-wide identification and expression analysis of SBP-box gene family in Sorghum bicolor L*. *Yi Chuan* 2016, **38**(6):569-580.

29. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P *et al*: *The genome of the cucumber, Cucumis sativus L*. *Nat Genet* 2009, **41**(12):1275-1281.

30. Liu SQ, Liu XH, Jiang LW: *Genome-wide identification, phylogeny and expression analysis of the lipoxygenase gene family in cucumber*. *Genet Mol Res* 2011, **10**(4):2613-2636.

31. Ling J, Jiang W, Zhang Y, Yu H, Mao Z, Gu X, Huang S, Xie B: *Genome-wide analysis of WRKY gene family in Cucumis sativus*. *BMC Genomics* 2011, **12**:471.

32. Hu L, Liu S: *Genome-wide identification and phylogenetic analysis of the ERF gene family in cucumbers*. *Genet Mol Biol* 2011, **34**(4):624-633.

33. Song S, Zhou HY, Sheng SB, Cao M, Li YY, Pang XM: *Genome-Wide Organization and Expression Profiling of the SBP-Box Gene Family in...*
34. Zhou Q, Zhang SS, Chen F, Liu BJ, Wu L, Li F, Zhang JQ, Bao MZ, Liu GF: **Genome-wide identification and characterization of the SBP-box gene family in Petunia.** *Bmc Genomics* 2018, 19.

35. Martin RC, Asahina M, Liu PP, Kristof JR, Coppersmith JL, Pluskota WE, Bassel GW, Goloviznina NA, Nguyen TT, Martínezandújar C: **The regulation of post-germinative transition from the cotyledon- to vegetative-leaf stages by microRNA-targeted SQUAMOSA PROMOTER-BINDING PROTEIN LIKE13 in Arabidopsis.** *Seed Science Research* 2010, 20(2):89-96.

36. Wang Z, Wang Y, Kohalmi SE, Amyot L, Hannoufa A: **SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 2 controls floral organ development and plant fertility by activating ASYMMETRIC LEAVES 2 in Arabidopsis thaliana.** *Plant Mol Biol* 2016, 92(6):661-674.

37. Hu L, Wang J, Liu H, Zhao J, Wang Q, Ji X: **Molecular Characteristic and Evolution Analysis of the SBP Gene Family in Common Bean.** *Molecular Plant Breeding* 2018.

38. Hongmin H, Jun L, Min G, Singer SD, Hao W, Linyong M, Zhangjun F, Xiping W: **Genomic organization, phylogenetic comparison and differential expression of the SBP-box family genes in grape.** *Plos One* 2013, 8(3):e59358.

39. Song S, Zhou H, Sheng S, Cao M, Li Y, Pang X: **Genome-Wide Organization and Expression Profiling of the SBP-Box Gene Family in Chinese Jujube (Ziziphus jujuba Mill.).** *Int J Mol Sci* 2017, 18(8):1734.

40. Zhou Q, Zhang S, Chen F, Liu B, Wu L, Li F, Zhang J, Bao M, Liu G: **Genome-wide identification and characterization of the SBP-box gene family in**
Petunia. *Bmc Genomics* 2018, **19**(S2):193.

41. Li J, Hou H, Li X, Xiang J, Yin X, Gao H, Zheng Y, Bassett CL, Wang X: Genome-wide identification and analysis of the SBP-box family genes in apple (*Malus x domestica Borkh*.). *Plant Physiol Biochem* 2013, **70**:100-114.

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

43. Xie K, Wu C, Xiong L: Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol* 2006, **142**(1):280-293.

44. Tao L, Zhang S, Liu T, Wang B, Guan H, Zhou Y, Duan Y, Wu W: Fine mapping and candidate identification of SST, a gene controlling seedling salt tolerance in rice (*Oryza sativa L*.). *Euphytica* 2015, **205**(1):269-274.

45. Preston JC, Hileman LC: Functional Evolution in the Plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family. *Front Plant Sci* 2013, **4**:80.

46. Si L, Chen J, Huang X, Gong H, Luo J, Hou Q, Zhou T, Lu T, Zhu J, Shangguan Y et al: *OsSPL13 controls grain size in cultivated rice*. *Nat Genet* 2016, **48**(4):447-456.

47. Perez-Garcia A, Romero D, Fernandez-Ortuno D, Lopez-Ruiz F, De Vicente A, Tores JA: The powdery mildew fungus *Podosphaera fusca* (synonym *Podosphaera xanthii*), a constant threat to cucurbits. *Mol Plant Pathol* 2009, **10**(2):153-160.

48. Tang M, Zhou C, Meng L, Mao D, Peng C, Zhu Y, Huang D, Tan Z, Chen C, Liu C
et al: Overexpression of OsSPL9 enhances accumulation of Cu in rice grain and improves its digestibility and metabolism. J Genet Genomics 2016, 43(11):673-676.

49. Bulmer M, . The selection-mutation-drift theory of synonymous codon usage. Genetics 1991, 129(3):897-907.

50. Carbone A, Zinov'ev A, Kepes F: Codon adaptation index as a measure of dominating codon bias. Bioinformatics 2003, 19(16):2005-2015.

51. Prabha R, Singh DP, Sinha S, Ahmad K, Rai A: Genome-wide comparative analysis of codon usage bias and codon context patterns among cyanobacterial genomes. Mar Genom 2017, 32:31-39.

52. Bernardi G, Bernardi G: Compositional Constraints and Genome Evolution. J Mol Evol 1986, 24(1-2):1-11.

53. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M: Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res 2013, 41(12):e121.

54. Chen C, Rui X, Hao C, He Y: TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. 2018.

55. Sun CZ, Li YQ, Zhao WS, Song XF, Lu M, Li XL, Li XX, Liu RY, Yan LY, Zhang XL: Integration of Hormonal and Nutritional Cues Orchestrates Progressive Corolla Opening in the female flower in cucumber. Plant Physiol 2016, 171(2):1209-1229.

56. Sharp PM, Li WH: The Codon Adaptation Index - a Measure of Directional Synonymous Codon Usage Bias, and Its Potential Applications. Nucleic Acids Research 1987, 15(3):1281-1295.

57. Wright F: The Effective Number of Codons Used in a Gene. Gene 1990,
### Tables

#### Table 1 The SBP gene family in *Cucumis sativus.*

| Gene Name | Gene ID      | Chromosome | Start   | End      | Gene Length | Amino Ac Length |
|-----------|--------------|------------|---------|----------|-------------|-----------------|
| CuSBP7b   | 105434392    | 3          | 6445615 | 6448269  | 2655        | 339             |
| CuSBP1b   | 101204516    | 6          | 27284737| 27286763 | 2027        | 141             |
| CuSBP9    | 101203301    | 3          | 30985523| 30991713 | 6191        | 382             |
| CuSBP12   | 101206747    | 1          | 3934091 | 3940111  | 6021        | 548             |
| CuSBP13b  | 101222787    | 1          | 236064  | 239806   | 3743        | 379             |
| CuSBP8    | 101204773    | 6          | 7313524 | 7317367  | 3844        | 297             |
| CuSBP16   | 101217032    | 6          | 6509781 | 6513088  | 3308        | 328             |
| CuSBP14   | 101219733    | 4          | 23212515| 23218369 | 5855        | 1031            |
| CuSBP1c   | 101215375    | 4          | 20657647| 20658784 | 1138        | 202             |
| CuSBP13c  | 101219589    | 3          | 22275396| 22277307 | 1912        | 344             |
| CuSBP1a   | 101203103    | 3          | 25952433| 25964518 | 12086       | 1013            |
| CuSBP6    | 101221286    | 1          | 2087516 | 2093881  | 6366        | 550             |
| CuSBP7a   | 101205874    | 3          | 10311215| 10321131 | 9917        | 793             |
| CuSBP13a  | 101221267    | 1          | 6001467 | 6004185  | 2719        | 314             |
| CuSBP3    | 101212580    | 1          | 7700070 | 7703680  | 3611        | 162             |

#### Figures

27
SBP domain alignment in CuSBPs. (A) Multiple alignment of SBP domain protein sequences. The Zn1 and Zn2 elements are boxed in red, and conserved zinc-finger structures (C3H, C2H) and NLS are marked. (B) Sequence logo of the SBP domain of CuSBPs.

Figure 1

Phylogenetic tree of cucumber, Arabidopsis and rice. The SBP domain protein sequences are divided into 6 groups.

Figure 2
Expression analysis of CuSBPs in different flowering process. Ovaries are used to...
Expression analysis of CuSBPs in ‘SSL508-28’ and ‘D8’ with and without inoculation.
Figure 5

Codon usage analysis. (A) Neutrality Plot. GC12 and GC3s values of CuSBPs are used to draw a scatterplot with a standard curve. (B) ENC plot. ENC and GC3s values of CuSBPs are used to draw a scatterplot with a standard curve.

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.

supplementary.rar