




Article

Genome-Wide Investigation and Co-Expression Network Analysis of SBT Family Gene in *Gossypium*

Tianxi Xue ¹, Lisen Liu ², Xinyi Zhang ¹, Zhongqiu Li ¹ , Minghao Sheng ¹, Xiaoyang Ge ², Wenying Xu ¹ and Zhen Su ^{1,*}

¹ State Key Laboratory of Plant Environmental Resilience, College of Biological Sciences, China Agricultural University, Beijing 100193, China

² State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China

* Correspondence: zhensu@cau.edu.cn; Tel.: +86-10-62734385

Abstract: Subtilases (SBTs), which belong to the serine peptidases, control plant development by regulating cell wall properties and the activity of extracellular signaling molecules, and affect all stages of the life cycle, such as seed development and germination, and responses to biotic and abiotic environments. In this study, 146 *Gossypium hirsutum*, 138 *Gossypium barbadense*, 89 *Gossypium arboreum* and 84 *Gossypium raimondii* SBTs were identified and divided into six subfamilies. Cotton SBTs are unevenly distributed on chromosomes. Synteny analysis showed that the members of SBT1 and SBT4 were expanded in cotton compared to *Arabidopsis thaliana*. Co-expression network analysis showed that six *Gossypium arboreum* SBT gene family members were in a network, among which five SBT1 genes and their *Gossypium hirsutum* and *Arabidopsis thaliana* direct homologues were down-regulated by salt treatment, indicating that the co-expression network might share conserved functions. Through co-expression network and annotation analysis, these SBTs may be involved in the biological processes of auxin transport, ABA signal transduction, cell wall repair and root tissue development. In summary, this study provides valuable information for the study of SBT genes in cotton and excavates SBT genes in response to salt stress, which provides ideas for cotton breeding for salinity resistance.



Citation: Xue, T.; Liu, L.; Zhang, X.; Li, Z.; Sheng, M.; Ge, X.; Xu, W.; Su, Z. Genome-Wide Investigation and Co-Expression Network Analysis of SBT Family Gene in *Gossypium*. *Int. J. Mol. Sci.* **2023**, *24*, 5760. <https://doi.org/10.3390/ijms24065760>

Academic Editors: Qian-Hao Zhu and Jie Sun

Received: 23 September 2022

Revised: 15 January 2023

Accepted: 26 January 2023

Published: 17 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: cotton; SBT; co-expression network; gene family; salt stress

1. Introduction

The precise regulation of protein levels ensures the normal function of plant cells, and the regulation of protein levels depends on the balance of protein synthesis and degradation. Since proteolysis is inherently irreversible, proteases control many aspects of plant growth, development and defense by selectively degrading regulatory proteins [1]. Most proteases in plants are of the catalytic serine peptidase type. Among serine peptidases, those related to bacterial subtilisin constitute the largest family, so it is very important to study the role of the SBT gene family in plant growth and development regulation and the response to environmental regulation [2,3].

Subtilisin-like serine protease (SBT) is a family of serine proteolytic enzymes that hydrolyze proteins into small peptides, which bind to receptors as signal molecules or ligands and participate in signal transduction. It has been reported in many plants for its involvement in various cellular processes such as protein activation. This unique region is used as a binding site for specific substrates [3]. There is a highly conserved catalytic triplet in the S8 peptidase domain, and amino acid residues consisting of aspartic acid (Asp), histidine (His) and serine (Ser) are characteristic of the subtilase family [4]. In addition, certain subtilase enzymes may also have a conserved catalytic asparagine (Asn) residue in the same S8 peptidase domain [4,5]. In addition, the protease-related (PA) domain, the

C-terminal fibronectin (Fn)-III-like domain and the inhibitory (I9) domain have also been found in plant SBT gene family members [2], and the PA domain is often described as a protein–protein interaction module [6]. The PA domain interacts with substrates and contributes to substrate selectivity in plant SBT [7]. The C-terminal fibronectin (Fn)-III-like domain is a more general feature of plant SBT, functioning in certain plants [8], but not all SBT have this domain. As to SlSBT3, the C-terminal was inserted into a hydrophobic pocket near the substrate binding channel, thus stabilizing the loop system near the active site [2,8]. Inhibitor I9 has two functions. First, it assists in folding and thus requires the enzyme to mature during its passage through the secretory pathway. Inhibitor I9 deletion mutants do not pass endoplasmic reticulum (ER) quality control and remain in cells [8]. Second, Inhibitor I9 inhibits the activity of self-acting proteases. The conserved domain of SBTs is closely related to its multifunctional function in plants [8].

The SBT gene exists throughout the plant kingdom and plays a variety of roles in plant growth and defense [2]. The SBT family has been extensively studied in *Arabidopsis thaliana*. A total of 56 SBT family members have been identified [9]. AtSBT1.2 is specifically expressed in stomatal precursor cells and has the function of regulating stomatal density [10,11]. AtSBT1.4 plays an important role in regulating ABA signaling and drought tolerance through its interaction with OST1 [12]. Subtilase, which promotes the formation of the signal peptide AtSBT2.4, was found to be involved in embryonic development. Transposon markers identified AtSBT2.4 as the gene responsible for impaired epidermal function and conditional pathogenicity in *ale1* mutants [13,14]. Overexpression of AtSBT5.4 is used as a means of densifying organs produced by plants, inflorescences or floral meristem [15]. This trait may have important implications for biomass, fruit or flower production in other plants. In addition, AIR3 (AtSBT5.3) promotes lateral root formation [16], ARA12 (AtSBT1.7) participates in seed coat development [17] and XSP1 (AtSBT4.14) regulates xylem differentiation [18]. They not only play a role in plant-specific developmental processes, but also participate in the response of *Arabidopsis thaliana* to environmental stress. AtSBT3.5 was shown to participate in PME processing during root development [19]. AtSBT6.1 encodes a protein required for the activity of bZIP17 in the endoplasmic reticulum stress signaling pathway and has been observed in plants in response to salt stress [20–23]. In addition, the loss of AtSBT3.3 function in the deletion mutant impaired innate immunity, while the overexpression of AtSBT3.3 enhanced plant resistance to pathogens [24]. In cotton SBTs, GbSBT1 protein interacted with GbVE1, a verticillium-wilt-resistant protein, and GbSBT1 gene was up-regulated in response to verticillium wilt, jasmonic acid and ethylene. GbSBT1 silencing increased the sensitivity of cotton plants to *Verticillium dahliae* infection [25].

Cotton is the main source of natural fiber for textile industries around the world. The genus *Gossypium* consists of 5 tetraploid ($2n = 4 \times 13 = 52$) and 45 diploid ($2n = 2 \times 13 = 26$) species, which are presumed to have originated from the common ancestor 5~10 MYA. *Gossypium hirsutum* ($2n = 4 \times 13 = 52$) is expected to undergo an allopolyploid event at about 1~2 MYA, involving A and D genome progenitors [26]. The ancestor of *Gossypium arboreum* (AA, $2n = 2 \times 13 = 26$) may have been the donor of A subgenomes, while the ancestor of *Gossypium raimondii* (DD, $2n = 2 \times 13 = 26$) may have evolved into A allotetraploid cotton species by adding D subgenomes to *Gossypium hirsutum* (AADD, $2n = 4 \times 13 = 52$). Polyploid not only changes genome organization and gene expression, but also leads to the generation of multiple duplicate genes in the genome [27–30]. The publication of the genomic data of *Gossypium raimondii*, *Gossypium arboreum*, *Gossypium hirsutum* and *Gossypium barbadense* provides an opportunity for comparative analysis and the identification of key gene families among cotton species. Cotton is a model crop for studying polyploid evolution, as with the putative cotton and allotetraploid cotton genome decoded, scientists can learn about the molecular mechanism of cotton genome size variation, the formation and evolution of allopolyploid, the development and utilization of genetic resources and important agronomic traits, the basis for the breeding of new varieties, and to provide guidance for cotton. At the same time, with the continuous development of transcriptome sequencing technology, more and more transcriptome data are available at different stages of cotton development

and under different environmental conditions. As genomics and transcriptomics data continue to grow, scientists are integrating them to build cotton databases such as CottonGen [31] and ccNET [32]. The CottonGen database provides reference genomes for different cotton species and the ccNET database can query the co-expression network of *Gossypium arboreum* and *Gossypium hirsutum* and infer the molecular functions and biological processes involved in cotton genes. These databases provide a basis for joint multi-omics analysis to explore candidate genes for key agronomic traits.

In general, drought and salt affect all levels of cotton, from the molecular to the biological levels, which leads to a decrease in plant growth, villus production and fiber quality. This effect varies with the timing and intensity of salt or drought stress, the growth stage and species. Drought has a negative effect on all cotton growth stages. However, seedlings, flowering and boll developmental stages are the most sensitive to water scarcity. Although several studies have been conducted to improve abiotic stress tolerance in cotton, breeding achievements have fallen short of expectations due to limited tolerance germplasm resources and poor understanding of the genetic basis of complex traits such as abiotic stress tolerance. Signal transduction plays a crucial role in plant responses to different stress conditions. Most SBTs are involved in signal transduction pathways. Plant SBTs affect all stages of the life cycle as they contribute to embryogenesis, seed development and germination, cuticle formation and epidermal patterning, vascular development, programmed cell death, organ abscission, senescence, and plant responses to their biotic and abiotic environments. We wish to identify some candidate SBT genes for cotton molecular breeding in resistance to abiotic stresses. At present, the function of the SBT gene in *Gossypium hirsutum*, *Gossypium arboreum*, *Gossypium raimondii* and *Gossypium barbadense* is rarely reported. Therefore, more detailed and comprehensive studies on the molecular evolution and gene function of the cotton SBT family are needed.

In this study, we first identified the SBT gene family members of *Gossypium arboreum* (GaSBT), *Gossypium raimondii* (GrSBT), *Gossypium hirsutum* (GhSBT) and *Gossypium barbadense* (GbSBT) and classified these SBT gene family members through phylogenetic analysis of *Arabidopsis thaliana* and four cotton species. Next, we used collinearity analysis to understand the underlying genetic mechanism of SBT gene family evolution. Then, the function of SBT gene family members in cotton was predicted by a co-expression network query. This study provides a reference for the structure and function of the SBT gene in cotton, and identifies genes related to salt stress by co-expression network analysis and the prediction of the biological processes involved by its members, which provides a research basis for the molecular breeding of cotton.

2. Results

2.1. Identification of SBT Gene Family Members in Cotton

To better explore the character of the SBT gene family in cotton, we selected two allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*, and their possible ancestral diploid group A donor *Gossypium arboreum* (A2 genome) and diploid group D donor *Gossypium raimondii* (D5 genome).

We applied multiple methods including keyword query, domain query and BLAST search, and identified 146 *Gossypium hirsutum*, 138 *Gossypium barbadense*, 89 *Gossypium arboreum* and 84 (130 isoforms) *Gossypium raimondii* SBT gene family members (Table 1). There were 69 SBT genes in *Gossypium hirsutum* A subgenome, 74 in D subgenome and 3 in scaffold, and there were 67 members in *Gossypium barbadense* A subgenome, 70 in D subgenome and 1 in scaffold, thus showing there are more SBT gene family members in the D subgenome than in the A subgenome in allotetraploid cotton. Compared with ancestral diploids and heterotetraploids, the number of subgenomic SBT gene family members in *Gossypium hirsutum* and *Gossypium barbadense* was less than that in *Gossypium arboreum* and *Gossypium raimondii*, indicating that there might be gene loss in SBT gene family members in cotton during the evolutionary process.

Table 1. Statistics of SBT gene family members of four cotton species.

	<i>Gossypium hirsutum</i>				<i>Gossypium barbadense</i>				<i>Gossypium arboreum</i> A Genome	<i>Gossypium raimondii</i> D Genome
	A Subgenome	D Subgenome	Scaffold	Total	A Subgenome	D Subgenome	Scaffold	Total		
SBT1	25	23	2	50	26	25	0	51	34	27
SBT2	7	7	0	14	7	8	0	15	8	8
SBT3	5	6	0	11	6	6	0	12	6	8
SBT4	14	24	0	38	11	15	1	27	27	23
SBT5	11	9	1	21	13	10	0	23	10	11
SBT6	7	5	0	12	4	6	0	11	4	7
Total	69	74	3	146	67	70	1	138	89	84

The SBT gene family has been identified in many species, including *Arabidopsis thaliana* (56 gene members) [9], rice (*Oryza sativa*, 63) [33], poplar (*Populus trichocarpa*, 90) [3], potato (*Solanum tuberosum*, 82) [34], pineapple (*Ananas comosus* L, 54) [35] and grape (*Vitis vinifera*, 82) [36,37]. Compared to these plants, the number of SBT gene family members is higher in cotton species. As to functional domains in SBT proteins, similar to other species, SBT gene family members had at least one key domain (Inhibitor I9, PA, Peptidase S8, FnIII-like domain, TPPII) (Table S1), indicating that the functional diversity of cotton SBTs is related to its multiple protein domains. The protein lengths of SBT gene family members were different in allotetraploid cotton. *Gossypium hirsutum* was 54 (Gh_Sca122440G01) to 1393 (Gh_A08G0882) aa and *Gossypium barbadense* was 89 (GB_scaffold13001_objG0001) to 3521 (GB_D10G2624) aa (Table S1). We found that *Gossypium barbadense* has a wider range of protein lengths than *Gossypium hirsutum*. However, the differences in protein length were small in the ancestral diploids of cotton. *Gossypium raimondii* was 77 (Gorai.011G157100.1) to 1351 (Gorai.004G002500.1) aa, and that of *Gossypium arboreum* was 84 (Cotton_A_30414) to 1415 (Cotton_A_35466) aa. In *Arabidopsis thaliana*, the protein sizes of 56 SBT members ranged from 671 to 1396 aa [9]. Compared with *Arabidopsis thaliana*, cotton has a wider range of amino acid lengths.

2.2. Phylogenetic Analysis of SBT Gene Family Members in Cotton

In order to determine the taxonomic and evolutionary relationships of SBT gene family members in cotton, we used a phylogenetic tree, multiple sequence alignment, BLAST search, domain analysis and manual correction to classify SBT gene family members. In addition, *Arabidopsis thaliana* was used as the classification standard for cotton SBT because it is a model species. We took the sequences of 91 *Arabidopsis thaliana* SBT proteins (56 genes), 146 *Gossypium hirsutum* SBT proteins, 138 *Gossypium barbadense* SBT proteins, 130 *Gossypium raimondii* SBT proteins (84 genes) and 89 *Gossypium arboreum* SBT proteins (Table S1) and constructed a maximum likelihood phylogenetic tree (Figure 1). The SBT gene family was divided into six subfamilies (SBT1, SBT2, SBT3, SBT4, SBT5 and SBT6), which is consistent with the previously reported classification of SBT families in *Arabidopsis thaliana* [9], *Ananas comosus* L [35] and *Vitis vinifera* [36,37]. These results confirm the high degree of the evolutionary conserved nature of the SBT family in plants. In each subfamily, we found that SBT of the four cotton species tended to cluster together, and the ancestral diploids and heterotetraploids all appeared in the same branch, indicating that the four cotton species were closely related and evolutionarily conserved.

As shown in Table 1, the number and distribution of SBT subfamilies of the four cotton species were similar. Among all SBT gene members, SBT1 gene family members were the most, and SBT6 gene family members were the least. Since SBT1 and SBT4 were more abundant in the four cotton species than in the other subfamilies, we hypothesized that the reason for the high number might be caused by gene duplication and expansion during the evolutionary process, so we studied the formation of the SBT gene family by collinearity analysis. Past studies have found that *Arabidopsis thaliana* SBT1 has nine gene family members and SBT6 has six [9]. Compared to *Arabidopsis thaliana*, the much greater number of SBTs in cotton indicates that the SBT1 members in cotton might have a broader function. However, cotton SBT6 is the smallest subfamily, which is consistent with the results of *Arabidopsis thaliana*, suggesting that this subfamily may not have undergone evolutionary

differentiation between dicotyledons and monocotyledons. In addition, compared with *Arabidopsis thaliana* in the SBT4 and SBT5 subfamilies' classification, some *Arabidopsis thaliana* SBT gene family members are in a branch, while the other subfamilies do not have the above results (Figure 1). In the SBT4 and SBT5 subfamilies, *Arabidopsis thaliana* and cotton may have great sequence differences.

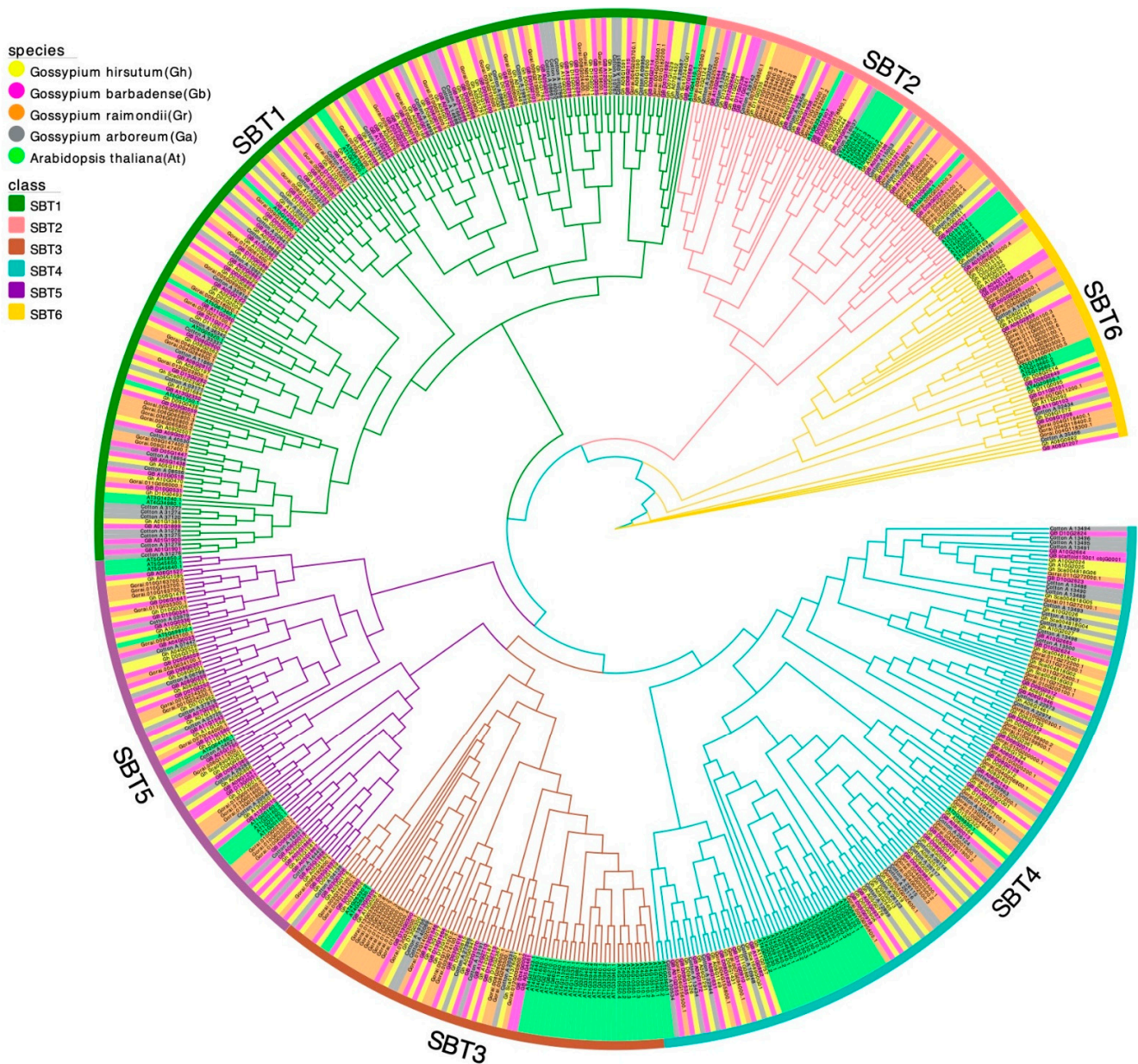


Figure 1. Phylogenetic relationships of SBT in different species. The circles around the phylogenetic tree (from center to periphery) represent each SBT subfamily and the distribution of species in different colors. Dark green, pink, brown, cyan, purple and dark yellow lines represent subfamilies from SBT1 to SBT6, respectively. Different gene names and background colors indicate different species. Yellow, dark pink, orange, grey and green indicate *Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium raimondii*, *Gossypium arboreum* and *Arabidopsis thaliana*, respectively.

We selected one protein each for the six subfamilies of *Arabidopsis thaliana* and *Gossypium arboreum* and obtained their predicted three-dimensional protein structures by using alphafold2. In previous studies, SBT has the Peptidases S8 domain, which has three enzyme active sites [2]. We compared the reported three-dimensional structure of the SBT protein

sequences with the SBT proteins of the six subfamilies of *Arabidopsis thaliana* and *Gossypium arboreum* to find the locations and domains of Asp, His and Ser residues in catalytic triplets in the predicted three-dimensional structure of proteins (Figures S1 and S2).

The predicted PA, FnIII, Peptidases S8 domains and catalytic triplet structures are similar to those reported for the protein three-dimensional structures of cucumber and SISBT3 [2], suggesting that these domains are structurally conserved in *Arabidopsis thaliana* and *Gossypium arboreum*. A comparison between the structures of each subfamily found that only the Peptidases S8 domain of SBT6 differed the most from the other subfamilies, indicating differences in function. The different positions of the Inhibitor I9 domain in different subclasses, such as SBT1 and SBT3 on the left of the Peptidases S8 domain and SBT2, SBT4 and SBT5 behind Peptidases S8, suggest that each subfamily may be different in inhibiting its own enzyme activity through the Inhibitor I9 domain. In addition, random curling around the β -sheet of the FnIII domain in each subclass was found to be different, and FnIII had a hydrophobic effect, further improving the stability of the enzyme (Figures S1 and S2).

2.3. Gene Structure Analysis of SBT Gene Family Members in Cotton

Divergences in coding regions, especially those that may change the function of the gene, can be caused by amino-acid-altering substitutions and/or alterations in the exon–intron structure. To further understand their gene structure diversity, we analyzed the exon–intron organization of cotton SBT genes.

The conserved gene structure of each subgroup suggests that SBT genes with high homologous sequence similarity tend to have the similar gene structure. There was little difference in the number of exons of the SBT subfamily among the four cotton species, indicating that SBT was evolutionarily conserved in the four cotton species. On the other hand, there were significant differences in the number of exons among the six SBT subfamilies among the four cotton varieties. The average number of exons in SBT6, SBT2– SBT5 (SBT2, SBT3, SBT4 and SBT5) and SBT1 was 14, 9 and 1, respectively. We found a significant difference in the number of exons between SBT1 and SBT6 (Figures 2 and S3), which suggests that SBT1 and SBT6 gene structures diverge to improve the fitness of the organisms [38]. These results are consistent with those of *Arabidopsis thaliana* (SBT1: average exon number 1, SBT6: average exon number 22).

To further understand the gene structure of SBT1 and determine whether it is conserved in other species. The number and proportion of SBT1 genes without introns were counted in four cotton species, and were 74% in *Gossypium hirsutum* (37/50), 70.59% in *Gossypium barbadense* (36/51), 55.5% in *Gossypium raimondii* (15/27) and 76.47% in *Gossypium arboreum* (26/34) (Figure 2 and Figure S3). These results suggest that most SBT1 gene family members of the four cotton species did not have introns, which was consistent with the structural characteristics of the SBT gene in *Arabidopsis thaliana* (7/9, 77.78%) [9] and *Ananas comosus* L (12/30, 40%) [35]. In addition to the gene structure, we also analyzed the domain of SBT1 and found that most proteins have four key domains (Inhibitor I9, Peptidases S8, PA domains and FnIII-like domain) (Figures S1, S2 and S4).

In SBT1 branches of the evolutionary tree, homologous genes share the same domain, gene length and exon number. For example, the SBT1.8 tree branches into two groups. AtSBT1.8 in a separate clade indicates that it evolved in parallel with cotton SBTs. Cotton_A_33337 in SBT1.8 is in a branch with the D genome and is included in the CottonGen database [31], with only one domain. Except for Cotton_A_33337, the subgenomes of the remaining genes were in a branch with the corresponding ancestral diploid genome, and the heterotetraploid had twice as many memories as the ancestral diploid. Compared with other SBT1 clades, SBT1.8 was mainly composed of two exons in the gene structure analysis. At the same time, SBT1.8 contained two domains in the domain analysis. The first domain was all four key domains (Inhibitor I9, Peptidases S8, PA domain and FnIII-like domain). The second was multiple AprE based on the four key domains (Figure S4).

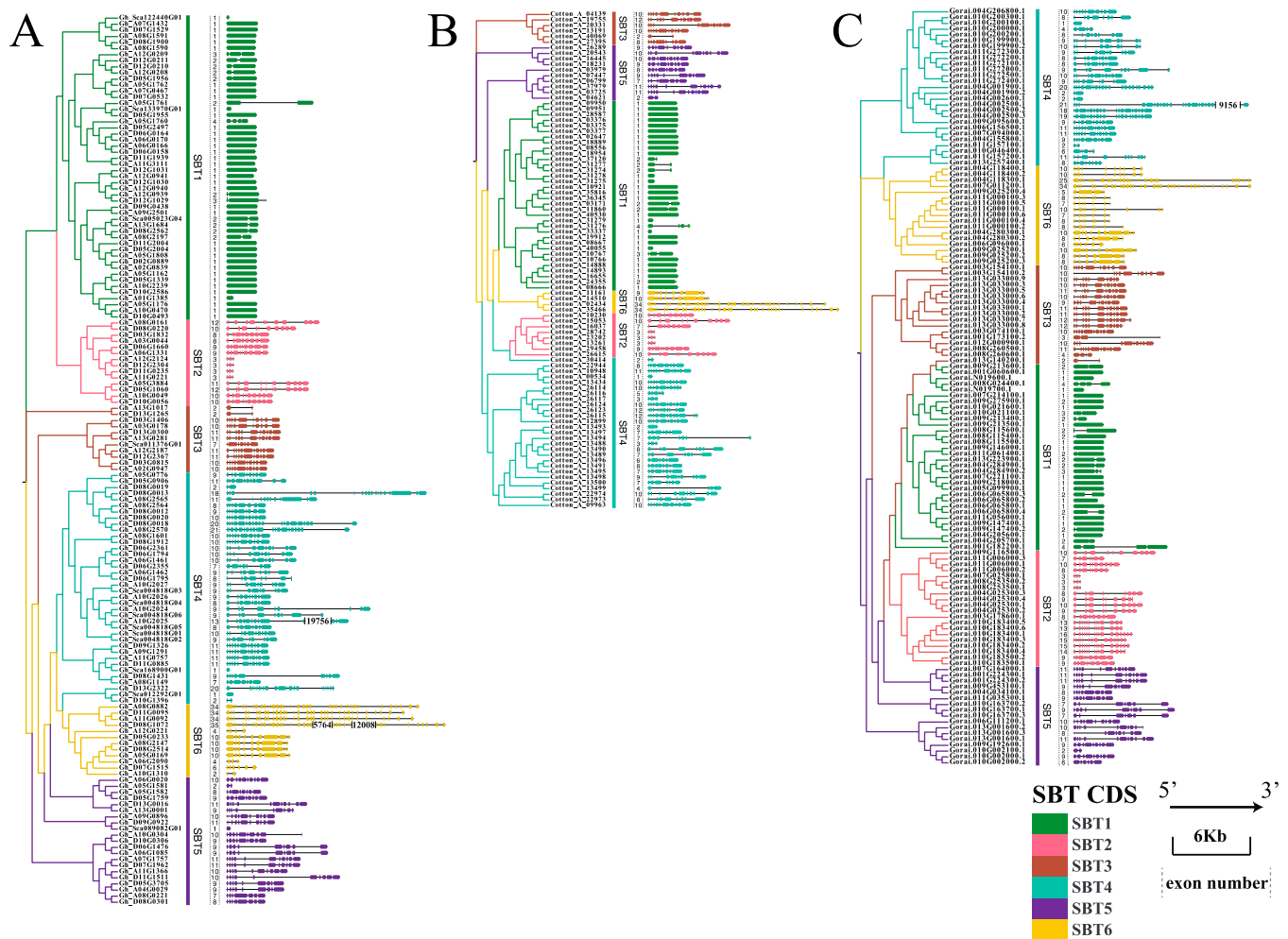


Figure 2. SBT gene structure of cotton. (A) SBT gene structure of *Gossypium hirsutum*. The phylogenetic tree is built by FastTree. (B) Genetic structure of *Gossypium arboreum*. The phylogenetic tree is built by FastTree. (C) Isoform structure of *Gossypium raimondii*. Different gene models as a result of alternative splicing or alternative start codons are numbered consecutively as Gorai.0XXGXXXXX.1, 2, etc. The evolution tree is built by FastTree. The squares are CDS, the horizontal lines are introns. Dark green, pink, brown, cyan, purple and dark yellow lines represent subfamilies from SBT1 to SBT6, respectively. The dotted line is exon number.

In addition, three Inhibitor I9, Peptidases S8, PA domains and FnIII-like domains were found in GB_A10G2665 and GB_D06G2013. GB_A10G2664 has five Inhibitor I9, Peptidases S8, PA domains and FnIII-like domains, and GB_D10G2624 has four Inhibitor I9, Peptidases S8 and PA domains, indicating that domain replication events may occur during the evolution of this gene (Figures S3 and S5 and Table S1). These genes belong to GbSBT4, and we suspect that *Gossypium barbadense* SBT4 has a gene expansion phenomenon.

2.4. Chromosomal Localization, Collinearity Analysis and Selective Pressure Analysis of the SBT Family Genes in Cotton

To better understand the distribution of SBT on the chromosome, genome and subgenome of cotton, a chromosome map of SBT was constructed based on the genome sequences of the four cotton species. We found that SBT is unevenly distributed on the chromosomes in the four cotton species (Figures 3 and S6). In the case of *Gossypium hirsutum*, first of all, the four cotton species differ in the number of genes on each chromosome. In all chromosomes, D08 chromosome has the most SBT genes, but D01 and D04 have no

SBT gene. Second, the distribution of SBT genes on a single chromosome is not uniform, and some SBT genes exist in gene clusters, most of which are SBT1 and SBT4.

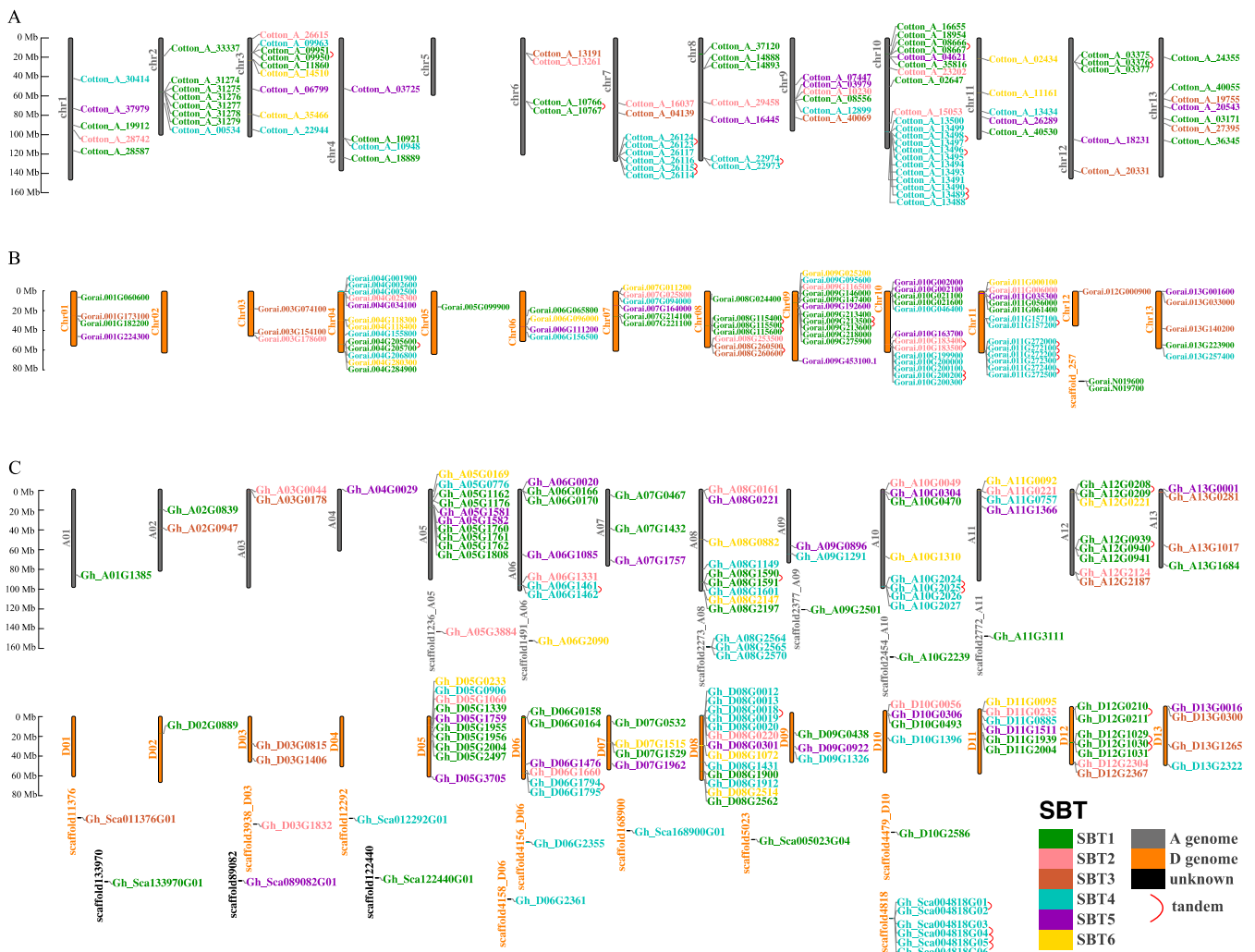


Figure 3. Chromosome positions of SBT gene family members in cotton. (A) *Gossypium arboreum*. (B) *Gossypium Raimondii*. (C) *Gossypium hirsutum*. The grey bars represent the chromosomes of the A subgenome. The orange bars show the chromosomes of the *Gossypium Raimondii* D subgenome. Black represents unpredicted subgenomic classifications. Different color gene names represent different classifications. Dark green, pink, brown, cyan, purple and dark yellow lines represent subfamilies from SBT1 to SBT6, respectively. The red curve connects tandem repeats, gene locations and chromosome sizes estimated in Mb on the left side of the figure.

Gene families commonly arise because of gene duplication events, mainly including tandem, segmental and whole-genome duplications. We used synteny analysis to understand how the SBT gene family was formed. We found that there were larger fragment duplications than tandem duplications in the four cotton species, indicating that the SBT gene family was formed mainly through segmental duplication. In tandem duplication, SBT1 and SBT4 accounted for more than any other subfamily (Figure 3 and Table S2). This result may indicate that most of the gene clusters on a single chromosome are SBT1 and SBT4. At the same time, collinearity analysis was performed on selected portions of tandem duplication genes of SBT1 and SBT4. In SBT1 tandem genes, the collinear genes corresponding to Cotton_A_03375, Cotton_A_03376 and Cotton_A_03377 are Gh_A12G0939, Gh_A12G0940 and Gh_A12G0941 on chromosome A12, meanwhile, they are Gh_D12G1029, Gh_D12G1030 and Gh_D12G1031 on chromosome D12. In SBT4 tandem genes, the collinear genes cor-

responding to Cotton_A_13490, Cotton_A_13495, Cotton_A_13497 and Cotton_A_13498 are Gh_A10G2024, Gh_A10G2025, Gh_A10G2026 and Gh_A10G2027 located on chromosome A10, as well as Gh_Sca004818G06, Gh_Sca004818G05, Gh_Sca004818G04 and Gh_Sca004818G03 located on chromosome scaffold4818. These results suggest that the selected tandem duplication genes of SBT1 and SBT4 in cotton are conserved in the evolution of tetraploid cotton and in diploid cotton (Figure S7).

In addition, the nonsynonymous (Ka) to synonymous (Ks) substitution rate ratio (Ka/Ks) was used to serve an estimator for selective pressure on DNA sequence evolution. Selective pressure analysis showed that most of the Ka/Ks pairs of homologous genes were less than 1 (Table S2), indicating that the cotton SBT gene was a purified selection.

2.5. Co-expression Network Analysis of *Gossypium arboreum* SBT

Compared with other cotton species, *Gossypium arboreum* has advantages in terms of its stress resistance [39]. In order to explore the function of the SBT gene in cotton, we used the ccNET [32] co-expression network database to analyze the expression pattern of the *Gossypium arboreum* SBT gene in root, stem and leaf tissues and under drought and salt stresses.

The *Gossypium arboreum* and *Gossypium hirsutum* SBT gene family members were divided into three groups by expression profile clustering analysis. The first type is the up-regulation of most genes under salt and drought stresses, the second type is the unchanged expression and the third type is the down-regulation of most genes under salt and drought stresses. In the first category, every subfamily is present and evenly distributed, in the second category, most are SBT4, and in the third category, every subfamily is present but the number of SBT1 is the highest (Figure 4 and Table S3). In addition, we found that there was little difference in the expression trend between the A subgenome and the D subgenome of *Gossypium hirsutum*. The gene expression trend of *Gossypium hirsutum* gene family members in salt treatment and PEG treatment was the same as that in *Gossypium arboreum*, which further indicated that diploid cotton and allotetraploid cotton were conserved (Figure S8 and Table S4).

As shown in Figure 4, genes with a similar variation trend of gene expression were gathered together, and we conducted a co-expression network query for the genes clustered in the same category in the ccNET database [32]. We found that six *Gossypium arboreum* SBT gene family members (Cotton_A_03171, Cotton_A_03377, Cotton_A_35816, Cotton_A_08556, Cotton_A_18889, Cotton_A_07447) were found in one network that consists of 139 co-expressed genes (Figures 4 and 5, Table S3). To explore the biological function of this network, we used AgriGOv2 [40] to screen the functional modules in the co-expression network and annotated the genes through a literature search. It was found that these six SBT gene family members of *Gossypium arboreum* were enriched in serine-type endopeptidase activity. In the DAG enrichment diagram (Figure S9), kinase activity and protein phosphorylation terms were significantly enriched, and protein kinases catalyzed the phosphoric reaction of proteins, and receptor protein kinases were found in these terms (Table S5). These results indicate that the network is involved in signal transduction and affects cotton growth and development.

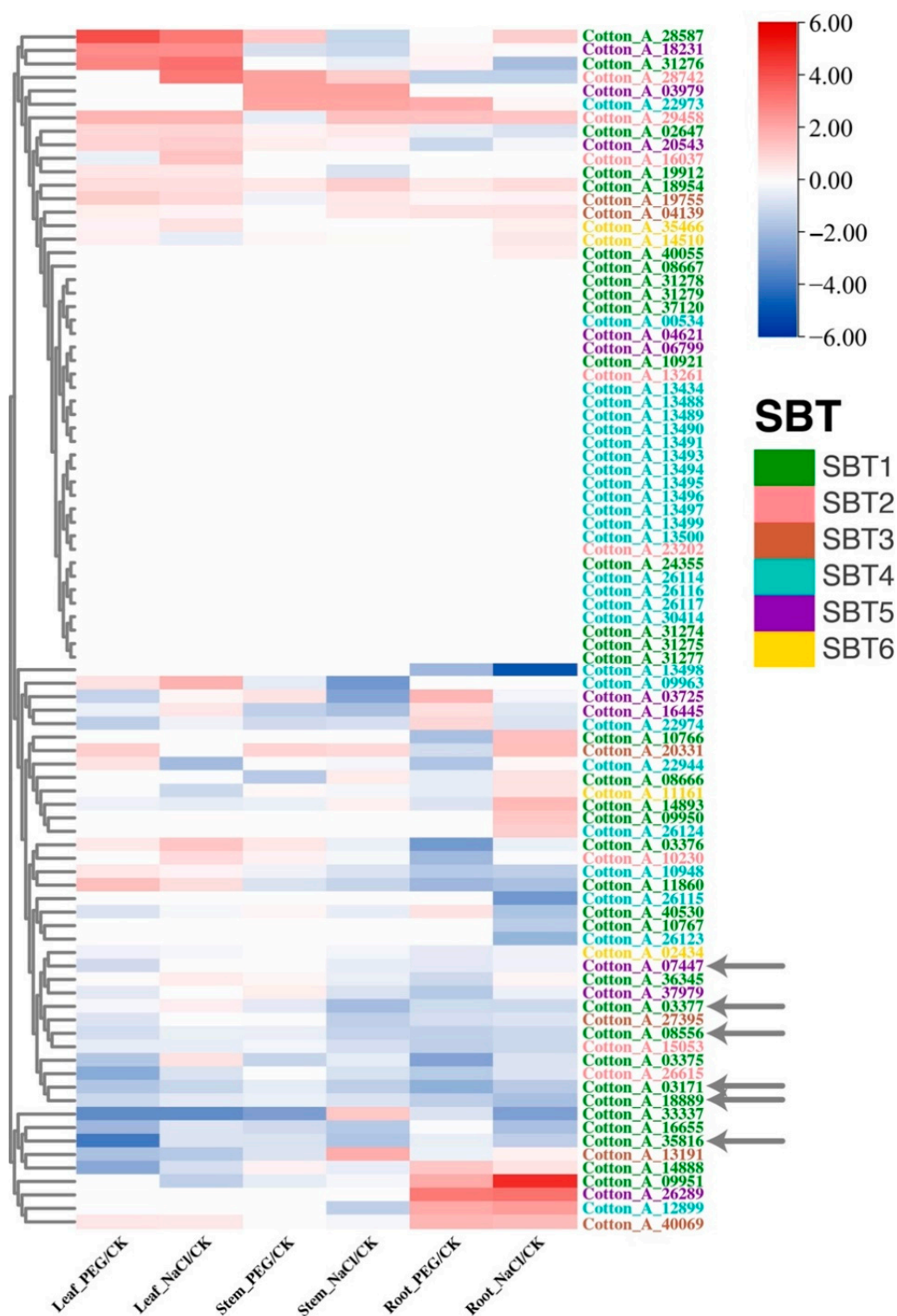


Figure 4. Expression profiles of 89 SBT gene families in *Gossypium arboreum*. There are 89 *Gossypium arboreum* gene family members in the vertical, and the ratio of PEG and NaCl treatment to control in leaf, stem and root tissues in the horizontal sequence, and the classification of each SBT gene family member is at the far right. Different color fonts indicate the subfamilies from SBT1 to SBT6. Dark green, pink, brown, cyan, purple and dark yellow lines represent subfamilies from SBT1 to SBT6, respectively.

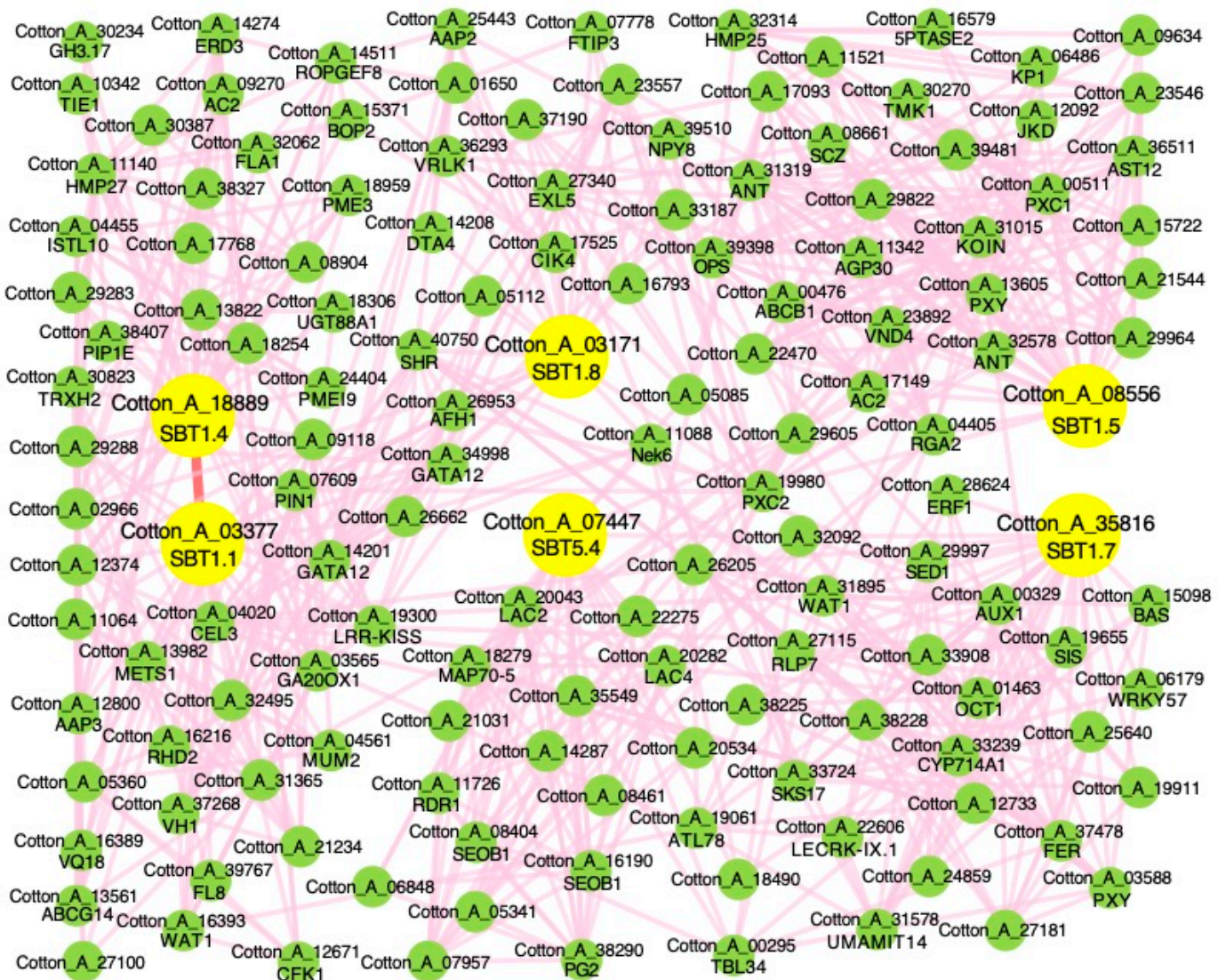


Figure 5. Gene co-expression network of six *Gossypium arboreum* SBT gene family members. The green nodes are co-expressed genes, and the large yellow nodes are *Gossypium arboreum* SBT genes. The pink line shows a co-expression relationship between two genes and the red line indicates SBT gene co-expression.

Because *Arabidopsis thaliana* is a dicotyledonous model plant, we searched for *Arabidopsis thaliana* orthologous genes co-expressed in *Gossypium arboreum* to further predict cotton gene function. We found many receptor kinases in gene sets with terms such as protein phosphorylation, transmembrane receptor protein tyrosine kinase signaling pathway, plasma membrane, ATP binding and plant-type cell wall (Figure S9 and Table S4). Among receptor kinases, AtTMK1 is related to root tissue development [41] and LRR-KISS responds to salt stress [42]. PXC1 may be involved in the regulation of cell elongation or swelling [43], and AtVH1 affects the proliferation and differentiation of many cells in young leaves [44]. Cheng Huang et al. found that down-regulation of AtVRLK1 could promote the thickening of the secondary cell wall, while up-regulation of AtVRLK1 could enhance cell elongation and inhibit the thickening of the secondary cell wall [45]. These results suggest that receptor kinases in the co-expression network of *Gossypium arboreum* may be involved in root development and the response to salt stress.

Therefore, we performed cluster analysis on the gene expression profiles of the six SBT co-expressed gene networks in different tissues under abiotic stresses and found that most genes were down-regulated in the roots, stems and leaves of *Gossypium arboreum*

after salt treatment and drought treatment. Five of the six SBT gene family members (Cotton_A_03171, Cotton_A_03377, Cotton_A_35816, Cotton_A_08556, Cotton_A_18889) were significantly down-regulated (Figure S10A, Table S6). We found that receptor kinases PXC1 [43], VH1 [44], VRLK1 [45] and lecrk-IX.1 [46] were down-regulated, SCZ [47] and SHR [48] related to root tissue development (Table S4) were down-regulated under salt treatment, and AUX1 [49] and ABCB1 [49] related to polar auxin transport were down-regulated (Figure S10b), suggesting that SBT1 was involved in auxin transport, root tissue development and the response to salt stress.

2.6. Comparison of Co-Expression Networks between *Arabidopsis thaliana* and Different Cotton Species

In addition to the five SBT1 genes in *Gossypium arboreum* that were down-regulated by salt treatment in the root tissue, we further investigated the co-expression network of five direct SBT1 orthologous genes in *Gossypium hirsutum* and *Arabidopsis thaliana*, in order to compare the network conservative types between species.

Firstly, five SBT1 were analyzed by collinearity analysis and a ccNET database [32] query to obtain direct homologous genes of *Gossypium hirsutum* corresponding to *Gossypium arboreum* (Figure S11) and to query the SBT co-expression network of *Gossypium hirsutum*. We found that Gh_D10G0493, Gh_A13G1684 and Gh_Sca005023G04 corresponding to Cotton_A_08556 and Cotton_A_03171 had collinearity in *Gossypium hirsutum* (Table S7), and were in the same co-expression network (Figure 6 and Table S8). The co-expression network and annotation analysis showed that the leucine-rich repeat protein kinase family (PXY and PXC2) of *Gossypium arboreum* and its homologous *Gossypium hirsutum* genes were involved in signal transduction. According to the literature notes, multiple families of leucine-rich repeat protein kinases were found to be involved in root growth and cell wall formation (Table S5). These results indicate that the co-expression networks of both *Gossypium hirsutum* and *Gossypium arboreum* are involved in signal transduction pathways.

Then, to compare the conservation of *Arabidopsis thaliana* and the cotton co-expression networks, we queried the *Arabidopsis thaliana* co-expression networks in the ATTED-II database. The *Arabidopsis thaliana* orthologous genes corresponding to *Gossypium arboreum* SBT genes were obtained from the CottonGen database [31], and we found that AT2G05920, AT5G67360, AT3G14067, AT1G01900 and AT3G14240 are on the same network (Figure S12a and Table S9). Finally, SEACOMPARE analysis was performed on *Gossypium arboreum*, *Gossypium hirsutum* and *Arabidopsis thaliana* to explore the functional conserved networks. The co-expression network between the species involved in the biological processes of protein serine/threonine kinase activity, plant-type cell wall, transmembrane receptor protein tyrosine kinase signaling pathway, transmembrane transporter activity and serine-type endopeptidase activity were significantly enriched in *Arabidopsis thaliana*, *Gossypium arboreum* and *Gossypium hirsutum* (Figure S12b). SBT gene family members were enriched into serine-type endopeptidase activity. The contents of the transmembrane receptor protein Tyrosine kinase signaling pathway and protein phosphorylation were significantly enriched in *Arabidopsis thaliana*, *Gossypium arboreum* and *Gossypium hirsutum*. Protein kinases catalyze the phosphoric acid reaction of proteins and receptor protein kinases are found in these terms. Apoplast and the integral component of the membrane were significantly enriched in *Arabidopsis thaliana*, *Gossypium arboreum* and *Gossypium hirsutum* (Figure S12b). These results suggest that both the cotton and *Arabidopsis thaliana* co-expression networks are involved in signal transduction and affect plant growth.

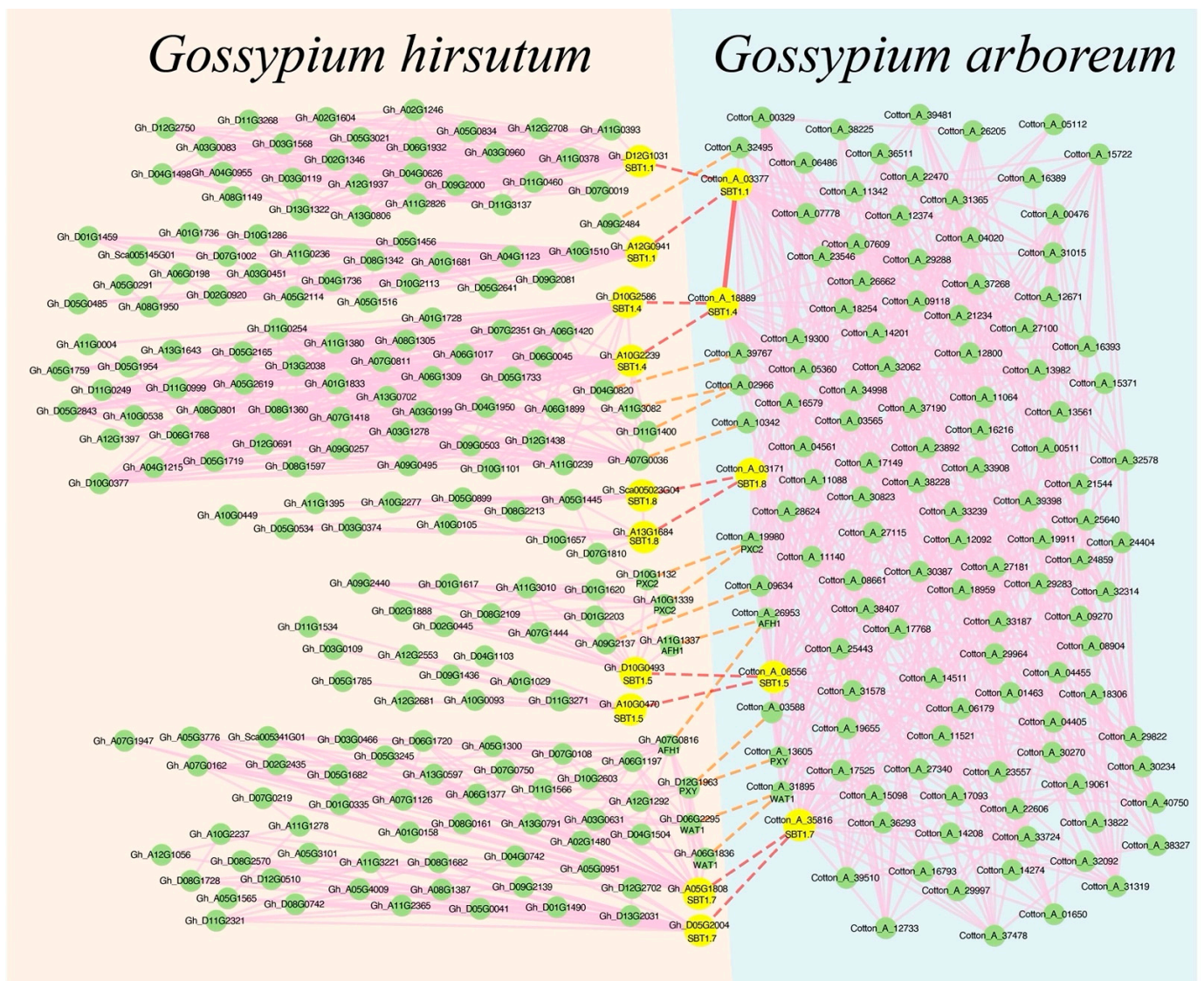


Figure 6. Comparison of SBT network between *Gossypium arboreum* and *Gossypium hirsutum*. The green nodes are co-expressed genes, and the large yellow nodes are *Gossypium arboreum* and *Gossypium hirsutum* SBT genes. *Gossypium arboreum* co-expressed network is blue background and *Gossypium hirsutum* is yellow. The pink line shows a co-expression relationship between two genes and the red line indicates SBT gene co-expression. The red dot line indicates that SBT is lineally homologous between two cotton species and the orange dot line indicates that co-expressed genes are lineally homologous between two cotton species.

As shown in Figure S12b, the response to water deprivation and the response to osmotic stress were enriched in *Gossypium hirsutum*. Therefore, expression profile clustering analysis was performed on transcriptome data related to the water stress of *Gossypium hirsutum* and *Arabidopsis thaliana* to determine the conserved type of network function.

The expression profile clustering analysis showed that most of the co-expressed genes in *Arabidopsis thaliana* and *Gossypium hirsutum* were down-regulated in salt treatment (Figures S10a and S12c,d, Tables S10 and S11). In addition, we found that five SBT1 genes in *Gossypium arboreum* and *Arabidopsis thaliana* were down-regulated by 6 h salt treatment in root tissues. Thus, we concluded that the co-expression network of *Gossypium arboreum* was conservative in *Gossypium hirsutum* and *Arabidopsis thaliana*.

3. Discussion

In this study, 146 *Gossypium hirsutum*, 138 *Gossypium barbadense*, 89 *Gossypium arboreum* and 84 *Gossypium raimondii* SBT gene family members were identified. There were 69 SBT genes in *Gossypium hirsutum* A subgenome, 74 in D subgenome and 3 in scaffold, as well as 67 members in the *Gossypium barbadense* A subgenome, 70 in the D subgenome and 1 in scaffold (Table 1). The SBT gene family was divided into six groups by BLAST and phylogenetic analysis. We found that most of the SBT1 gene family members in the four cotton species had no introns (Figure S4). Chromosome location and synteny analysis showed that SBT gene family members were not evenly distributed on each chromosome. The SBT gene family was mainly formed by segmental duplication, and SBT1 and SBT4 genes expanded significantly (Figures 3 and S6). Combined with the results of gene structure analysis and collinearity analysis, it was found that cotton SBT1 expanded significantly, and the function of cotton SBT1 was mined and predicted. Co-expression network analysis revealed six SBT gene family members (Cotton_A_03171, Cotton_A_03377, Cotton_A_35816, Cotton_A_08556, Cotton_A_18889, Cotton_A_07447) in a network (Figures 4 and 5). The annotation of the network revealed that the network was involved in signal transduction and root tissue growth and development. Five SBT1 (Cotton_A_03171, Cotton_A_03377, Cotton_A_35816, Cotton_A_08556, Cotton_A_18889) were down-regulated in response to salt stress in root tissue (Figure S10). We found five direct orthologous genes of SBT1 corresponding to *Gossypium hirsutum* (Figure S11) and *Arabidopsis thaliana* and queried the co-expression network. By enrichment analysis and expression profile clustering analysis, we found that the function of the network was conserved in *Gossypium arboreum*, *Gossypium hirsutum* and *Arabidopsis thaliana* (Figures S10 and S12).

3.1. Evolution of Cotton SBT Family Genes

In this study, we discussed the evolution of SBT family genes. Among the identified SBT members, we found that the number of allotetraploid subgenomes was less than the number of ancestral diploids. In allotetraploid cotton, there are more SBT gene family members in the D subgenome than in the A subgenome. These results indicate that there is asymmetry in gene loss among allotetraploid subgenomes, and the loss of the A subgenome is more common than that of the D subgenome. These are consistent with the study of Zhang et al. [39]. At the same time, the number of each subfamily was compared, and it was found that SBT1 and SBT4 lost more genes. Based on chromosomal location and collinearity analysis, it was found that there was an uneven distribution of the SBT gene on the A chromosome, and the SBT gene alignment of subgenome A to *Gossypium raimondii* and subgenome D to *Gossypium arboreum* showed that the subgenome had great structural variation, chromosomal cross swap and structural rearrangement after the formation of tetraploids. In conclusion, the SBT gene family analysis showed that the two subgenomes of tetraploid cotton showed asymmetric evolution in terms of the loss of genome structure rearrangement genes after the polyploidy event.

In addition, SBT gene structure analysis showed that there was little difference in the gene structure and exon number among the four cotton species in the same subfamily. In the protein domain analysis, all the SBT gene family members of the four cotton species had at least one key SBT domain. In transcriptional change analysis, there was little difference between the expression trends of subgenome A and subgenome D (Figure S8). SBT1 has the least number of exons among the six subfamilies. The simpler the gene structure means the easier it is for gene amplification; thus, SBT1 has the most gene family members. Among the five down-regulated SBT1 genes found in the co-expression network, SBT1.8 has two exons and very short introns on average, which may be the structure formed after gene expansion (Figures S4 and S10).

3.2. Cotton SBT Gene Function during Environmental Response

In order to understand the function of cotton SBT under salt stress, we analyzed the *Gossypium arboreum* co-expression network of SBT gene family members and found that

six SBT were in the same network, among which five SBT1 genes were down-regulated in response to salt stress (Figure S10b). GO enrichment analysis and literature annotation analysis of the network showed that the network was involved in root tissue development in response to salt stress (Figure S9 and Table S5).

We identified genes associated with phytohormone synthesis in the *Gossypium arboreum* co-expression network. These genes play key roles in salt stress adaptation. We found that auxin-transport-related genes WAT1 [50], PIN1 [51], AUX1 [49] and ABCB1 [49] and root-tissue-development-related genes SHR [48], SCZ [47], VH1 [44], TMK1 [41] and CEL3 [52] were down-regulated in root tissue salt treatment. These results suggested that *Gossypium arboreum* SBT may be involved in auxin transport and root tissue development. Moreover, it was further demonstrated that auxin levels and the expression of auxin transporters were decreased under salt stress, which resulted in decreased root meristem activity and growth inhibition of primary roots [53]. Therefore, understanding the mechanisms regulating auxin signaling under salt stress will be useful for future biotechnology applications to improve plant salt tolerance [54]. In addition to auxin, the ABA signal transduction pathway is a research hotspot in a plant's response to salt stress. *Gossypium arboreum* SBT1.4 was co-expressed with SBT1.1. The AtSBT1.4 mutant SASP showed enhanced ABA sensitivity during seed germination and seedling growth, which enhanced ABA-mediated leaf senescence [12]. SCR can directly regulate the expression of several transcription factors controlling ABA response [47]. These results suggest that both SBT1.1 and SBT1.4 in *Gossypium arboreum* may be involved in the ABA signal transduction pathway in response to salt stress affecting root growth and development. In the co-expression network of *Arabidopsis thaliana*, *Gossypium hirsutum* and *Gossypium arboreum*, many co-expressed genes were found to be receptor kinases (Figure S12b, Tables S5, S8 and S9). According to the literature query, the receptor kinase FER binds to RALF peptide, the homologous receptor of TDIF peptide is PXY and the small signal peptide formed by SBT hydrolysis may be RALF or TDIF [55]. In addition, we found AT3G13510 and AQI peptides in the *Arabidopsis thaliana* co-expression network, and these peptides may be obtained by SBT hydrolysis (Table S9).

SEACOMPARE analysis showed that *Gossypium arboreum*, *Gossypium hirsutum* and *Arabidopsis thaliana* were enriched to the cell wall (Figure S12b). In plants, cell walls not only determine cell expansion and growth but also provide mechanical protection against abiotic stress [56]. Accumulated evidence suggests that maintaining cell wall integrity is essential for plant salt tolerance and root tissue development. Most SBTs are targeted to the cell wall, where they contribute to the control of growth and development by regulating the properties of the cell wall and the activity of extracellular signaling molecules. AtSBT1.7 triggers the accumulation and/or activation of cell-wall-modifying enzymes that are required for external primary cell wall loosening or promoting mucus expansion, as indicated by increased pectin methyl esterase activity during AtSBT1.7 mutant seed development. FERONIA (FER) is a local receptor kinase in the plasma membrane, which is considered to be one of the cell wall integrity sensors. FER is involved in the general stress response by interacting with the abscisic acid (ABA) signaling pathway. The mutation of the FER gene leads to enhanced leaf bleaching and root cell swelling under salt stress [55]. It is proposed that FER is involved in the regulation of cell wall repair under salt stress through Ca²⁺-mediated signaling pathway. In the network, FER expression in *Gossypium arboreum* was down-regulated by salt treatment in root tissues (Figure S10b), suggesting that SBT may participate in the Ca²⁺-mediated signaling pathway in response to salt stress and participate in regulating cell wall biological processes [57]. In conclusion, SBT may be involved in the biological processes of auxin transport, ABA signal transduction, cell wall repair and root tissue development. We mined *Gossypium arboreum* SBT1 genes and their corresponding *Gossypium hirsutum* genes, and they will be verified in the future. Since cotton is not a food crop but a cash crop, it is important to mine the SBT gene associated with salinity tolerance to allow more fertile land to be used for food crops.

There are deficiencies in this study: First, at present, the co-expression network of *Gossypium barbadense* and *Gossypium raimondii* has not been constructed. More omics data should be provided to ensure the accuracy of the network and improve the prediction accuracy, so as to provide a reference for the subsequent mining of key agronomic trait genes and provide a basis for further comparison between diploid ancestors and heterotetraploids. Second, SBT genes related to root tissue development in response to salt stress need to be further verified by experiments. For example, compared with the four SBT1 down-regulated by other salt treatments, most SBT1.8 had two exons (Figure S4), and the gene down-regulated by salt treatment was obvious in the expression profile analysis (Figures 4 and S8, Tables S3 and S4). With multiple omics data, we are able to study SBT to integrate genomics and transcription data analysis, and by combining the apparent omics three genomics data analysis, through the analysis of more restructuring, we could find cotton SBT under different stress levels and states of transcriptional expression and the contact between them, and further exploring more genes related to agronomic traits.

4. Materials and Methods

4.1. Identification of SBT Gene Family Members in Cotton

Genome sequence data for *Gossypium* ssp. were obtained from CottonGen database [31] (<https://www.cottongen.org/>, accessed on 15 September 2021) (*Gossypium hirsutum*: NAU-NBI_v1.1_a1.1, *Gossypium barbadense*: ZJU_v1.1_a1, *Gossypium arboreum*: CGP-BGI_v2_a1 and *Gossypium raimondii*: JGI_v2_a2.1). We identified SBT gene family members in the following three ways. First, we ran a genetic search on CottonGen (<https://www.cottongen.org/>, accessed on 30 September 2021) by typing “Subtilisin” into the search box. Second, downloaded Peptidase_S8 model from PFAM website and conducted HMMER search through HMMER software. Third, the *Arabidopsis thaliana* SBT protein sequences confirmed in the literature were aligned with cotton protein sequences by BLASTP, where the parameter of the e-value was less than 5. Gene family members were obtained by combining the above three methods. Molecular weight (MW) and theoretical isoelectric point (pI) were investigated with the TBtools [58]. Subcellular locations were predicted by online tools CELLO [59] (<http://cello.life.nctu.edu.tw/>, accessed on 3 March 2022). The domain of SBT gene family member was confirmed by NCBI-CDD [60] (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>, accessed on 3 March 2022), and the e-value was less than 1.

4.2. Sequence Alignment and Phylogenetic Analysis

The SBT gene family members of four cotton species and *Arabidopsis thaliana* were constructed using Muscle 5.1. Osx64 (muscle -super5 input.fa -output aln.afa) for multi-sequence comparison of only domain proteins, and the tree was constructed with the maximum likelihood method (model: VT+I+I+R4, predicted by Modelfinder) IQtree2.2 using domain protein sequence muscle comparison [61,62].

Phylogenetic tree of gene structure was built by FastTree [63]. *Gossypium hirsutum* and *Gossypium arboreum* used WAG model and *Gossypium barbadense* and *Gossypium raimondii* used JTT model.

4.3. Cotton Gene Structure, Location Display, Cotton Collinearity Analysis, Calculation of Selection Pressure for Duplicated Gene Pairs

TBtools [58] was used to display the introns and exons of gene structure. The location information of all confirmed cotton SBT, including the starting position, located chromosomes and chromosome length, was obtained from CottonGen [31] (<https://www.cottongen.org/>, accessed on 15 September 2021) database. The evolutionary tree was constructed by FastTree. For the other cotton species, protein sequence alignment was completed. The cotton protein sequence was BLASTP, and the parameter e-value E-5 and max target seqs was 5. With MCSanX [64] software, use parameters MATCH_SCORE: 50, MATCH_SIZE: 0, GAP_PENALTY: 0, OVERLAP_WINDOW: 0, E_VALUE: 1, MAX GAPS:

25. Collinearity analysis was performed on four cotton species, and the visual collinearity display was completed by TBtools of unlimited synteny visualization. Circos was made TBtools advanced circos. The input file of the circos graph is obtained by modules “GXF gene position & Info. extract” and “Text Transformat for Micro-Synteny View”. Ka/Ks analysis was performed using TBtools’ Simple Ka/Ks calculator.

4.4. Co-Expression Network Analysis and Functional Enrichment Analysis

Cotton transcriptomic data and gene co-expression relationship were obtained from ccNET database [32], and *Arabidopsis thaliana* co-expression gene relationship was obtained from ATTED-II [65] (<http://atted.jp/>, accessed on 8 June 2022) database, with parameters of M.c7.1 and MR30. Amount of gene expression data from paper [66], cytoscape drawing co-express network, the differentially expressed genes were the log₂-fold change in FPKM value between treatment and WT (NaCl/CK or PEG/CK) ≤ -1 or ≥ 1 . The obtained cotton co-expressed genes were queried by CottonGen [31] (<https://www.cottongen.org/>, accessed on 15 September 2021) database to obtain *Arabidopsis thaliana* lineal homologous genes, and enrichment analysis and SEACOMPARE analysis were performed on agriGOV2 [40] (<http://systemsbiology.cau.edu.cn/agriGOv2/index.php>, accessed on 10 June 2022) platform. The most detailed entry in the DAG was selected for SEACOMPARE analysis.

4.5. Three-Dimensional Structure Prediction for SBT Protein

Firstly, the protein sequence was BLASTp (<https://www.uniprot.org/blast>, accessed on 8 January 2023) on the UniProt website to find the protein ID number with the highest score of the corresponding species, where the parameter target database was UniProtKB with 3D structure predictions (alphafold2), and the rest parameters were default. Then enter the obtained protein ID into the alphafold2 website (<https://alphafold.ebi.ac.uk>, accessed on 8 January 2023) [67,68] and download the PDB file. Finally, the three-dimensional structure of the protein was visualized on pymol.

5. Conclusions

In this study, 146 *Gossypium hirsutum*, 138 *Gossypium barbadense*, 89 *Gossypium arboreum* and 84 *Gossypium raimondii* were identified by phylogenetic analysis and domain analysis, and they were divided into six subfamilies. Gene structure analysis of each subfamily showed that most SBT1 gene family members of four cotton varieties did not have introns. The distribution of the SBT gene family members on the chromosomes of the four cotton species was found to be uneven. Through co-expression network analysis, Cotton_A_03171, Cotton_A_03377, Cotton_A_35816, Cotton_A_08556, Cotton_A_18889 and Cotton_A_07447 were found to be in the same network. Most of the genes in the roots, stems and leaves of *Gossypium arboreum* were down-regulated after salt and drought treatment. This network is involved in root tissue development and the response to salt stress in cotton. Co-expressed genes of *Gossypium hirsutum* and *Arabidopsis thaliana* were mostly down-regulated after salt treatment, indicating that the networks were functionally conserved. We predicted that these five *Gossypium arboreum* SBT1 and their corresponding *Gossypium hirsutum* ortholog genes responded to salt stress. This study provides a basis for understanding the biological functions of the SBT family genes in cotton under breeding and abiotic stress conditions. It provides a reference for mining the key genes for breeders to provide candidate genes related to salt and alkali resistance among many genes, and reduce the time taken for breeding varieties.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24065760/s1>.

Author Contributions: Conceptualization, Z.S. and W.X.; methodology, T.X., L.L., X.G. and Z.S.; software, T.X., Z.L. and Z.S.; formal analysis, T.X., W.X., X.Z., M.S. and Z.S.; data curation, Z.S., W.X., X.Z., M.S., L.L. and T.X.; writing—review and editing, T.X., Z.S., W.X. and X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was sponsored by the State Key Laboratory of Cotton Biology Open Fund (CB2021B04).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and materials that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: Thanks to Qunlian Zhang for her help in cotton planting.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vierstra, R.D. The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 385–397. [[CrossRef](#)] [[PubMed](#)]
2. Schaller, A.; Stintzi, A.; Rivas, S.; Serrano, I.; Chichkova, N.V.; Vartapetian, A.B.; Martínez, D.; Guiamét, J.J.; Sueldo, D.J.; van der Hoorn, R.; et al. From structure to function—A family portrait of plant subtilases. *New Phytol.* **2018**, *218*, 901–915. [[CrossRef](#)] [[PubMed](#)]
3. Schaller, A.; Stintzi, A.; Graff, L. Subtilases—Versatile tools for protein turnover, plant development, and interactions with the environment. *Physiol. Plant* **2012**, *145*, 52–66. [[CrossRef](#)] [[PubMed](#)]
4. Dodson, G.; Wlodawer, A. Catalytic triads and their relatives. *Trends Biochem. Sci.* **1998**, *23*, 347–352. [[CrossRef](#)]
5. Jordá, L.; Coego, A.; Conejero, V.; Vera, P. A genomic cluster containing four differentially regulated subtilisin-like processing protease genes is in tomato plants. *J. Biol. Chem.* **1999**, *274*, 2360–2365. [[CrossRef](#)]
6. Mahon, P.; Bateman, A. The PA domain: A protease-associated domain. *Protein Sci.* **2000**, *9*, 1930–1934. [[CrossRef](#)]
7. Murayama, K.; Kato-Murayama, M.; Hosaka, T.; Sotokawauchi, A.; Yokoyama, S.; Arima, K.; Shirouzu, M. Crystal structure of cucumisin, a subtilisin-like endoprotease from *Cucumis melo* L. *J. Mol. Biol.* **2012**, *423*, 386–396. [[CrossRef](#)]
8. Cedzich, A.; Huttenlocher, F.; Kuhn, B.M.; Pfannstiel, J.; Gabler, L.; Stintzi, A.; Schaller, A. The protease-associated (PA) domain and C-terminal extension are required for zymogen processing, sorting within the secretory pathway, and activity of tomato subtilase 3 (SISBT3). *J. Biol. Chem.* **2009**, *284*, 14068–14078. [[CrossRef](#)]
9. Rautengarten, C.; Steinhäuser, D.; Büssis, D.; Stintzi, A.; Schaller, A.; Kopka, J.; Altmann, T. Inferring hypotheses on functional relationships of genes: Analysis of the *Arabidopsis thaliana* subtilase gene family. *PLoS Comput. Biol.* **2005**, *1*, e40. [[CrossRef](#)]
10. Berger, D.; Altmann, T. A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in *Arab. Thaliana*. *Genes Dev.* **2000**, *14*, 1119–1131. [[CrossRef](#)]
11. Von Groll, U.; Berger, D.; Altmann, T. The subtilisin-like serine protease SDD1 mediates cell-to-cell signaling during Arabidopsis stomatal development. *Plant Cell* **2002**, *14*, 1527–1539. [[CrossRef](#)] [[PubMed](#)]
12. Wang, Q.; Guo, Q.; Guo, Y.; Yang, J.; Wang, M.; Duan, X.; Niu, J.; Liu, S.; Zhang, J.; Lu, Y.; et al. Arabidopsis subtilase SASP is involved in the regulation of ABA signaling and drought tolerance by interacting with OPEN STOMATA 1. *J. Exp. Bot.* **2018**, *69*, 4403–4417. [[CrossRef](#)] [[PubMed](#)]
13. Tanaka, H.; Onouchi, H.; Kondo, M.; Hara-Nishimura, I.; Nishimura, M.; Machida, C.; Machida, Y. A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. *Development* **2001**, *128*, 4681–4689. [[CrossRef](#)] [[PubMed](#)]
14. Xing, Q.; Creff, A.; Waters, A.; Tanaka, H.; Goodrich, J.; Ingram, G.C. ZHOUP1 controls embryonic cuticle formation via a signalling pathway involving the subtilisin protease ABNORMAL LEAF-SHAPE1 and the receptor kinases GASSHO1 and GASSHO2. *Development* **2013**, *140*, 770–779. [[CrossRef](#)]
15. Liu, J.X.; Srivastava, R.; Howell, S. Overexpression of an Arabidopsis gene encoding a subtilase (AtSBT5.4) produces a clavata-like phenotype. *Planta* **2009**, *230*, 687–697. [[CrossRef](#)]
16. Neuteboom, L.W.; Veth-Tello, L.M.; Clijdesdale, O.R.; Hooykaas, P.J.; van der Zaal, B.J. A novel subtilisin-like protease gene from *Arabidopsis thaliana* is expressed at sites of lateral root emergence. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* **1999**, *6*, 13–19. [[CrossRef](#)]
17. Rautengarten, C.; Usadel, B.; Neumetzler, L.; Hartmann, J.; Büssis, D.; Altmann, T. A subtilisin-like serine protease essential for mucilage release from Arabidopsis seed coats. *Plant J. Cell Mol. Biol.* **2008**, *54*, 466–480. [[CrossRef](#)]

18. Zhao, C.; Johnson, B.J.; Kositsup, B.; Beers, E.P. Exploiting secondary growth in Arabidopsis. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. *Plant Physiol.* **2000**, *123*, 1185–1196. [[CrossRef](#)]
19. Sénéchal, F.; Graff, L.; Surcouf, O.; Marcelo, P.; Rayon, C.; Bouton, S.; Mareck, A.; Mouille, G.; Stintzi, A.; Höfte, H.; et al. Arabidopsis PECTIN METHYLESTERASE17 is co-expressed with and processed by SBT3.5, a subtilisin-like serine protease. *Ann. Bot.* **2014**, *114*, 1161–1175. [[CrossRef](#)]
20. Liu, J.X.; Srivastava, R.; Che, P.; Howell, S.H. An endoplasmic reticulum stress response in Arabidopsis is mediated by proteolytic processing and nuclear relocation of a membrane-associated transcription factor, bZIP28. *Plant Cell* **2007**, *19*, 4111–4119. [[CrossRef](#)]
21. Liu, J.X.; Srivastava, R.; Che, P.; Howell, S.H. Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *Plant J.* **2007**, *51*, 897–909. [[CrossRef](#)]
22. Ghorbani, S.; Hoogewijs, K.; Pečenková, T.; Fernandez, A.; Inzé, A.; Eeckhout, D.; Kawa, D.; De Jaeger, G.; Beeckman, T.; Madder, A.; et al. The SBT6.1 subtilase processes the GOLVEN1 peptide controlling cell elongation. *J. Exp. Bot.* **2016**, *67*, 4877–4887. [[CrossRef](#)]
23. Srivastava, R.; Liu, J.X.; Guo, H.; Yin, Y.; Howell, S.H. Regulation and processing of a plant peptide hormone, AtRALF23, in Arabidopsis. *Plant J.* **2009**, *59*, 930–939. [[CrossRef](#)] [[PubMed](#)]
24. Ramírez, V.; López, A.; Mauch-Mani, B.; Gil, M.J.; Vera, P. An extracellular subtilase switch for immune priming in Arabidopsis. *PLoS Pathog.* **2013**, *9*, e1003445. [[CrossRef](#)] [[PubMed](#)]
25. Duan, X.; Zhang, Z.; Wang, J.; Zuo, K. Characterization of a Novel Cotton Subtilase Gene GbSBT1 in Response to Extracellular Stimulations and Its Role in Verticillium Resistance. *PLoS ONE* **2016**, *11*, e0153988. [[CrossRef](#)] [[PubMed](#)]
26. Wendel, J.F.; Albert, V.A. Phylogenetics of the Cotton Genus (*Gossypium*): Character-State Weighted Parsimony Analysis of Chloroplast-DNA Restriction Site Data and Its Systematic and Biogeographic Implications. *Syst. Bot.* **1992**, *17*, 115–143. [[CrossRef](#)]
27. Sunilkumar, G.; Campbell, L.M.; Puckhaber, L.; Stipanovic, R.D.; Rathore, K.S. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18054–18059. [[CrossRef](#)]
28. Wendel, J.F. Genome evolution in polyploids. *Plant Mol. Biol.* **2000**, *42*, 225–249. [[CrossRef](#)]
29. Blanc, G.; Wolfe, K.H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **2004**, *16*, 1667–1678. [[CrossRef](#)]
30. Liu, Z.; Adams, K.L. Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Curr. Biol.* **2007**, *17*, 1669–1674. [[CrossRef](#)]
31. Yu, J.; Jung, S.; Cheng, C.H.; Lee, T.; Zheng, P.; Buble, K.; Crabb, J.; Humann, J.; Hough, H.; Jones, D.; et al. CottonGen: The Community Database for Cotton Genomics, Genetics, and Breeding Research. *Plants* **2021**, *10*, 2805. [[CrossRef](#)] [[PubMed](#)]
32. You, Q.; Xu, W.; Zhang, K.; Zhang, L.; Yi, X.; Yao, D.; Wang, C.; Zhang, X.; Zhao, X.; Provar, N.J.; et al. ccNET: Database of co-expression networks with functional modules for diploid and polyploid *Gossypium*. *Nucleic Acids Res.* **2017**, *45*, D1090–D1099. [[CrossRef](#)] [[PubMed](#)]
33. Tripathi, L.P.; Sowdhamini, R. Cross genome comparisons of serine proteases in Arabidopsis and rice. *BMC Genom.* **2006**, *7*, 200. [[CrossRef](#)] [[PubMed](#)]
34. Norero, N.S.; Castellote, M.A.; de la Canal, L.; Sergio, E.F. Genome-Wide Analyses of Subtilisin-Like Serine Proteases on *Solanum tuberosum*. *Am. J. Potato Res.* **2016**, *93*, 485–496. [[CrossRef](#)]
35. Jin, X.; Liu, Y.; Hou, Z.; Zhang, Y.; Fang, Y.; Huang, Y.; Cai, H.; Qin, Y.; Cheng, Y. Genome-Wide Investigation of SBT Family Genes in Pineapple and Functional Analysis of AcoSBT1.12 in Floral Transition. *Front. Genet.* **2021**, *12*, 730821. [[CrossRef](#)]
36. Cao, J.; Han, X.; Zhang, T.; Yang, Y.; Huang, J.; Hu, X. Genome-wide and molecular evolution analysis of the subtilase gene family in *Vitis vinifera*. *BMC Genom.* **2014**, *15*, 1116. [[CrossRef](#)]
37. Figueiredo, J.; Costa, G.J.; Maia, M.; Paulo, O.S.; Malhó, R.; Sousa Silva, M.; Figueiredo, A. Revisiting *Vitis vinifera* Subtilase Gene Family: A Possible Role in Grapevine Resistance against *Plasmopara viticola*. *Front. Plant Sci.* **2016**, *7*, 1783. [[CrossRef](#)]
38. Xu, Y.; Wang, S.; Li, L.; Sahu, S.K.; Petersen, M.; Liu, X.; Melkonian, M.; Zhang, G.; Liu, H. Molecular evidence for origin, diversification and ancient gene duplication of plant subtilases (SBTs). *Sci. Rep.* **2019**, *9*, 12485. [[CrossRef](#)]
39. Zhang, T.; Hu, Y.; Jiang, W.; Fang, L.; Guan, X.; Chen, J.; Zhang, J.; Saski, C.A.; Scheffler, B.E.; Stelly, D.M.; et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **2015**, *33*, 531–537. [[CrossRef](#)]
40. Tian, T.; Liu, Y.; Yan, H.; You, Q.; Yi, X.; Du, Z.; Xu, W.; Su, Z. agriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **2017**, *45*, W122–W129. [[CrossRef](#)]
41. Chen, H.; Kong, Y.; Chen, J.; Li, L.; Li, X.; Yu, F.; Ming, Z. Crystal structure of the extracellular domain of the receptor-like kinase TMK3 from *Arabidopsis thaliana*. *Acta crystallographica. Sect. F Struct. Biol. Commun.* **2020**, *76*, 384–390. [[CrossRef](#)]
42. Julkowska, M.M.; Klei, K.; Fokkens, L.; Haring, M.A.; Schranz, M.E.; Testerink, C. Natural variation in rosette size under salt stress conditions corresponds to developmental differences between Arabidopsis accessions and allelic variation in the LRR-KISS gene. *J. Exp. Bot.* **2016**, *67*, 2127–2138. [[CrossRef](#)]
43. Kim, S.; Kim, S.J.; Shin, Y.J.; Kang, J.H.; Kim, M.R.; Nam, K.H.; Lee, M.S.; Lee, S.H.; Kim, Y.H.; Hong, S.K.; et al. An atypical soybean leucine-rich repeat receptor-like kinase, GmLRK1, may be involved in the regulation of cell elongation. *Planta* **2009**, *229*, 811–821. [[CrossRef](#)]
44. Clay, N.K.; Nelson, T. VH1, a provascular cell-specific receptor kinase that influences leaf cell patterns in Arabidopsis. *Plant Cell* **2002**, *14*, 2707–2722. [[CrossRef](#)]

45. Huang, C.; Zhang, R.; Gui, J.; Zhong, Y.; Li, L. The Receptor-Like Kinase AtVRLK1 Regulates Secondary Cell Wall Thickening. *Plant Physiol.* **2018**, *177*, 671–683. [[CrossRef](#)]
46. Wang, Y.; Nsibo, D.L.; Juhar, H.M.; Govers, F.; Bouwmeester, K. Ectopic expression of Arabidopsis L-type lectin receptor kinase genes LecRK-I.9 and LecRK-IX.1 in *Nicotiana benthamiana* confers *Phytophthora* resistance. *Plant Cell Rep.* **2016**, *35*, 845–855. [[CrossRef](#)]
47. Begum, T.; Reuter, R.; Schöfl, F. Overexpression of AtHsfB4 induces specific effects on root development of Arabidopsis. *Mech. Dev.* **2013**, *130*, 54–60. [[CrossRef](#)]
48. Hernández-Coronado, M.; Ortiz-Ramírez, C. Root Patterning: Tuning SHORT ROOT Function Creates Diversity in Form. *Front. Plant Sci.* **2021**, *12*, 745861. [[CrossRef](#)]
49. Moreno-Piovan, G.S.; Moreno, J.E.; Cabello, J.V.; Arce, A.L.; Otegui, M.E.; Chan, R.L. A role for LAX2 in regulating xylem development and lateral-vein symmetry in the leaf. *Ann. Bot.* **2017**, *120*, 577–590. [[CrossRef](#)]
50. Ranocha, P.; Dima, O.; Nagy, R.; Felten, J.; Corratgé-Faillie, C.; Novák, O.; Morreel, K.; Lacombe, B.; Martinez, Y.; Pfrunder, S.; et al. Arabidopsis WAT1 is a vacuolar auxin transport facilitator required for auxin homeostasis. *Nat. Commun.* **2013**, *4*, 2625. [[CrossRef](#)]
51. Prabhakaran Mariyamma, N.; Clarke, K.J.; Yu, H.; Wilton, E.E.; Van Dyk, J.; Hou, H.; Schultz, E.A. Members of the Arabidopsis FORKED1-LIKE gene family act to localize PIN1 in developing veins. *J. Exp. Bot.* **2018**, *69*, 4773–4790. [[CrossRef](#)]
52. Lewis, D.R.; Olex, A.L.; Lundy, S.R.; Turkett, W.H.; Fetrow, J.S.; Muday, G.K. A kinetic analysis of the auxin transcriptome reveals cell wall remodeling proteins that modulate lateral root development in Arabidopsis. *Plant Cell* **2013**, *25*, 3329–3346. [[CrossRef](#)]
53. Liu, W.; Li, R.J.; Han, T.T.; Cai, W.; Fu, Z.W.; Lu, Y.T. Salt stress reduces root meristem size by nitric oxide-mediated modulation of auxin accumulation and signaling in Arabidopsis. *Plant Physiol.* **2015**, *168*, 343–356. [[CrossRef](#)]
54. Ribba, T.; Garrido-Vargas, F.; O'Brien, J.A. Auxin-mediated responses under salt stress: From developmental regulation to biotechnological applications. *J. Exp. Bot.* **2020**, *71*, 3843–3853. [[CrossRef](#)]
55. Deslauriers, S.D.; Larsen, P.B. FERONIA is a key modulator of brassinosteroid and ethylene responsiveness in Arabidopsis hypocotyls. *Mol. Plant* **2010**, *3*, 626–640. [[CrossRef](#)] [[PubMed](#)]
56. Wolf, S.; Hématy, K.; Höfte, H. Growth control and cell wall signaling in plants. *Annu. Rev. Plant Biol.* **2012**, *63*, 381–407. [[CrossRef](#)]
57. Feng, W.; Kita, D.; Peaucelle, A.; Cartwright, H.N.; Doan, V.; Duan, Q.; Liu, M.C.; Maman, J.; Steinhorst, L.; Schmitz-Thom, I.; et al. The FERONIA Receptor Kinase Maintains Cell-Wall Integrity during Salt Stress through Ca²⁺ Signaling. *Curr. Biol.* **2018**, *28*, 666–675.e5. [[CrossRef](#)] [[PubMed](#)]
58. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **2020**, *13*, 1194–1202. [[CrossRef](#)] [[PubMed](#)]
59. Yu, C.S.; Lin, C.J.; Hwang, J.K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci. A Publ. Protein Soc.* **2004**, *13*, 1402–1406. [[CrossRef](#)]
60. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268. [[CrossRef](#)]
61. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [[CrossRef](#)] [[PubMed](#)]
62. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.; von Haeseler, A.; Jermini, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [[CrossRef](#)]
63. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)]
64. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic. Acids Res.* **2012**, *40*, e49. [[CrossRef](#)] [[PubMed](#)]
65. Obayashi, T.; Aoki, Y.; Tadaka, S.; Kagaya, Y.; Kinoshita, K. ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. *Plant Cell Physiol.* **2018**, *59*, e3. [[CrossRef](#)]
66. Yu, J.; Zhang, Y.; Di, C.; Zhang, Q.; Zhang, K.; Wang, C.; You, Q.; Yan, H.; Dai, S.Y.; Yuan, J.S. JAZ7 negatively regulates dark-induced leaf senescence in Arabidopsis. *J. Exp. Bot.* **2016**, *67*, 751–762. [[CrossRef](#)]
67. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)] [[PubMed](#)]
68. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.