Identification of SBP Gene Family and Analysis of Expression Patterns under salt stress in Quinoa 1

Wei-hai Hou
Tibet Agricultural and Animal Husbandry University
Jian-ling Wang
Tibet Agricultural and Animal Husbandry University
Zong-yu Zhang
Tibet Agricultural and Animal Husbandry University
Inzamam Ul Haq
Gansu Agricultural University
Xibo Feng (fxb750217@126.com)
Tibet Agricultural and Animal Husbandry University

Research Article

Keywords: Quinoa, SBP gene family, Bioinformatics, Salt stress, Expression analysis

Posted Date: June 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1742404/v1

Abstract

This study identified the SQUAMOSA Promoter Binding Proteins transcription factor SBP from the whole genome of Quinoa by analyzing the phylogenetic relationship, gene structure, chromosome location and gene replication, upstream cis-regulatory element, tissue expression, and construction of the SBP protein interaction network, as well as analyzing the characteristics of the SBP tissue expression pattern under salt stress. It serves as the foundation for the study and application of Quinoa SBP. The results showed that 23 genes of the CqSBP family were identified by bioinformatics analysis, and their amino acid physical and chemical properties showed diversity. The family proteins were hydrophilic, and the subcellular localization showed that 23 CqSBPs were located in the nucleus. The SBP family genes are unevenly distributed on the 12 chromosomes of Quinoa; the upstream cis-acting element analysis shows that there are 49 elements with plant hormones, stress, light response, and tissue-specific expression, and all CqSBPs contain one or more TATA box elements. Protein interaction network analysis shows that all CqSBP proteins appear in the known interaction network of Arabidopsis; the expression of different SBP genes in different organs and periods of Quinoa is different, and the expression of SBP genes has certain tissue specificity. The expression of CqSBP showed many changes under salt stress.

1. Introduction

Transcription factors are a class of proteins that regulate transcription. They can activate or inhibit the transcription of target genes by binding to specific sequences of DNA, thereby regulating gene expression (Riese et al.2007). SQUAMOSA Promoter Binding Proteins (SBPs) are plant-specific transcription factors with multiple members. Klein (Klein et al.1996) first isolated two proteins from the SBP-box family, AMPSBP1 and AMPSBP2, while studying the regulation of gene expression networks. With the completion of more and more plant genome sequencing, the SBP transcription factor family has been widely identified in the whole genome. SBP-box family genes have also been identified in Arabidopsis, rice, and other plants (Yang et al. 2007; Zhang et al. 2017). Members of this family are widely involved in plant growth, development, and various physiological and biochemical processes (Guo et al.2008).

The sequence analysis of these SBP proteins revealed that all known SBP transcription factor member proteins contained a highly conserved DNA binding domain consisting of 79 amino acid residues, which included two zinc finger domains (Zinc finger domain) and 1 Two-way nuclear localization signal (NLS) (Birkenbihl et al.2005). The DNA binding domain of the Arabidopsis SBP transcription factor comprises two separate zinc finger structures, composed of 8 amino acid residues and Zn^{2+}, and the 8 amino acid residues include histidine and cysteine acid. The two zinc fingers are composed of (Cys3HisCys2HisCys) and (Cys6HisCys). Zn^{2+} binds to the front and rear 4 amino acid residues (Yamasaki et al.2004). Zn^{2+} and nuclear localization signals are necessary for the process of protein-DNA binding. The C-terminal NLS partially overlaps with the C2HC-type zinc finger structure sequence, and it has the function of guiding the SBP-box gene into the nucleus to regulate the transcription and expression of its downstream genes. The revelation of these structural features will help to analyze further their regulatory role in the growth and development of higher plants. At the same time, studies have pointed out that miR156 regulates many SBP transcription factor family members. At present, 11 SPL genes have been found to contain miR156 recognition sites in both Arabidopsis and rice (Rhoades et al.2002). The spatiotemporal expression of SPL3 during the vegetative development of Arabidopsis is regulated by miR156 (Schwab et al.2005). Tissue-specific interactions between MIR156 and certain OsSPL genes also occur in rice (Xie et al.2006).

Currently, the functions of many SBP-box genes have been identified in many species. For example, SPL2, SPL10, and SPL11 genes are involved in bud morphogenesis in Arabidopsis Thaliana (Shikata et al.2009). And overexpression of ATSPL1 or AtSPL2 enhanced heat tolerance in Arabidopsis Thaliana and tobacco (L.M. Chao et al.2017). AtSPL7

induces the flowering of grasses. ZmLG1 controlled the ligule development in maize (Moreno et al.1997). SPL13 regulates tomato inflorescence structure and side branch yield (Cui et al.2020). OsSPL18 regulates grain weight and number in rice (Yuan et al.2019). In wheat, yield-related traits are controlled by TASPL20 and TASPL21 genes (Zhang et al.2017). The SBP-box gene family is also involved in the stress response process of plants. Such as, SBP transcription factor CRR1 plays an important role in maintaining the copper homeostasis in its body in *Chlamydomonas reinhardtii* (Kropat et al.2005). Therefore, SBP protein plays an important role in flower formation and development, leaf morphogenesis, fruit development, environmental signal response (abiotic stress), and signal transduction. The SBP gene in cabbage and chrysanthemum responds to hormone treatment and abiotic stress (Tan et al. 2015; Song et al. 2016), which indicates that the SBP gene is an important gene related to drought resistance and stress tolerance.

Quinoa (*Chenopodium willd*) is an annual dicotyledon of the family Amaranthaceae. Quinoa is cold-tolerant, drought-tolerant, salt, alkali, and barren-tolerant, with high protein content in the grain, coordinated amino acid ratio, and rich in vitamin (A, B2, E) and minerals(Ga, Fe, Cu, Mg, Zn). It has the title of mother of grain, golden grain, and sacred food, which has attracted the attention of domestic and foreign agricultural and food experts and consumers. Quinoa genome sequencing was completed in 2017 (Jarvis et al.2017). This makes it possible to reveal the functions of important gene families in Quinoa at the whole genome level. The SBP-box gene plays an important role in plant type, yield, stress resistance, etc. Therefore, the isolation and identification of important genes in the SBP-box family can provide candidate genes for crop genetic improvement. This study used bioinformatics methods to identify the quinoa SBP-box gene family and analyzed the family members' sequence characteristics, chromosome location distribution, and gene structure. The RNA-seq data in public databases were used to study the expression patterns of the family members in different tissues and environments. At the same time, the expression patterns of the family genes under abiotic stress were analyzed by qRT-PCR. This provides important information for the in-depth study of the quinoa SBP-box gene family and the cloning of important genes.

2. Materials And Methods

2.1. Search and Identification of SBP Members in Quinoa

Downloaded the complete genome sequence, CDS sequence, protein sequence, and gene annotation files from the Phytozome V12.1 database (Goodstein et al.2012) (https: Phytozome.jgi.doe.gov pzportal. HTML). We downloaded the file of the Hidden Markov Model sequence spectrum (PF03110) of the SBP protein family and constructed Hidden Markov Model (HMM) with Hmmer (V3.1B2, http: HMMER.org), searched for candidate sequences with SBP-box domain (E-value set to 1) in Quinoa protein database. The candidate sequences were passed through the Pfam (http://pfam.xfam.org/family), NCBI-CDD (https://www.ncbi.nlm.nih.gov/cdd/) and SMART (http://smart.embl-heidelberg.de/). The SBP-box gene of Quinoa was obtained by removing the incomplete reading frame by hand and checking whether it had an SBP domain.

2.2. Basic physical and chemical properties of protein and phylogenetic tree analysis

The isoelectric point and relative molecular mass of all SBP amino acid sequences were analyzed by ExPASy (https://web.expasy.org/protparam/), and the subcellular localization was performed by Psort-Prediction (http://psort1.hgc.jp/form.html).

Known amino acid sequences of Arabidopsis and tomato SBP were downloaded from PlantTFDB v5.0 (http://planttfdb.gao-lab.org/tf.php?sp=Ppe&did=Prupe.I004500.1.p). The phylogenetic analysis of Arabidopsis Thaliana, tomato, and Quinoa SBP protein was performed using ClustalX (V2.1) (Thompson et al.2002). The

phylogenetic tree was constructed by the adjacency method, and the results were compared using Mega (V6.0) for phylogenetic analysis (Tamura et al.2013), and the test parameter bootstrapping was repeated 1000 times, with other parameters default.

2.3. Gene structure and conservative motif analysis

According to the quinoa genome annotation file, the gene structure display system GSDS (http://gsds.cbi.pku.e-du.cn/) was used to construct the gene structure of SRS gene exons/introns (Clouse.1996). Used MEME (http://meme-suite.org/) to predict and analyze the amino acid conserved domains of the quinoa SBP transcription factor protein sequence. The upper limit of the number of conserved domains obtained by the search was 10, and other parameters were defaulted (Suyama et al.2006).

2.4. Chromosome location and gene duplication analysis

Used the annotation information of the SRS gene in the quinoa database to determine the position of the family member on the chromosome and used the MapInspect tool (http://www.plantbreeding.wur.nl/uk/software_mapinspect.html) to mark each SBP gene position on the chromosome, to get the distribution of each SBP-box gene in the genome. MCScanx was used for gene family replication analysis (Wang et al.2012). The conditions for determining gene duplication events were based on the identification method of the Plant Genome Duplication Database (Lee et al.2013); that is, gene duplication events must meet the following conditions at the same time (Zhou et al.2004): (1) The length of the matching part of the two gene sequences is greater than the length of the longer sequence. (2) The similarity of the matching parts of the two gene sequences is greater than 80%; (3) The closely connected genes only participate in one replication event. At the same time, combined with the gene's position on the chromosome to determine whether a tandem copy or a fragmented copy has occurred.

2.5. Analysis of cis-acting elements and construction of protein interaction network diagram

According to the Quinoa annotation information file, 2000bp upstream of the SBP gene transcription initiation site was extracted by the sequence extraction function of TBtools (Chen et al.2020) tool, which was used as the promoter region, and the SBP promoter region was analyzed by the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al.2002). String(http://STRING db.org/) was used to construct the protein-protein interaction network diagram (Szklarczyk et al.2015), and the SBP network was constructed using the STRING software (confidence greater than 0.8) based on *Arabidopsis thaliana*.

2.6. Secondary structure analysis and tertiary modeling prediction

The secondary structure of the family proteins was analyzed on NPS@: GOR4: (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html) (Combet et al.2000). At the same time, we used the Swiss Model (https://swissmodel.expasy.org/) (Kelley et al.2015) to predict the tertiary structure of proteins.

2.7. Plant material and treatments

Quinoa "Honghua Dajingyuan (HHDJY)" was used as the material which were grown in seed germination pouches(height:width = 30 cm:25.5 cm) .Uniform seeds of Quinoa with the same degree of fullness were subjected to surface sterilization and germinated in distilled water, with a germination box (10 × 8 × 6 cm) used as a bed for the seeds.After 2–3 d, seedlings were transferred to the seed germination pouches and placed in the trough of the paper wick and in a growth room (relative humidity 60–70%, light time 12h, day and night temperature 28℃/18℃) .Before transferring seedlings to the pouch,100 ml of MS culture medium was added to the pouch,When the seedlings grew for one months, transferring seedlings to the pouch again,100 ml of distilled water or 200 mmol/L Nacl solution was added

to the pouch, with distilled water treatment as a control. The leaves were collected at 0, 2, 4, 8, and 12 hours after treatment. Three biological replicates were performed at each time point. The collected leaves were temporarily stored in liquid nitrogen and then stored in a -80℃ refrigerator for further analysis.

2.8. SBP gene expression analysis, RNA extraction, and real-time quantitative PCR

The quinoa SBP gene expression data was obtained from the transcriptome data of different tissues of Quinoa (number: PRJNA394651) and drought, high temperature, salt, and low phosphorus stress (number: PRJNA306026). The data was standardized by the Log2 method.

Total RNA was extracted with RNA plant Plus reagent (Beijing, China), and cDNA was prepared with Superscript TM III reverse transcriptase kit (Invitrogen). RNA and cDNA were detected by ultra-micro ultraviolet spectrophotometer (UV-Vis Spectrophotometer Q5000). qRT-PCR analysis was carried out in the abi-viia 7 real-time PCR system of American applied biosystems using 2× quantitect-sybr-green-pcr-mix(qiagen) (Qiagen). The procedure was as follows: Denaturation at 95°C for 3 min, denatured for 10 s at 95°C for 40 cycles, then annealed/extended at 60°C for 1.

3. Results

3.1. Analysis of basic physical and chemical properties

23 SBP genes were identified in Quinoa. Analysis of basic properties showed that the size of the amino acids encoded by this family is between 189–1106, and the molecular weight is between 21298.92-121757.34 Da (Table 1). The theoretical isoelectric point variation was relatively large, ranging from 5.74 to 10.24, the instability index is between 43.72 with 81.14, and both are greater than 40, which are unstable proteins. The aliphatic index was between 56.30 with 87.87. GRAVY (grand average of hydropathicity) had differences, but they were all less than zero. Therefore, all proteins in this family belong to hydrophilic proteins. Subcellular location showed that 23 CqSBPs proteins are located in the nucleus. The secondary structure prediction showed that Alpha Helix (HH) Extended Strand (Ee) and Random coil (Cc) mainly constituted the secondary structure of CqSBP, and Random coil accounted for the largest proportion of all proteins (all more than 45.54%).

Table 1
Information of the 23 SBP gene members in quinoa

Gene accession No	Gene	Size (aa)	Molecular weight (Da)	Isoelectric point	Instability index	Aliphatic index	GRAVY	Subcellular Localization
AUR62004146-RA	CqSBP01	189	21298.92	10.24	81.14	56.83	-0.803	nucleus
AUR62013707-RA	CqSBP02	189	21300.88	10.24	81.14	56.30	-0.796	nucleus
AUR62028919-RA	CqSBP03	462	50273.96	8.90	60.23	58.53	-0.643	nucleus
AUR62005629-RA	CqSBP04	462	50311.10	8.67	56.91	61.65	-0.580	nucleus
AUR62002563-RA	CqSBP05	978	108454.93	6.44	49.36	84.30	-0.358	nucleus
AUR62029983-RA	CqSBP06	1189	132135.72	7.01	48.63	87.87	-0.305	nucleus
AUR62029984-RA	CqSBP07	980	108865.55	6.30	46.74	84.13	-0.306	nucleus
AUR62024322-RA	CqSBP08	349	37731.32	7.64	61.14	53.04	-0.680	nucleus
AUR62019452-RA	CqSBP09	202	22122.73	9.81	68.39	58.86	-0.864	nucleus
AUR62012061-RA	CqSBP10	373	39563.29	8.45	59.28	51.98	-0.654	nucleus
AUR62032118-RA	CqSBP11	562	62082.24	6.87	50.08	71.25	-0.468	nucleus
AUR62039662-RA	CqSBP12	545	60397.78	6.26	55.10	64.88	-0.645	nucleus
AUR62011728-RA	CqSBP13	984	110161.02	5.76	50.96	78.58	-0.461	nucleus
AUR62029416-RA	CqSBP14	347	38017.36	8.76	62.86	58.70	-0.633	nucleus
AUR62003425-RA	CqSBP15	349	38422.77	8.74	65.21	58.11	-0.668	nucleus
AUR62035190-RA	CqSBP16	1106	121757.34	7.06	57.61	70.34	-0.510	nucleus
AUR62042534-RA	CqSBP17	922	101713.51	6.75	59.37	67.16	-0.541	nucleus
AUR62005645-RA	CqSBP18	359	38997.57	9.36	46.94	59.22	-0.574	nucleus
AUR62028905-RA	CqSBP19	317	34782.05	9.05	43.72	62.74	-0.653	nucleus

Note: GRAVY represents Grand average of hydropathicity.

Gene accession No	Gene	Size (aa)	Molecular weight (Da)	Isoelectric point	Instability index	Aliphatic index	GRAVY	Subcellular Localization
AUR62003075-RA	CqSBP20	499	55390.33	5.89	50.89	74.43	-0.448	nucleus
AUR62007890-RA	CqSBP21	423	46751.15	6.57	51.37	66.88	-0.671	nucleus
AUR62042853-RA	CqSBP22	716	80133.15	6.32	51.17	74.85	-0.378	nucleus
AUR62042654-RA	CqSBP23	707	79251.95	5.74	48.34	75.12	-0.404	nucleus
Note: GRAVY represents Grand average of hydropathicity.								

3.2. Phylogenetic tree analysis

To understand the phylogenetic relationship and classification of CqSBP genes. We used 71 protein sequences from Arabidopsis (31), tomato (17), and Quinoa (23) to construct a phylogenetic tree based on multiple sequence alignments (Fig. 1, Table S1). The SBP gene was divided into 10 subgroups (Group1-Group10). The SBP protein in 10 subgroups ranged from 2 to 16, and the largest number of CqSBP genes was found in Group10. Compared with other subfamilies, Group8 protein contained a relatively long amino acid sequence, indicating that there may be some functional differences between this subfamily and other subfamilies. In CqSBP genes, we found 8 pairs of paralogous genes (CqSBP01/CqSBP02, CqSBP06/CqSBP07, CqSBP08/ CqSBP10, CqSBP14/ CqSBP15, CqSBP16/ CqSBP17, CqSBP18/ CqSBP19, CqSBP20/ CqSBP21, and CqSBP22/ CqSBP23). Paralogous genes showed high sequence similarity between two genes. This similarity indicates that they may have originated from a single gene in a distant common ancestor.

3.3. Gene structure and conserved domain analysis

Intron-exon structure, intron type, and number are typical evolutionary imprints of a gene family. Further exploring the gene structure of the 23 sequences of the quinoa SBP gene family, the gene structure of the same branch of the phylogenetic tree was relatively similar, but the different gene structures of different branches (Fig. 2). The number of introns in the quinoa SBP genes family is between 2 with 14. According to statistics, there are 8 CqSBP genes contain 2 introns (34.78%), 5 CqSBP genes contain 3 introns (21.74%), 2 CqSBP genes contain 4 introns (8.7%), 8 CqSBP genes contain 9 or more introns (34.78%). The MEME analysis tool was used to predict all quinoa SBP protein sequences, and the results showed that the 10 conservative functional motifs are statistically significant, and the e-value of each conservative motif is less than $1e^{-300}$. The predictions of the conservative motifs of quinoa SBP genes mostly supported the classification of SBP gene family phylogenetic analysis. The sequence characteristics and amino acid length of these conserved motifs are shown in Fig. 3. Motif 3 and Motif 4 are conserved SBP domains in all quinoa proteins. 15 SBP genes contain only three conserved motifs, Motif 1, Motif 3, and Motif 4, corresponding to the simple gene structure. CqSBP22 and CqSBP23 contain motif 2, Motif 3, Motif 4, Motif 6, and Motif 9. Some genes (CqSBP05, CqSBP06, CqSBP07, CqSBP013, CqSBP16, and CqSBP17) had 10 conservative motifs.

An unrooted phylogenetic tree was constructed based on the full-length sequences of CqSBP proteins using the N-J method in MEGA7. Bootstrap values based on 1,000 replications were calculated. (A) The distribution of motif in SRS proteins. (B) The exon-intron structure of the SRS gene.

2.4. Chromosome location and gene duplication analysis

23 CqSBP genes were located on 12 chromosomes of Quinoa (Fig. 3) by chromosome mapping (a total of 18 chromosomes of Quinoa). The most distributed SBP genes (4) on chromosome 11, and other genes distributed on the remaining chromosomes were between 1 and 3. A Gene family is a group of genes derived from the same ancestor, composed of a gene through gene duplication; usually, they have obvious similarities in structure and function and code similar protein products. Genes in the same family can be arranged closely together to form a gene cluster, but they are often scattered in different positions on the same chromosome or exist on different chromosomes (1). Gene duplication has a great effect on gene families. Gene duplication provides raw materials for new genes and promotes the production of new functions (5). Gene duplication mainly includes genome fragment duplication, tandem duplication, and rearrangement at the gene and chromosome level. Tandem duplication and fragment duplication often occur in the evolution and expansion of gene families (7). Tandem duplication usually causes gene clusters, and fragmented duplication may cause the dispersion of family members.

At the same time, we detected the replication event of the CqSBP-box gene family. Using information from chromosome fragments and genomic repeats, 12 pairs of homologous (Table 2) genes with fragment repeats and tandem repeats were found, and their sequence homology was high. Eleven pairs of repetitive genes belong to fragment duplication (91.67% of all replicated gene pairs). The pair (CqSBP06/CqSBP07) belongs to tandem repeats, and the gene pairs of tandem repeats have high sequence similarity, often higher than 60%. The gene pairs with repeated fragments were distributed in each subfamily of the phylogenetic tree, and their distribution had a certain preference. Ten pairs of them were distributed in the same phylogenetic subfamily, which may be caused by the polyploidization process of the quinoa genome. At the same time, we also found that there were 8 pairs of paralogous genes in the CqSBP gene. Further calculating the Ka/Ks between genes with homology, we found that all Ka/Ks are less than 1; it shows that these genes have undergone purification selection after duplication, and the gene function is relatively conservative.

Table 2
Gene duplication in CqSBP family

Duplicated CAMTA gene1	Duplicated CAMTA gene2	Ka	Ks	Ka/Ks	Date(mya)T = Ks/2λ	Selective pressure	Duplicate type
CqSBP01	CqSBP02	0.006	0.123	0.048	7.240	Purifying selection	Segmental
CqSBP03	CqSBP04	0.017	0.077	0.225	4.536	Purifying selection	Segmental
CqSBP05	CqSBP06	0.040	0.103	0.392	6.066	Purifying selection	Segmental
CqSBP05	CqSBP07	0.040	0.098	0.410	5.776	Purifying selection	Segmental
CqSBP06	CqSBP07	0.032	0.081	0.400	4.767	Purifying selection	Segmental
CqSBP08	CqSBP10	0.013	0.077	0.164	4.551	Purifying selection	Segmental
CqSBP11	CqSBP12	0.028	0.062	0.445	3.664	Purifying selection	Segmental
CqSBP14	CqSBP15	0.014	0.061	0.230	3.577	Purifying selection	Segmental
CqSBP16	CqSBP17	0.022	0.094	0.228	5.563	Purifying selection	Segmental
CqSBP18	CqSBP19	0.072	0.135	0.533	7.945	Purifying selection	Segmental
CqSBP20	CqSBP21	0.018	0.072	0.257	4.237	Purifying selection	Segmental
CqSBP22	CqSBP23	0.033	0.114	0.286	6.738	Purifying selection	Segmental

Note: The non-synonymous (Ka) and synonymous substitution rate (Ks); millions of years ago (mya

2.5. Analysis of Homeopathic Elements and Construction of Protein Interaction Network Diagram

The promoter sequence 2000bp upstream of the start codon of the CqSBP gene was obtained through TBtools and analyzed by Plantcare. A total of 49 functional elements were found (Fig. 4, Table S2) related to tissue-specific expression, light response, hormones, and stress. There were 22 light response elements: ACE, AE-Box, Box-4, G-Box, I-Box, MRE, SP1, and LAMP. Some elements only exist in specific genes (RY-element only exists in CqSBP13, AAAC-motif only in CqSBP05, and Gap-box only in CqSBP08), G-box, and G-box GT1-motif were found in almost all CqSBP genes. There were 10 types of action elements that respond to hormones. Among them, 10 elements were involved in the methyl jasmonate reaction, and 9 elements were involved in the gibberellin reaction (except SARE). There were 8 types of elements involved in the abscisic acid reaction (except SARE and AuxRR-core). At the same time, a special element SARE (salicylic acid response element), was discovered, which only exists in CqSBP22. In addition, some elements involved in stress have been discovered, including low-temperature response elements (LTR, DRE), drought response elements MBS, and WUN-motif related to wound response. There are also the necessary action elements ARE involved in participating in anaerobic induction. Some functional elements rarely occur the element CAT-box, which is related to

meristem expression, the GCN4-motif involved in endosperm expression, and the HD-Zip1, which is related to the differentiation of palisade mesophyll cells, and zeol. O_2-site is an element related to the regulation of protein metabolism.

All promoter sequences (2000bp) were analyzed. In order to get a closer understanding of the functions of this family of proteins, *Arabidopsis thaliana* was used as a template to construct a protein interaction network diagram (Fig. 5). We found that 22 CqSBP proteins (except CqSBP17) appeared in the known Arabidopsis interaction network diagram. Among them, the functions of Arabidopsis genes (AT1G69170 and AT1G76580) homologous to the 7 CqSBP genes of Quinoa are unknown, and the functions of other Arabidopsis genes that are homologous to Quinoa have been studied. It can be seen from the figure below that AtSPL7 is at the center of the interaction network, and it works by interacting with multiple proteins. Studies have shown that AtSPL7 directly upregulates SEPALLATA3 (SEP3) and MADS32 genes to induce phase transition and flowering of gramineous forages (Gou et al.2019). In addition, this gene can regulate copper homeostasis-related genes in *Arabidopsis thaliana* (Araki et al.2018). Therefore, the homologous CqSBP genes (CqSBP22 and CqSBP23) may have similar functions. AtSPL9 participates in the transition of nutrition to the reproductive stage and inhibits the germination of new leaves in the shoot apex meristem. Therefore, the CqSBP genes (CqSBP8 and CqSBP10) may also have similar functions. The overexpression of ATSPL1 (CqSBP05, CqSBP06, CqSBP07, and CqSBP13) enhanced the heat tolerance of *Arabidopsis Thaliana* and tobacco (Chao et al.2017). ATSPL13B (CqSBP14, CqSBP15, CqSBP18, and CqSBP19) is involved in the vegetative growth and reproductive development of *Arabidopsis thaliana*. At the same time, AtSNZ can interact with SPL1/SPL3/SPL13B/SPL9/AGL8/AT1G69170 and so on, so CqSBP genes which are homologous with these genes may be regulated by CqSNZ genes.

2.6. Analysis of expression patterns

The analysis of quinoa gene transcriptome data obtained gene expression profiles of 23 quinoa CqSBP genes in 13 quinoa tissues and 10 different treatments. Use TBtools software to draw a gene expression map (Fig. 6, Table S3). In the figure, red and blue indicate the intensity of gene expression, red indicates strong signal strength and blue indicates weak signal strength. The results showed that most of the genes in the SBP family have strong expression under different treatments (CqSBP03/CqSBP04, CqSBP05/CqSBP06, CqSBP07 CqSBP08/CqSBP10, CqSBP13, CqSBP16/CqSBP17, CqSBP22/CqSBP23). The expression of some SBP genes was very low under different treatments (CqSBP01/CqSBP02, CqSBP14/CqSBP15, CqSBP18/CqSBP19, and CqSBP09).

There are differences in expressing different SBP genes in different organs and periods of Quinoa. Further analysis found that the expression of SBP genes has certain tissue specificity. Gene expression in the same branch of the evolutionary tree is similar. For example, CqSBP11/CqSBP12, CqSBP14/CqSBP15, and CqSBP20/CqSBP21 genes are expressed very low in different tissues of Quinoa; CqSBP13 is expressed very high in various tissues. CqSBP01/CqSBP02 genes were highly expressed in Apical meristems, stems, Flowers, immaturity seeds, and Leaf petioles of Quinoa. Also highly expressed in Apical meristems were CqSBP09, CqSBP13, CqSBP08/CqSBP10, and CqSBP16/SBP18. (A) CqSBP expression patterns at different treatments. (B) CqSBP expression patterns at different developmental stages and tissues. FPKM calculated gene expression. We standardized the data using the Log2 method.

2.7. Q-RTPCR analysis

To study whether the CqSBP gene family has a certain expression pattern under salt stress. We used qRT-PCR to detect the changes in the expression of each gene in leaves under salt stress (Fig. 7, Table S4-S5). The results showed that most of the repeated gene pairs have similar expression patterns, such as CqSBP01/CqSBP02, CqSBP05/CqSBP06,

CqSBP16/CqSBP17, CqSBP18/CqSBP19, and the expression patterns of individual repeated genes are also quite different (CqSBP08/CqSBP10, CqSBP11/CqSBP12, CqSBP14/CqSBP15). Most genes respond significantly to salt stress. The expression levels of CqSBP01, CqSBP02, CqSBP12, CqSBP13, and CqSBP16 genes under salt stress increased by 10 times or even 100 times. Among them, the expression level of CqSBP13 under salt stress was 78 times that of the control. This gene plays an extremely important role in salt stress, but its function still needs further research. Most of the genes (CqSBP03, CqSBP04, CqSBP05, CqSBP06, CqSBP07, CqSBP08, CqSBP09, CqSBP10, CqSBP14, CqSBP17, CqSBP20, CqSBP21, CqSBP22, and CqSBP23) expressed more frequently after stress treatment, at the same time, we found that some genes also showed down-regulation under salt stress (CqSBP11, CqSBP15, CqSBP18, and CqSBP19), which suggested that these genes might play a negative role in salt stress. These results suggest that the CqSBP gene in Quinoa may participate in salt-stress adaptation through complex mechanisms.

Values represented the mean ± standard error of the mean (SEM) of three biological replicates with three technical replicates at different treatments. Error bars indicated the SEM among the three experiments.

4. Discussion

The SBP-box protein family is a type of plant-specific transcription factor which regulates the expression of downstream genes by combining with cis-acting elements in the promoter region of downstream genes. The number of members of this family varies greatly among different species, and 49 SBP genes have been identified in soybean (Schmutz et al.2010). This study identified 23 quinoa SBP genes. This number variation makes the function of SBP transcription factors more diversified. More SBP genes are in the soybean genome, indicating that the soybean SBP gene family has experienced more complex amplification, loss, and evolutionary processes. The 23 CqSBP genes are distributed on 12 chromosomes. The number of amino acids, isoelectric points, and molecular weights of CqSBP proteins is quite different. This may be due to the family members' different functions during growth and development, amino acid composition, and protein structure. The difference may lead to differences in the function of SBP gene family members. Interestingly, we found that all CqSBP genes are located in the nucleus, indicating that CqSBP can be a transcription factor in the nucleus.

Gene duplication is of great significance to the evolution of gene families, mainly because gene duplication can provide the most primitive material for generating new genes, and the generation of new genes promotes new functions (Rensing., 2014). There are three main ways of plant gene replication: fragment replication, tandem replication, transposition events such as retrotransposition, and repeated transcription. Fragment duplication is the most important way because most plants undergo a chromosome doubling process and retain many repeated chromosome fragments in the genome (Wang et al. 2012). This study identified that 11 pairs of homologous genes were generated by fragment duplication, and 1 pair of homologous genes was amplified by tandem duplication, indicating that the amplification of the quinoa gene family was mainly amplified by fragment duplication. SPL10/SPL11 in Arabidopsis has the same gene structure, the sequence is highly similar (82.1%), and they are adjacent to each other on the chromosome. The tandem repeat mechanism may produce them. The corresponding SBP-box in Quinoa has tandem replication. The family members CqSBP06/CqSBP07 also have high homology. Quinoa is a heterotetraploid plant, and its genome has undergone a process of doubling during evolution, so many genes exist in multiple copies in the quinoa genome. Recent studies (Jarvis et al.2017) have shown that a whole-genome duplication event occurred in Quinoa between 3.3–6.3 Mya (million years ago), supporting the idea that genes exist in multiple copies. This study found that multiple copies of SBP genes are scattered on different chromosomes. Microcollinearity analysis showed that there are 8 pairs of SBP paralogous genes in the quinoa genome, which provides favorable evidence for the doubling of the quinoa genome.

From a phylogenetic point of view, SBP genes are composed of multiple members in most plants. For example, 3, 4, and 8 SBP homologous gene pairs were identified in Arabidopsis, rice, and soybean genomes. They belong to the horizontal

homologous genes, are in the same evolutionary branch, and are formed after the occurrence of species. Compared with Arabidopsis and rice, Quinoa has more SBP homologous gene pairs, which fully indicates that more repetitive events of SBP genes occur after quinoa speciation. Generally speaking, SBP genes with complete SBP functional domains can often find EST sequences, which means that these SBP genes have transcriptional activity. The 23 CqSBP genes in this study all included typical SBP functional domains.

The C-terminus of the conserved domain of the SBP transcription factor is the nuclear localization signal region. When analyzing the conserved domains of SBP, this study found that the domain of quinoa SBP protein contains about 79 amino acid residues and has two zinc finger structures, which are C3H (C-C-C-H) C2HC (C-C-H-C) types. In addition, most transcription factors have an NLS site at the C-terminus of the SBP conserved domain. In addition, CqSBP genes located in the same branch share similar intron/exon structures. And most CqSBPs of the same branch have similar motifs. Therefore, genes in the same phylogenetic group may have similar roles in Quinoa. This indicates that the evolution of the SBP-box gene family may be closely related to the diversity of gene structure. In addition to the conserved CqSBP motif, several unique group-specific motifs have also been observed, such as CqSBP05, CqSBP06, CqSBP07, and CqSBP13 subfamily The motifs in 2, 5, 6, 7, 8, 9 and 10. These specific motifs may be important for the specific role of the CqSBP genes, and their functional differentiation may have appeared during the evolution of different lineages.

Gene expression profile analysis showed that the expression patterns of genes in this family are different in different tissues. These results may be an important research resource for further revealing the function of CqSBP genes in quinoa development. Most SBP genes are widely expressed in the meristems, flowers, inflorescences, petioles, internodes, stems, and leaves of Quinoa, which indicates their key role in these biological processes. At the same time, most CqSBP genes are highly expressed in the apical meristem, which indicates that they are widely involved in cell differentiation. Studies have shown that AtSPL8 mediates anther development, flowering, cell differentiation, floral organ development, and stamen filament elongation (Zhang et al. 2007; Unte et al. 2003). BrcSPL8 plays an important role in developing Chinese cabbage flowers (Zhang et al.2017). In this study, CqSBP01 and CqSBP02, homologous to AtSPL8, are highly expressed in inflorescence and apical meristems. It is inferred that they may be involved in distinguishing meristems and flower formation. CqSBP08 and CqSBP10 are highly expressed in stems, and they may participate in stem development and regulation of organ formation, which has been verified in the study of the homologous gene AtSPL9 (Zhang et al.2020).

AtSPL14 plays a role in developing normal plant structures and is sensitive to fumonisin B1 (Stone et al.2005). It is highly expressed in all organizations. OsSPL6 and OsSPL15 are also highly expressed in all tissues. In our study, CqSBP16 (the homologous gene of AtSPL14) is highly expressed in all tissues, and it may affect the plant structure of Quinoa. The qualitative analysis of the role of the quinoa SBP transcription factor in the growth and development of Quinoa is of great significance to the study of the function of the Quinoa SBP gene. It provides a certain theoretical basis for studying the regulation mechanism of cotton fiber (with economic value). Therefore, increasing the research on these SBP genes may help better understand specific physiological processes and subsequent agricultural genetics research. Q-TR-PCR showed that CqSBP01, CqSBP02, and CqSBP13 expressed extremely significantly under salt stress, indicating that these three genes may play a key role in salt stress. Wang (Wang et al.2009) analyzed the regulatory network related to SPL genes and found that 112 genes are closely related to SPL genes, and the promoters of these genes all contain a core motif of GATC, speculating that SBP transcription factors may be involved in plant tissues development, biotic and abiotic stress response, and activation of other transcription factors and membrane proteins. At the same time, the SPL gene may also be involved in the metabolism of glucose, inorganic salts, ATP, and the transport of carbohydrates. The SBP gene was discovered in the gene regulation network that studies the path of flower formation, so it is considered to be closely related to flower development. In recent years, it has been discovered that the

SBP gene has a wide range of biological functions. The SBP gene has been found in many plants, but its function is still poorly understood. One of the reasons is that almost all functional studies are carried out through mutants, for example, by constructing over-expression or silencing vectors of related genes, then using transgenic technology to transfer them into corresponding plants, and finally by observing the phenotype of the transgenic plants. Explain the function of genes.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (32060482), China Agricultural University- Tibet Agriculture and animal Husbandry University Joint Project (604419037;

766660010)

Author contributions

Xibo Feng:Designed the research, Funding acquisition and Project administration. Weihai Hou:conceived and designed the research,conducted experiments,as well as wrote the manuscript. Jianling Wang:prepared seed materials. Zongyu Zhang :Data analysis for the research. Inzamam Ul Haq: Help with experiments.

References

1. Araki R, Mermod M, Yamasaki H, Kamiya T, Fujiwara T, Shikanai T (2018) SPL7 locally regulates copper-homeostasis-related genes in Arabidopsis. J Plant Physiol 224:137–143
2. Birkenbihl RP, Jach G, Saedler H, Huijser P (2005) Functional dissection of the plant-specific sbp-domain: overlap of the DNA-binding and nuclear localization domains. J Mol Biol 352:585–596
3. Chao LM, Liu YQ, Chen DY, Xue XY, Mao YB, Chen XY, Chen XY (2017))Arabidopsis Transcription Factors SPL1 and SPL12 Confer Plant Thermotolerance at Reproductive Stage. Mol Plant 10:735–748
4. Chen CJ, Xia R, Chen H, He YH (2020) TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. Mol Plant 13:1194–1202
5. Clouse SD (1996) Molecular genetic studies confirm the role of brassinosteroids in plant growth and development. Plant J 10:1–8
6. Combet C, Blanchet C, Geourjon C, Deléage G (2000) NPS@: network protein sequence analysis. Trends Biochem Sci 25:147–150
7. Cui L, Zheng FY,Wang JF, Zhang CL, Xiao FM, Ye J, Li CX, Ye ZB,Zhang JH (2020) miR156a-targeted SBP-Box transcription factor SlSPL13 regulates inflorescence morphogenesis by directly activating SFT in tomato. J Genet Genomics 18:1670–1682
8. Goodstein DM, Shu SQ, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186
9. Gou JQ, Tang CR, Chen NC, Wang H, Debnath S, Sun L, Flanagan A, Tang YH, Jiang QZ, Allen RD, Wang ZY (2019) SPL7 and SPL8 represent a novel flowering regulation mechanism in switchgrass. New Phytol 222:1610–1623
10. Guo AY, Zhu QH, Gu XC, Ge S, Yang J, Luo JC (2008) Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family. 418:1–8

11. Jarvis D, Ho Y, Lightfoot D, Schmöckel S, Li B, Borm T, Ohyanagi H, Mineta K, Michell C, Saber N, Kharbatia N, Rupper R, Sharp A, Dally N, Boughton B, Woo Y, Gao G, Schijlen E, Guo X, Momin A, Negrão S, Al-Babili S, Gehring C, Roessner U, Jung C, Murphy K, Arold S, Gojobori T, Linden C, Loo E, Jellen E, Maughan P, Tester M (2017) The genome of Chenopodium quinoa. 542:307–3127641
12. Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, Ohyanagi H, Mineta K, Michell CT, Saber N, Kharbatia NM, Rupper RR (2017) Correction: Corrigendum: The genome of Chenopodium quinoa. Nature: Int Wkly J Sci 545:510–510
13. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845
14. Klein J, Saedler H, Huijser P (1996) A new family of dna binding proteins includes putative transcriptional regulators of the Antirrhinum majus floral meristem identity gene SQUAMOSA. Mol Gen Genet 250:7–16
15. Kropat J, Tottey S, Birkenbihl RP, Depege N, Huijser P, Merchant S (2005) 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 102:18730–18735
16. Lee TH, Tang HB, Wang XY, Paterson AH (2013) PGDD: a database of gene and genome duplication in plants. Nucleic Acids Res 41:D1152–D1158
17. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van PY, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30:325–327
18. Moreno MA, Harper LC, Krueger RW, Dellaporta SL, Freeling M (1997) Liguleless1 encodes a nuclear-localized protein required for induction of ligules and auricles during maize leaf organogenesis. Gene Dev 11:616–622
19. Rensing SA (2014) Gene duplication as a driver of plant morphogenetic evolution. Curr Opin Plant Biol 17:43–48
20. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP (2002) Prediction of plant microRNA targets. Cell 110:513–520
21. Riese M, Höhmann S, Saedler H, Münster T, Huijser P (2007) Comparative analysis of the SBP-box gene families in P. patens and seed plants. Gene 401:28–37
22. Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ Chen J L,Xu D.
23. Hellsten U (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183
24. Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D (2005) Specific effects of microRNAs on the plant transcriptome. Dev Cell 8:517–527
25. Shikata M, Koyama T, Mitsuda N, Ohme-Takagi M (2009) Arabidopsis SBP-box genes SPL10, SPL11 and SPL2 control morphological change in association with shoot maturation in the reproductive phase. Plant Cell Physiol 50:2133–2145
26. Song AP, Gao TW, ,Wu D, Xin JJ, Chen SM, Guan ZY, Wang HB, Jin LL, Chen FD (2016) Transcriptome-wide identification and expression analysis of chrysanthemum SBP-like transcription factors. Plant Physiol Biochem 102:10–16
27. Stone JM, Liang XW, Nekl ER, Stiers JJ (2005) Arabidopsis AtSPL14, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1. Plant J 41:744–754
28. Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612
29. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P (2015) STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic

Acids Res 43:D447–D452

30. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6. 0. Mol Biol Evol 30:2725–2729
31. Tan HW, Song XM, Duan WK, Wang Y, Hou XL (2015) Genome-wide analysis of the SBP-box gene family in Chinese cabbage (Brassica rapa subsp. pekinensis). Genome 58:463–477
32. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinf 2:1–22
33. Unte US, Sorensen AM, Pesaresi P, Gandikota M, Leister D, Saedler H, Huijser P (2003) SPL8, an SBP-Box Gene That Affects Pollen Sac Development in Arabidopsis. Plant Cell 15:1009–1019
34. Wang Y, Hu ZL, Yang YX, Chen XQ, Chen GP (2009) Function annotation of an SBP-box gene in Arabidopsis based on analysis of co-expression networks and promoters. Int J Mol Sci 10:116–132
35. Wang YP, Tang HB, Debarry JD, Tan X, Li JP, Wang XY, Lee TH, Jin HZ, Marler B, Guo H, Kissinger JC, Paterson AH (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40:e49–e49
36. Wang YP, Wang XY, Paterson AH (2012) Genome and gene duplications and gene expression divergence: a view from plants. Ann N Y Acad Sci 1256:1–14
37. Xie KB, Wu CQ, Xiong LZ (2006) Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. Plant Physiol 142:280–293
38. Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Nunokawa E, Ishizuka Y, Terada T (2004) A novel zinc-binding motif revealed by solution structures of DNAbinding domains of Arabidopsis SBP-family transcription factors. J Mol Biol 337:49–63
39. Yang ZF, Wang XF, Gu SL, Hu ZQ, Xu H, Xu CW (2007) Comparative study of SBP-box gene family in Arabidopsis and rice. Gene 407:1–11
40. Yuan H, Qin P, Hu L, Zhan SJ, Wang SF, Gao P, Li J, Jin MY,Xu ZY, Gao Q, Du A, Tu B, Chen WL, Ma BT, Wang YP, Li SG (2019) OsSPL18 controls grain weight and grain number in rice. J Genet Genomics 46:41–51
41. Zhang B, Xu WN, Xia L, Mao XG, Li A, Wang JY, Chang XP, Zhang XY, Jing RL (2017) Functional conservation and divergence among homoeologs of taspl20 and taspl21, two SBP-box genes governing yield-related traits in hexaploid wheat. Plant Physiol 174:1177–1191
42. Zhang J, Ping AM, Wang XT, Li GZ, Zhu ZJ, Li ML, Xing GM, Hou LP (2017) Cloning and expression analysis of SPL8 homolog from pak choi (Brassica rapa subsp. chinensis). Biotechnol Biotechnol Equip 31:1132–1138
43. Zhang QQ, Wang JG,Wang LY, Wang JF, Wang Q, Yu P, Bai MY, Fan M (2020) Gibberellin repression of axillary bud formation in Arabidopsis by modulation of DELLA-SPL9 complex activity. J Integr Plant Biol 62:421–432
44. Zhang Y, Schwarz S, Saedler H, Huijser P (2007) SPL8, a local regulator in a subset of gibberellin-mediated developmental processes in Arabidopsis. Plant Mol Biol 63:429–439
45. Zhou T, Wang YX, Chen JQ, Araki H, Jing ZQ, Jiang K, Shen JD, Tian D (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol Genet Genomics 271:402–415

Figures

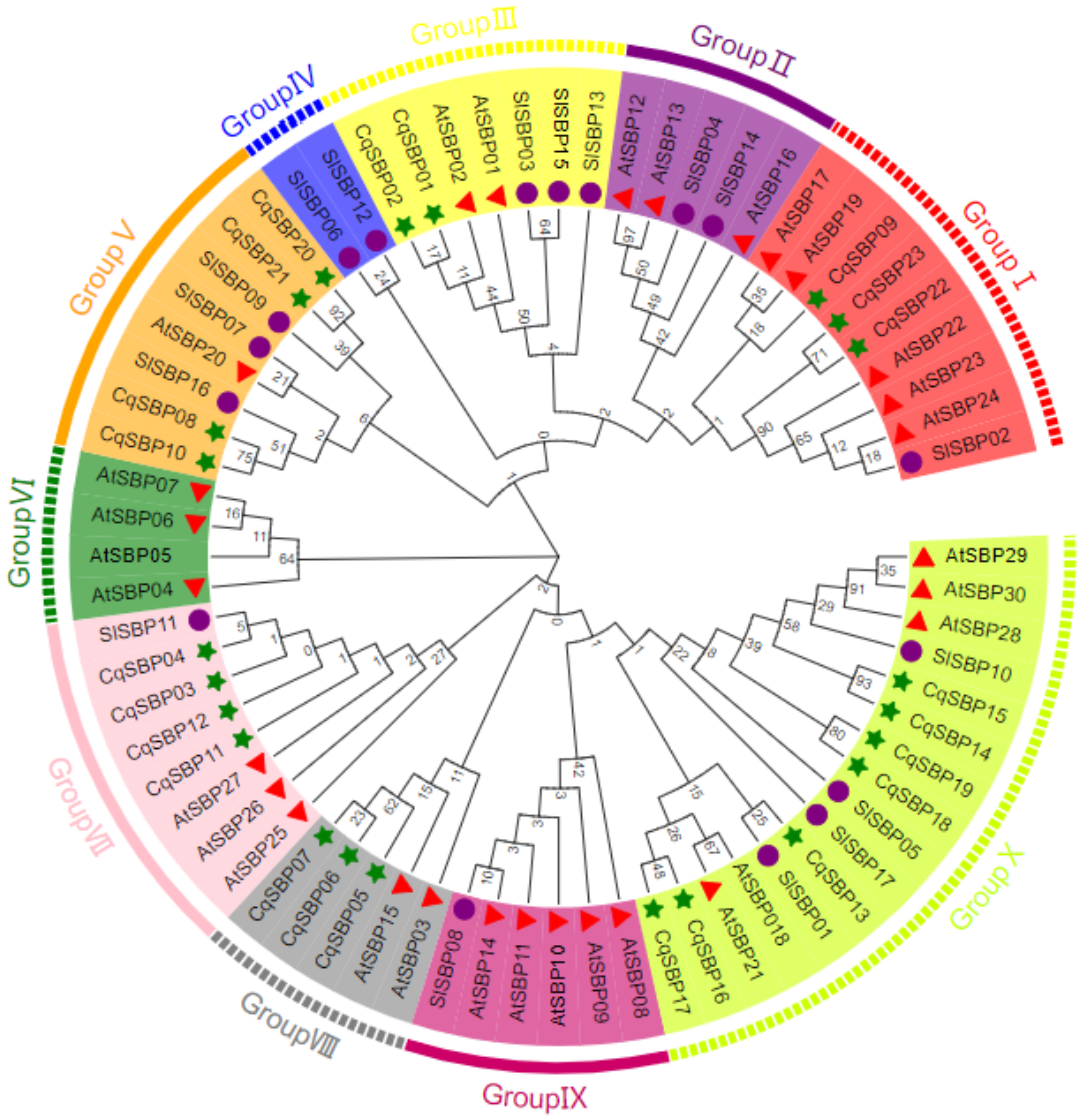

Figure 1

Phylogenetic relationships of the SBP homologs in different species

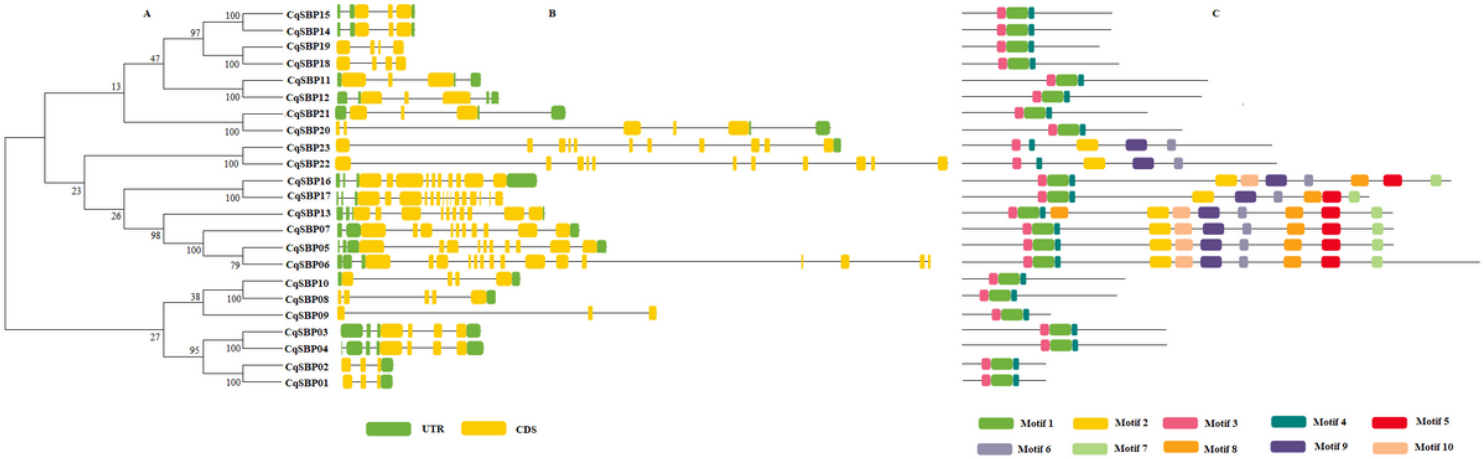

Figure 2

Structural analysis of CqSBP genes in Quinoa.

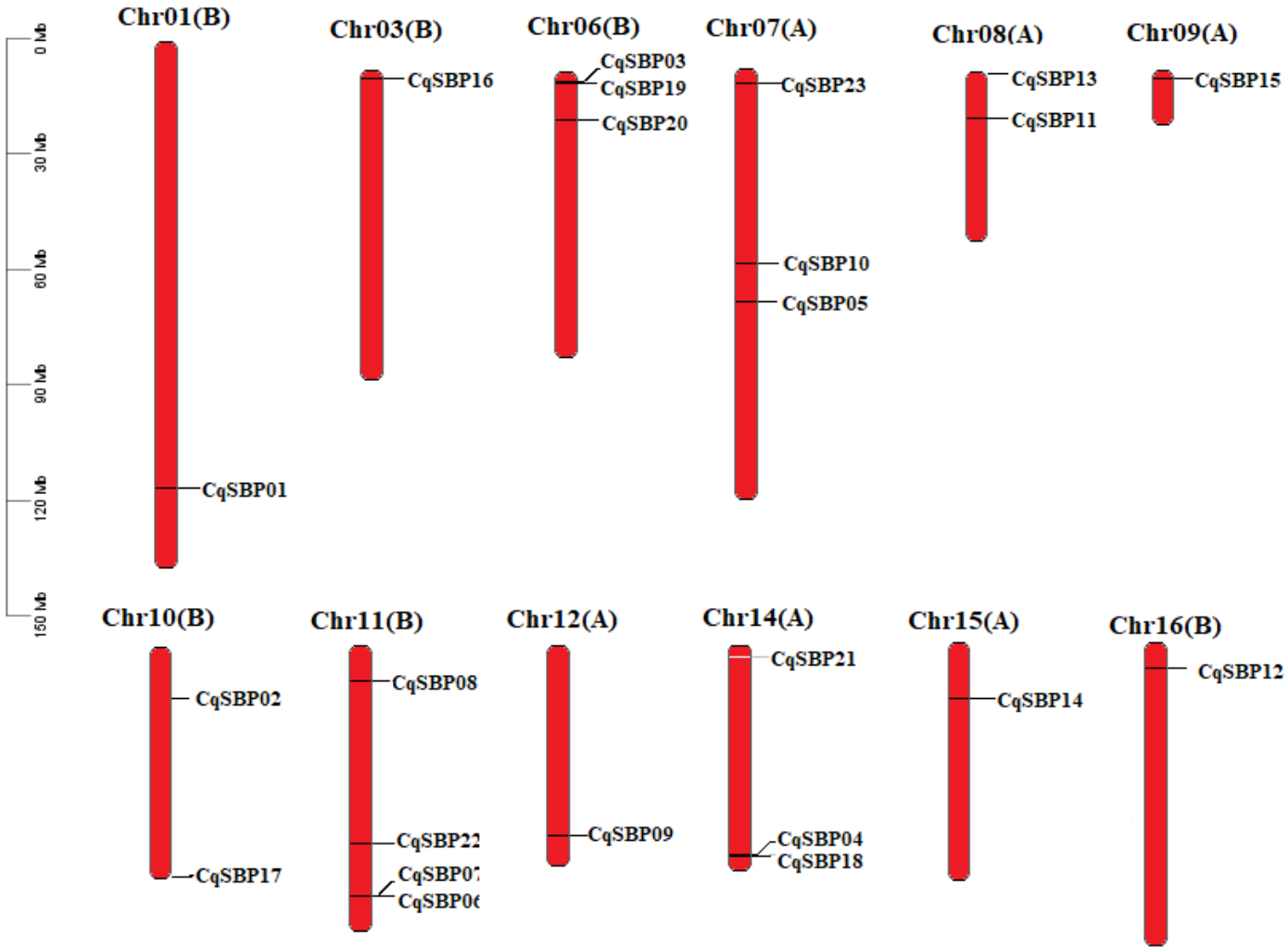

Figure 3

Chromosome mapping of CqSRS genes in Quinoa.

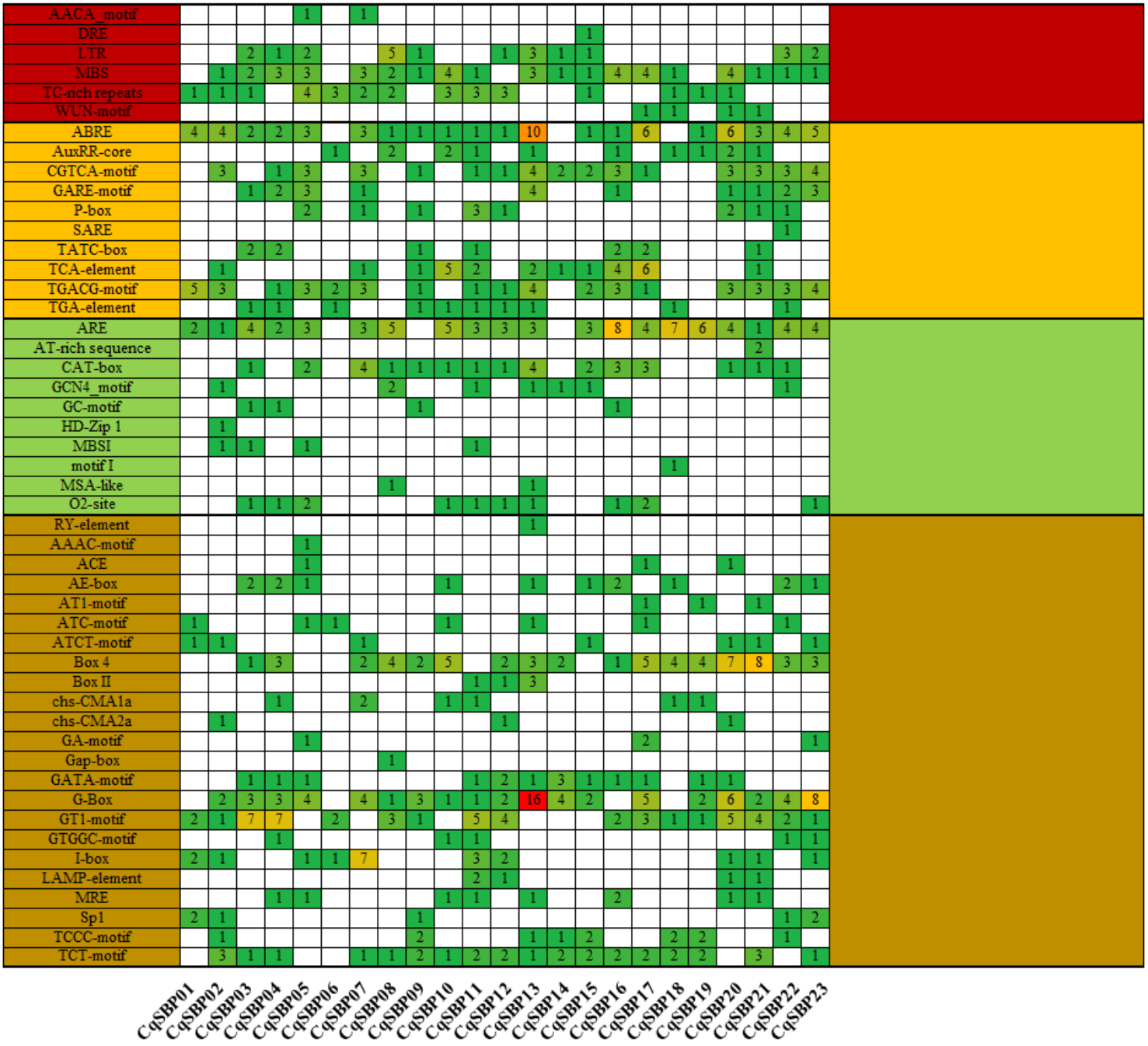

Figure 4

Cis-acting components of Quinoa SBP genes.

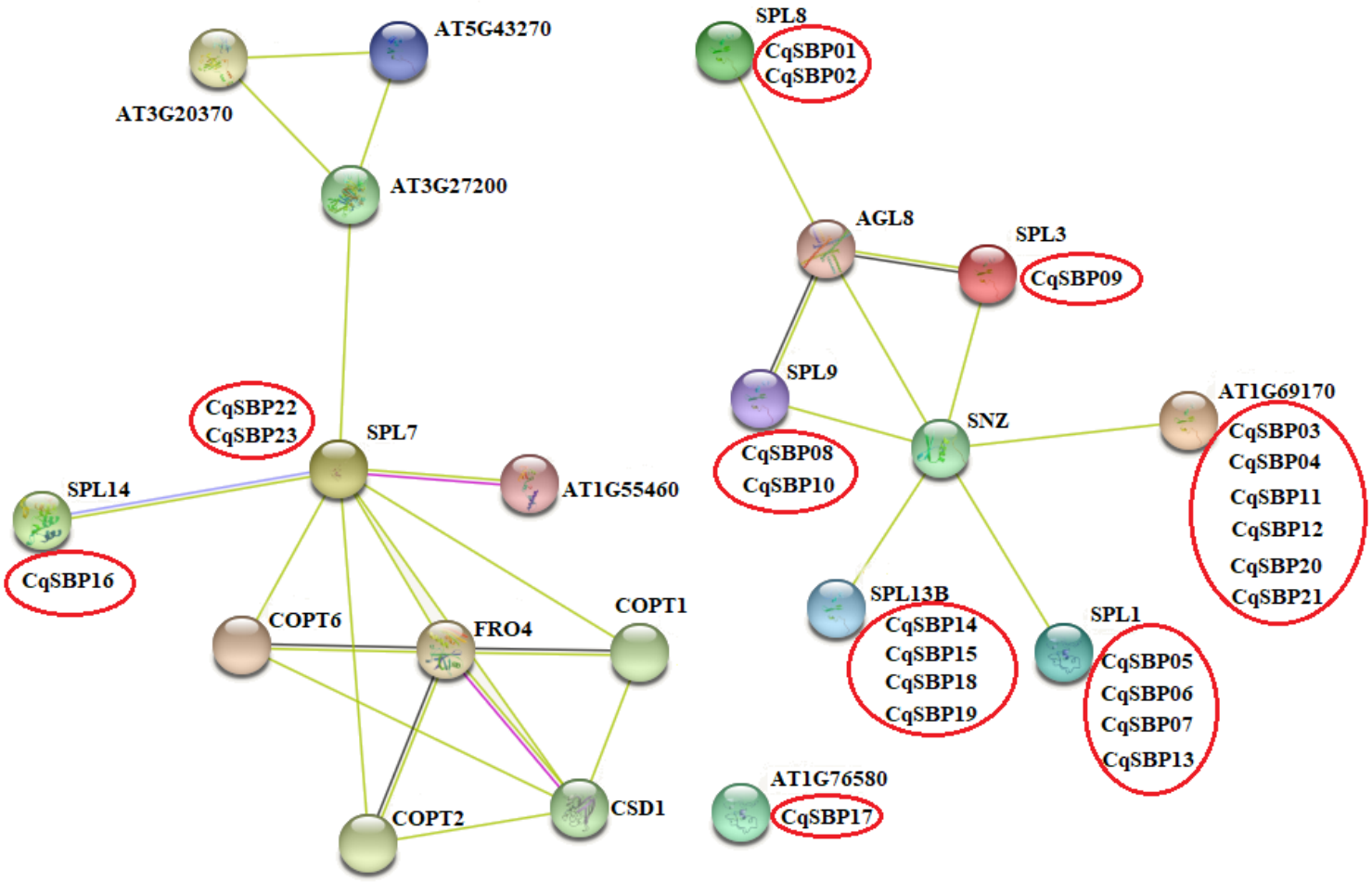

Figure 5

The potential interaction network of CqSBP based on the *Arabidopsis* and Quinoa.

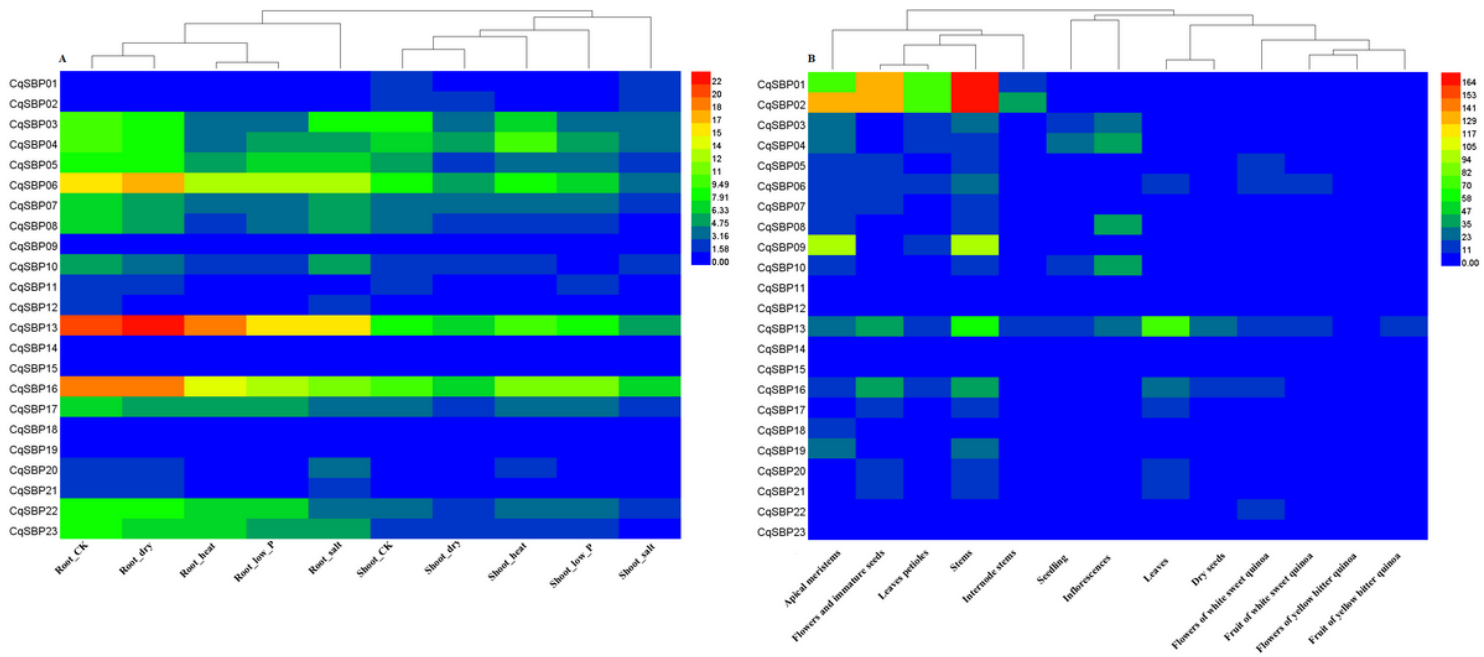

Figure 6

The expression profiles of SBP genes in different treatments and developmental stages and tissues of Quinoa.

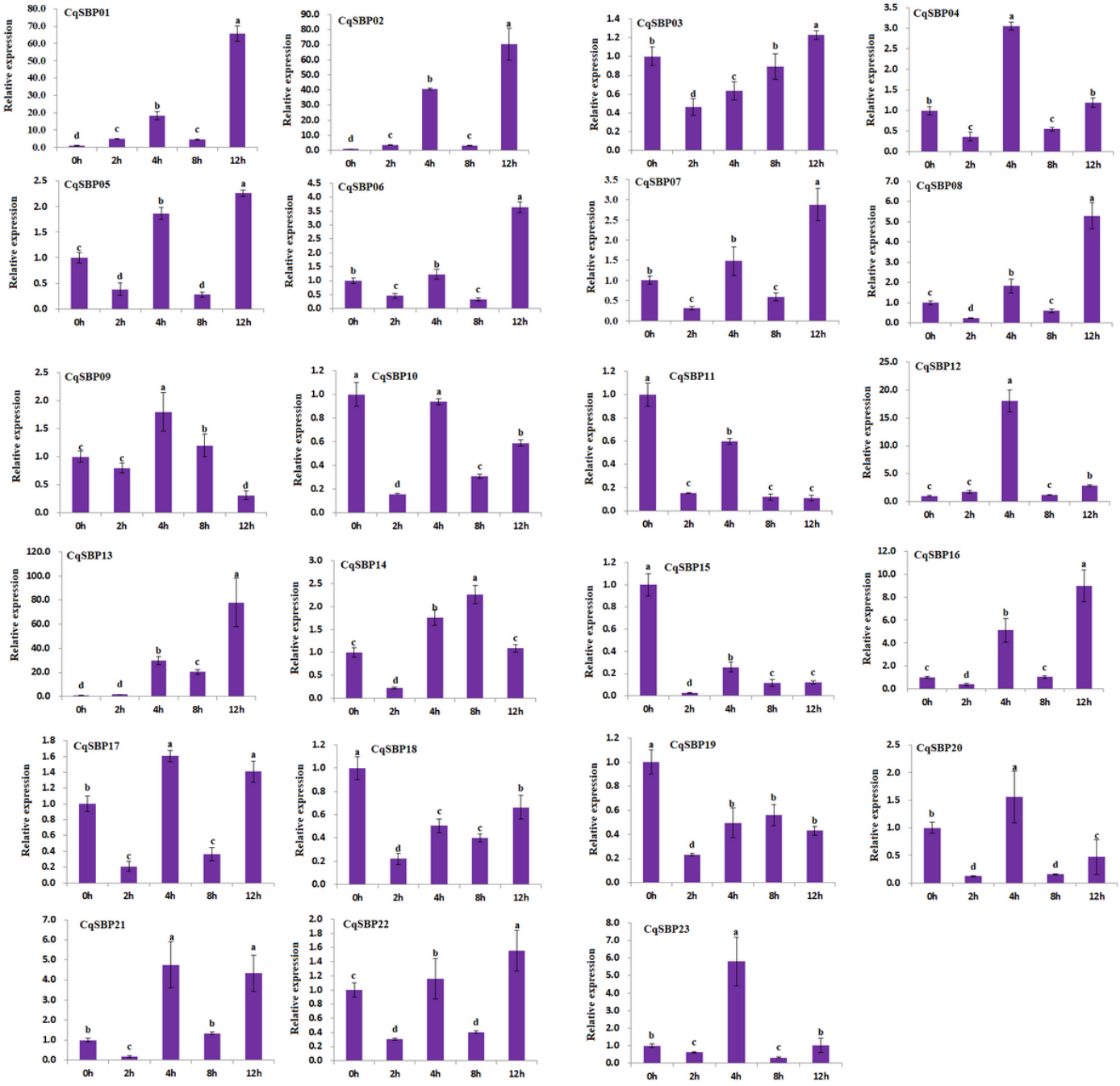

Figure 7

Supplementary Files

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

- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx

- TableS5.xlsx