agronomy

MDPI

Article

B-box Proteins in *Arachis duranensis*: Genome-Wide Characterization and Expression Profiles Analysis

Hanqi Jin †, Mengge Xing †, Chunmei Cai and Shuai Li *

College of Life Sciences, Key Lab of Plant Biotechnology in Universities of Shandong Province, Qingdao Agricultural University, Qingdao 266109, China; jinhanqi1@163.com (H.J.); xmg1024@163.com (M.X.); caichunmei0902@163.com (C.C.)

* Correspondence: li2014shuai@qau.edu.cn; Tel.: +86-0532-58957640

† These authors have contributed equally to this work.

Received: 5 November 2019; Accepted: 19 December 2019; Published: 23 December 2019

Abstract: B-box (BBX) proteins are important factors involved in plant growth and developmental regulation, and they have been identified in many species. However, information on the characteristics and transcription patterns of *BBX* genes in wild peanut are limited. In this study, we identified and characterized 24 *BBX* genes from a wild peanut, *Arachis duranensis*. Many characteristics were analyzed, including chromosomal locations, phylogenetic relationships, and gene structures. *Arachis duranensis* B-box (AdBBX) proteins were grouped into five classes based on the diversity of their conserved domains: I (3 genes), II (4 genes), III (4 genes), IV (9 genes), and V (4 genes). Fifteen distinct motifs were found in the 24 AdBBX proteins. Duplication analysis revealed the presence of two interchromosomal duplicated gene pairs, from group II and IV. In addition, 95 kinds of cis-acting elements were found in the genes' promoter regions, 53 of which received putative functional predictions. The numbers and types of *cis*-acting elements varied among different *AdBBX* promoters, and, as a result, *AdBBX* genes exhibited distinct expression patterns in different tissues. Transcriptional profiling combined with synteny analysis suggests that *AdBBX8* may be a key factor involved in flowering time regulation. Our study will provide essential information for further functional investigation of *AdBBX* genes.

Keywords: peanut; B-box (BBX); *Arachis duranensis*; flowering time; gene expression

1. Introduction

Transcription factors are essential components of signal transduction pathways in plants. They often work as activators or suppressors, binding cis-acting elements in target promoter regions to regulate downstream gene expression [1,2]. Various transcription factor families are found in plants, which are involved in many different response pathways. Among these gene families, zinc-finger transcription factors, possess a conserved domain that can bind metal ions like zinc and interact with DNA, RNA or proteins, and members of that large transcription factor family play critical roles in plant growth and development [3–6]. Based on the diversification of their protein sequences and structures, zinc-finger genes are further classified into several distinct subfamilies [3]. A subgroup of zinc-finger proteins containing B-box (BBX) conserved domains, which are thought to be involved in protein–protein interactions, is designated as the BBX family and is highly conserved across all multicellular species [4,7–9].

Two types of BBX domains, B-box1 and B-box2, have been identified based on their consensus sequences and the spacing of their zinc-binding residues [4,10–12]. In *Arabidopsis*, 21 and 11 of the 32 BBX proteins contain two and one BBX domains, respectively [3,4]. In addition to the conserved BBX domain, some BBX family members contain additional specific domains, such as the CCT (for

CONSTANS, CONSTANS-LIKE, TOC1) and VP (valine-proline) domains. Seventeen *Arabidopsis* BBX proteins contain a CCT domain close to their C terminus. *Arabidopsis* BBX proteins are grouped into five classes based on differences in the type and number of BBX and CCT domains. Classes I and II members have two BBX domains and one CCT domain, while class III members have one BBX and one CCT domain. Class IV members have two different BBX domains, and group V members have only one BBX domain [3,4,13].

BBX proteins are involved in the regulation of many signaling pathways, including those associated with flowering time, the circadian clock, seedling photomorphogenesis, and abiotic stress [4]. CONSTANS (CO/BBX1) was the first BBX protein identified in *Arabidopsis*. Overexpression of *CO* accelerates flowering time under both long day (LD) and short day (SD) conditions. Mutation of *CO* significantly delays flowering time under LD, while the flowering time of *CO* is similar to wild-type plants under SD [14–16]. CO binds directly to *CO*-responsive elements (CORE) and CCAAT-box elements in the promoter region of *FLOWERING LOCUS T* (*FT*) to promote *FT* expression, which is in turn responsible for the acceleration of flowering time [17–21]. The *CO-FT* flowering time regulatory pathway is conserved in other species [22–31]. For example, *Heading date 1* (*Hd1*), the rice *CO* ortholog, participates in the regulation of the rice *FT* orthologs *Heading date 3a (OsHd3a)* and *Flowering Locus T1 (OsRFT1)* [24,25]. *Hd1* promotes flowering time under SD but delays flowering time under LD [22]. In addition, some CO-like (COL) proteins in *Arabidopsis* are also involved in flowering time or circadian clock regulation, including *BBX2*, *BBX3*, *BBX4*, *BBX6*, and *BBX7*. *BBX18*, *BBX24*, and *BBX32* are also thought to be involved in abiotic and biotic stress responses [4].

Peanut is an important oil crop widely cultivated in many countries. The allotetraploid *Arachis hypogaea* (AABB, $2n = 4x = 40$) is thought to be derived from the hybridization and polyploidization of two diploid peanuts, *Arachis duranensis* (AA genome) and *Arachis ipaensis* (BB genome) [26,32,33]. Many genes in cultivated peanut might retain some functions of the original gene in wild peanut, and the investigation of wild peanut genes will provide useful information for further analysis of cultivated peanut gene functions. The development of the peanut genome database in recent years now permits investigation of peanut gene functions [33–37]. Although there is considerable information on *BBX* gene function in other species, there is less research on the role of *BBX* genes in peanut development. In this study, we identified and characterized 24 *BBX* proteins from one of the wild peanut species, *A. duranensis*. We investigated many characteristics of its *BBX* genes, including structures, conserved motifs, chromosome locations, phylogenetic relationships, gene duplications, cis-acting elements, and tissue expression profiles. Our study provides essential information for further functional characterization of *BBX* proteins in wild and cultivated peanut.

2. Materials and Methods

2.1. Identification of BBX Members in A. duranensis

The amino acid sequences of the BBX conserved domain (PF00643) and *Arabidopsis* BBX proteins, downloaded from Arabidopsis genome database TAIR, were used as blast queries against the peanut genome database to search for *A. duranensis*, *Arachis ipaensis*, and *Arachis hypogaea BBX* genes (Peanut genome database version: *Arachis hypogaea* Tifrunner 1.0, *Arachis duranensis* 1.0 and *Arachis ipaensis* 1.0) [33]. All output candidate genes were assessed using the Pfam database (Pfam 32.0, September 2018, 17,929 entries) [38] and the InterPro program [39] to confirm the presence of BBX domains and remove genes without conserved BBX domains. The B-box1 and B-box2 domains were determined as described by Crocco et al. [12]. The first B-box that appeared in the N terminal position was called B-box1 and the second was termed B-box2. The positions of the BBX and CCT domains in each *Arachis duranensis* B-box (AdBBX) protein were analyzed with Pro Scan program. The genomic lengths, coding sequence (CDS) lengths, and amino acid numbers of the *AdBBX* genes were obtained from the peanut genome database. The gene GC contents were determined using MegAlign software

of Lasergene (DNASTAR, Madison, WI, USA). Chemical features of the AdBBX proteins, including molecular weight and theoretical iso-electric point, were obtained using ProtParam [40].

2.2. Sequence Alignment and Phylogenetic Relationship Analyses

The amino acid sequences of BBX and CCT domains were aligned with ClustalW2 software (2.1, Nanyang Technological University, Singapore) [41], and the results were used to generate an alignment map using DNAMAN software (Version 5.2.2, LinnonBiosoft, Vaudreuil, Quebec, Canada). To analyze the evolutionary relationships among the *AdBBX* genes and other well-studied *BBX* genes, full-length BBX amino acid sequences from *A. duranensis*, rice, and *Arabidopsis* were aligned by Clustal-X2, and the alignment results were used to construct a phylogenetic tree in MEGA7.0 using the neighbor-joining method [42]. The BBX proteins from *A. duranensis*, *Arachis ipaensis* and *Arachis hypogaea* were also aligned to construct a phylogenetic tree in MEGA7.0 using the neighbor-joining method.

2.3. Gene Structure, Conserved Motif, and Sequence Logo Analyses

The genomic and CDS sequences of the *AdBBX* genes were downloaded from the peanut genome database [33] and used for the analysis of gene structure using the Gene Structure Display Server (GSDS) [43]. To investigate conserved motifs in these proteins, full-length AdBBX amino acid sequences were obtained from the peanut genome database and analyzed using MEME tools [44]. In addition, the sequence logos of BBX (B-box1 and B-box2) and CCT conserved domains were investigated using the online WebLogo platform [45].

2.4. Chromosomal Location, Synteny, and Gene Duplication Analyses

AdBBX genes were mapped to the peanut genome to identify their chromosomal locations and physical positions. A chromosomal distribution map was generated using Mapinspect software (Mike Lischke, Berlin, Germany). To investigate the synteny relationships of related genome regions in different species, putative orthologous genes surrounding *CO* orthologs/homologs from soybean and *A. duranensis* were identified by a BLASTP search as previously described [46,47], and the blast results were used to generate a synteny map. To analyze the duplicated gene pairs, we clustered *AdBBX*, *AiBBX*, and *AhBBX* genes using OrthoMCL software [48] with amino acid sequences more than 70% similar, and visualized the duplicated gene pairs using Circos software as previously described, respectively [49,50].

2.5. Cis-Acting Element Analysis

Promoter sequences, 2 kb upstream of the initiation codon of each gene, were obtained from the peanut genome database [33] and used for cis-acting element analysis using PlantCARE [51]. The *cis*-acting elements were classified based on their putative functions.

2.6. Analysis of Gene Expression in Different Tissues

To investigate the expression of *AdBBX* genes in different *A. duranensis* tissues, RNA-seq datasets were downloaded from the peanut genome database [33–37,52,53] and used for transcription analysis. The *A. hypogaea* gene expression profiles were mapped to *A. duranensis* for the construction of a heatmap with MeV 4.9.0 (Multiple Experiment Viewer) [52,53].

3. Results

3.1. Identification of BBX Genes in A. duranensis

To identify *BBX* genes in *A. duranensis*, we used the amino acid sequences of the BBX conserved domain (PF00643) and 32 *Arabidopsis* BBXs as blast queries against the peanut genome database. We then used Pfam and Pro Scan to confirm the presence of conserved BBX domains in the resulting candidate genes. A total of 24 *BBX* genes were found in *A. duranensis*. The characteristics of these *AdBBX*

genes were investigated, and detailed information is provided in Table 1. *AdBBX* genes exhibited diversity in genomic length, CDS length, amino acid number, isoelectric point, molecular weight, and GC content (Table 1). The genomic lengths of *AdBBX* genes ranged from 468 bp (*AraduF08JS*) to 4,564 bp (*AraduBV95P*), the CDS lengths ranged from 468 bp (*AraduF08JS*) to 1,641 bp (*AraduVV0J1*), and the numbers of deduced amino acids ranged from 155 to 546 (Table 1). GC content represents the stability of the genes to some degree, and we therefore investigated GC content of the *AdBBX* genes. The GC contents in *AdBBX* genes ranged from 32.13% (*Aradu1V7PF*) to 52.72% (*Aradu28KTI*). The isoelectric points of the proteins varied from 4.16 (AraduF08JS) to 8.85 (Aradu28KTI), and their molecular weights ranged from 17,020.4 (AraduF08JS) to 61,012.7 (AraduVV0J1).

Table 1. Identification of *Arachis duranensis* B-box *(AdBBX*) members.

Gene ID	Genomic Length/bp	CDS Length/bp	No.of AA	GC%	pI	Mol. Wt (Da)	Strand	Chr	5′ End	3′ End	Gene Name
AraduV9D7S	1837	1098	365	48.89%	6.59	39,720.5	+	A01	91,890,657	91,892,493	*AdBBX1*
Aradu7ZG0H	2519	735	244	35.03%	5.09	26,830.5	−	A03	3,490,043	3,492,561	*AdBBX2*
Aradu7V6T8	1535	651	216	36.81%	5.63	23,896.5	+	A03	4,936,073	4,937,607	*AdBBX3*
Aradu1V7PF	2390	639	212	32.13%	5.83	23,720.7	−	A03	5,102,892	5,105,281	*AdBBX4*
AraduJ5IAH	3841	1209	402	33.66%	4.98	44,409.8	+	A03	15,759,939	15,763,779	*AdBBX5*
Aradu8CI8S	915	612	203	47.98%	4.48	22,189.3	+	A03	16,814,094	16,815,008	*AdBBX6*
AraduWJ2ZP	1792	1293	430	43.81%	5.48	48,100.6	+	A03	33,948,193	33,949,984	*AdBBX7*
Aradu43J56	1894	933	310	40.60%	7.12	33,922.1	−	A04	120,041,625	120,043,518	*AdBBX8*
Aradu3ZR52	2253	570	189	37.28%	6.23	20,945.9	−	A05	11,384,528	11,386,780	*AdBBX9*
AraduXPS1Y	1674	1200	399	37.46%	5.5	45,440.5	+	A05	93,078,336	93,080,009	*AdBBX10*
AraduHU5GE	2523	1113	370	38.13%	6.03	41,278.9	+	A06	8,755,561	8,758,083	*AdBBX11*
Aradu23R92	1711	900	299	39.22%	6.38	32,697.9	+	A06	16,881,156	16,882,866	*AdBBX12*
AraduJ7TMC	4164	1140	379	37.58%	6.57	41,756.3	+	A06	94,914,152	94,918,315	*AdBBX13*
AraduBV95P	4564	1248	415	37.12%	5.14	45,093.2	−	A07	27,306,397	27,310,960	*AdBBX14*
AraduWL99W	1873	696	231	39.14%	6.77	25,455.6	−	A07	56,363,517	56,365,389	*AdBBX15*
AraduM999T	2300	903	300	39.78%	4.98	32,359.4	−	A07	75,340,469	75,342,768	*AdBBX16*
AraduVV0JI	3222	1641	546	38.08%	6.44	61,012.7	+	A08	23,797,353	23,800,574	*AdBBX17*
AraduHJ8Q1	1249	1065	354	40.30%	4.61	39,302.6	+	A08	29,919,055	29,920,303	*AdBBX18*
Aradu28KTI	882	882	293	52.72%	8.85	32,390.3	+	A09	592,000	592,881	*AdBBX19*
AraduZT2KF	4052	1389	462	37.66%	5.03	51,996.2	−	A09	113,442,542	113,446,593	*AdBBX20*
AraduJ9KV2	1368	1272	423	46.78%	6.24	47,360.9	−	A09	114,055,561	114,056,928	*AdBBX21*
AraduF08JS	468	468	155	42.52%	4.16	17,020.4	−	A10	3,740,421	3,740,888	*AdBBX22*
Aradu5RF5F	1287	984	327	51.83%	6.51	35,690.8	−	A10	88,380,486	88,381,772	*AdBBX23*
AraduS156I	2354	1185	394	35.39%	4.81	43,580	+	A10	104,730,376	10,473,2729	*AdBBX24*

Chr, chromosome number; CDS, coding sequence; AA, amino acid; pI, isoelectric point; Mol. Wt, molecular weight.

3.2. Chromosomal Distribution of AdBBXs

To investigate the chromosomal locations of *AdBBXs*, we mapped the genes to the peanut genome to determine their physical positions. The *AdBBX* genes were named from *AdBBX1* to *AdBBX24* based on their chromosomal distribution (Figure 1, Table 1). Among the 24 *AdBBX* genes, 13 were located on the plus strand and 11 were located on the minus strand (Table 1). Nine of the 10 peanut chromosomes contained *AdBBX* genes, and only chromosome 2 contained none (Figure 1 and Table 1). Chromosome 3 contained the largest number of *AdBBX* genes, with six *AdBBX* members, followed by chromosomes 6, 7, 9, and 10, with three *AdBBXs* on each. *AdBBX2*, *AdBBX3*, and *AdBBX4* were located close together on chromosome 3 (the distance between *AdBBX2* and *AdBBX3* was 1,443,512 bp and the distance between *AdBBX3* and *AdBBX4* was 165,285 bp), as were *AdBBX5* and *AdBBX6* (the distance between *AdBBX5* and *AdBBX6* was 1,050,315 bp). *AdBBX20* and *AdBBX21* (the distance between them was 608,968 bp) were also located close together on chromosome 9. Most *AdBBX* genes were located on the chromosome arms, but three (*AdBBX14*, *AdBBX17*, and *AdBBX18*) were found on the middle portions of the chromosomes.

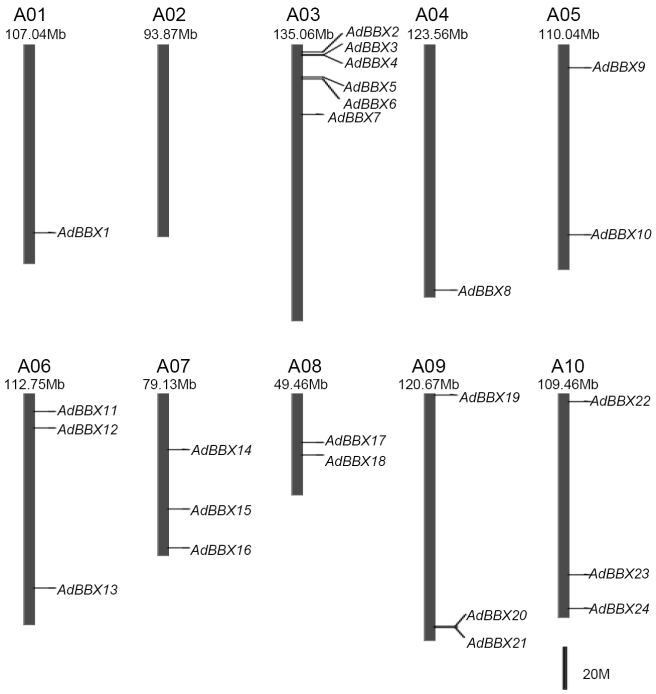


Figure 1. Chromosomal distribution of *AdBBX* genes. Chromosome number and length are shown.

3.3. Protein Sequence and Classification Analysis of AdBBX Genes

BBX proteins are classified into five subgroups based their conserved domains, including the types and numbers of BBX and CCT domains [3,4]. We analyzed conserved domains in the AdBBX proteins and found two distinct BBX domains (B-box1 and B-box2) and one CCT domain. To further investigate the conserved amino acid sequences in these domains, logos of *A. duranensis* B-box1 ($CX_2CX_8CX_2DXAXLCX_2CDX_2VHX_2NXLX_3H$, where X represents any amino acid), B-box2 ($CX_2CX_4AX_3CX_7CX_2CDX_3HX_9H$), and CCT ($RX_5RX_3KX_7KX_2RYX_2RKX_2AX_2RXRXKGRFXK$) were produced using Weblogo [45] (Figure 2). The amino acid sequences of the B-box1, B-box2, and CCT domains were also aligned to analyze corresponding positions of the conserved amino acid sequences (Figure 3).

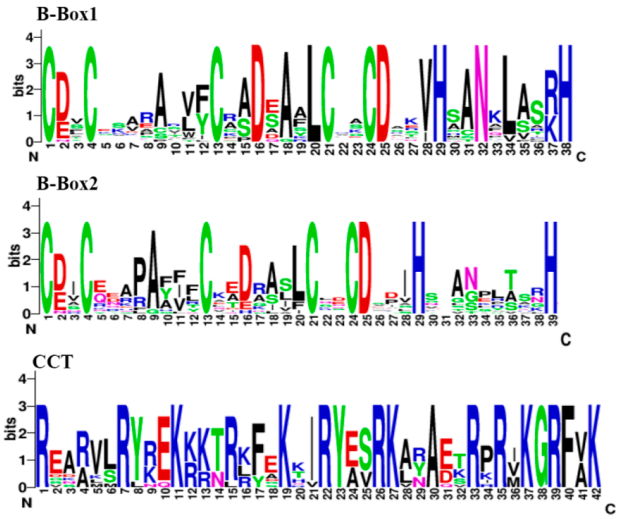


Figure 2. The motif logos of B-box1, B-box2, and CCT (CONSTANS, CONSTANS-LIKE, TOC1) domains. The *x*-axis indicates the conserved sequences of these domains, and the height of each letter indicates the conservation of each residue.

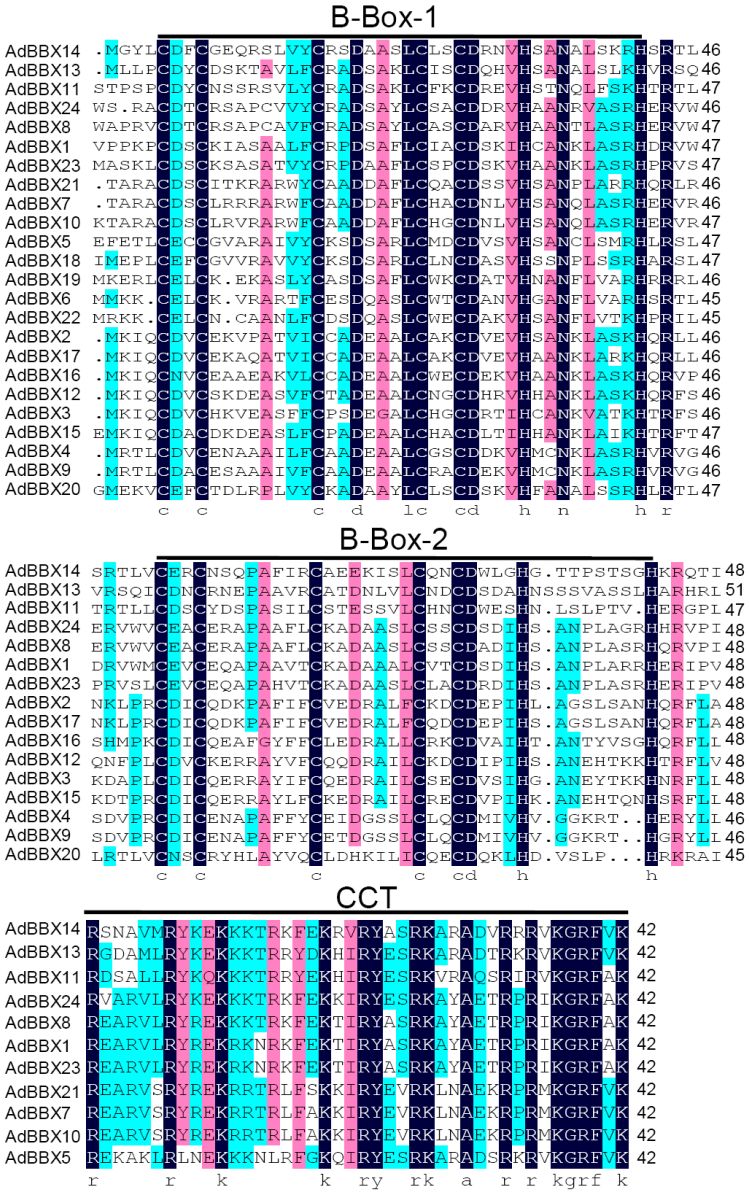
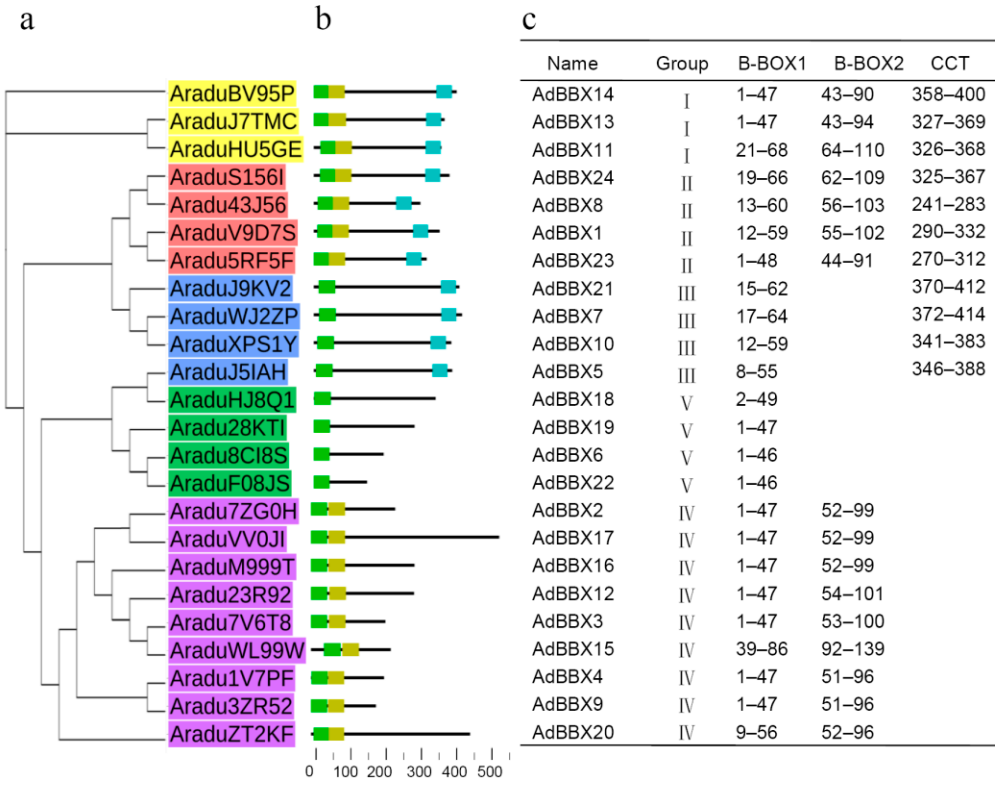


Figure 3. Multiple sequence alignments of B-box1, B-box2, and CCT domains. The amino acid sequences of the conserved B-box1, B-box2, and CCT domains were aligned, and identical and similar amino acids are shaded.

To further investigate the evolutionary relationships among the *AdBBXs*, we created a phylogenetic tree based on their amino acid sequences (Figure 4a). Conserved domain analysis revealed that 16 of the 24 AdBBX proteins had two BBX domains (i.e., B-box1 and B-box2), 11 members contained a CCT domain, and seven had both BBX domains and a CCT domain (Figure 4b). The AdBBX proteins were also grouped into five subfamilies based on the diversity of their conserved domains (Figure 4c). Group I and II, which differed in the details of their B-box2 domain, had a B-box1, B-box2, and CCT domain. There were three and four members of classes I and II, respectively, in *A. duranensis*. Class III had a B-box1 and a CCT domain and contained four members. Group IV had both a B-box1 and B-box2 domain and was the largest group in *A. duranensis* with nine members. Most *AdBBX* genes in the same class were clustered together, except for *AdBBX5* (*AraduJ5IAH*), which had a distinct relationship with other class III genes and a closer relationship with class V subfamily members (Figure 4a). However, the number of *BBX* genes from *A. ipaensis* and *A. hypogaea* in each group showed differences from *A. duranensis*. Group I, II, III, IV, and V contained 2, 4, 5, 9, and 5 members in *A. ipaensis*, and 1, 2, 10,

8, and 19 in *A. hypogaea*, respectively (Figures S1 and S2). To analyze the evolutional relationship of peanut *BBX* genes, a phylogenetic tree was constructed using *BBX* genes from *A. duranensis*, *A. ipaensis*, and *A. hypogaea* (Figure S3). Most of the genes from *A. duranensis* and *A. ipaensis* in each group were clustered into the same clades, while many genes from *A. hypogaea* in the same group were not clustered together, indicating *BBX* genes in cultivated peanut might have changed a lot during evolution.



Name	Group	B-BOX1	B-BOX2	CCT
AdBBX14	I	1–47	43–90	358–400
AdBBX13	I	1–47	43–94	327–369
AdBBX11	I	21–68	64–110	326–368
AdBBX24	II	19–66	62–109	325–367
AdBBX8	II	13–60	56–103	241–283
AdBBX1	II	12–59	55–102	290–332
AdBBX23	II	1–48	44–91	270–312
AdBBX21	III	15–62		370–412
AdBBX7	III	17–64		372–414
AdBBX10	III	12–59		341–383
AdBBX5	III	8–55		346–388
AdBBX18	V	2–49		
AdBBX19	V	1–47		
AdBBX6	V	1–46		
AdBBX22	V	1–46		
AdBBX2	IV	1–47	52–99	
AdBBX17	IV	1–47	52–99	
AdBBX16	IV	1–47	52–99	
AdBBX12	IV	1–47	54–101	
AdBBX3	IV	1–47	53–100	
AdBBX15	IV	39–86	92–139	
AdBBX4	IV	1–47	51–96	
AdBBX9	IV	1–47	51–96	
AdBBX20	IV	9–56	52–96	

Figure 4. Phylogenetic and structure analyses of AdBBX proteins. (**a**) A phylogenetic tree was generated from protein sequences in MEGA 7 using the neighbor-joining method. The yellow, red, blue, purple, and green gene names indicate class I, II, III, IV, and V members, respectively. (**b**) The positions of conserved domains in each AdBBX protein are represented and the green, pale brown, and cyan boxes indicate B-box1, B-box2, and CCT domains, respectively. (**c**) The classification of AdBBX members and the position of each domain.

To obtain information from well-studied *BBX* genes in other species, we also analyzed the evolutionary relationships among *BBX* genes from *Arabidopsis*, rice, and *A. duranensis* (Figure S4). Phylogenetic analysis revealed that *AdBBX8* and *AdBBX24* grouped with the well-studied flowering time genes *At5g15840* (*CO*) and *Os06g16370* (*Hd1*) from *Arabidopsis* and rice, respectively [4], indicating that one or both of these genes may play an important role in flowering time regulation in *A. duranensis*.

3.4. Gene Structures and Conserved Motifs of AdBBX Genes

To investigate exon–intron organization, the genomic and CDS sequences of *AdBBXs* obtained from the peanut genome database were analyzed using the Gene Structure Display Server [43]. The exon numbers of *AdBBX* genes ranged from one (*AdBBX19* and *AdBBX22*) to five (*AdBBX4*), and the intron numbers ranged from zero (*AdBBX19* and *AdBBX22*) to six (*AdBBX4* and *AdBBX5*). Nine *AdBBX* genes had both 5′ and 3′ untranslated regions (UTRs), four genes had only 3′ UTRs, three genes had only 5′ UTRs, and eight genes had no predicted UTRs (Figure 5a). To further investigate the conservation and diversity of AdBBX protein structures, the putative motifs of these genes were predicted using MEME tools. Fifteen distinct motifs were found across the AdBBX proteins (Figure 5b, Figure S5). Among these motifs, motifs 1 and 5 were found in all the AdBBX proteins. Conservation of AdBBX

protein structures was observed among genes that clustered into the same classes. For example, all class I members shared five motifs, including motifs 1, 2, 3, 4, and 5, and all class II members shared six motifs, including motifs 1, 2, 4, 5, 14, and 15. In addition, structural diversity was also found among the AdBBX proteins. The motif numbers in AdBBX proteins varied from two (AdBBX19) to seven (AdBBX7, AdBBX10, and AdBBX21), and some motifs were only found in specific AdBBX proteins. For example, motif 2 was specific to class I, II, and III proteins and was considered to be the CCT domain, and motif 15 was only found in class II proteins.

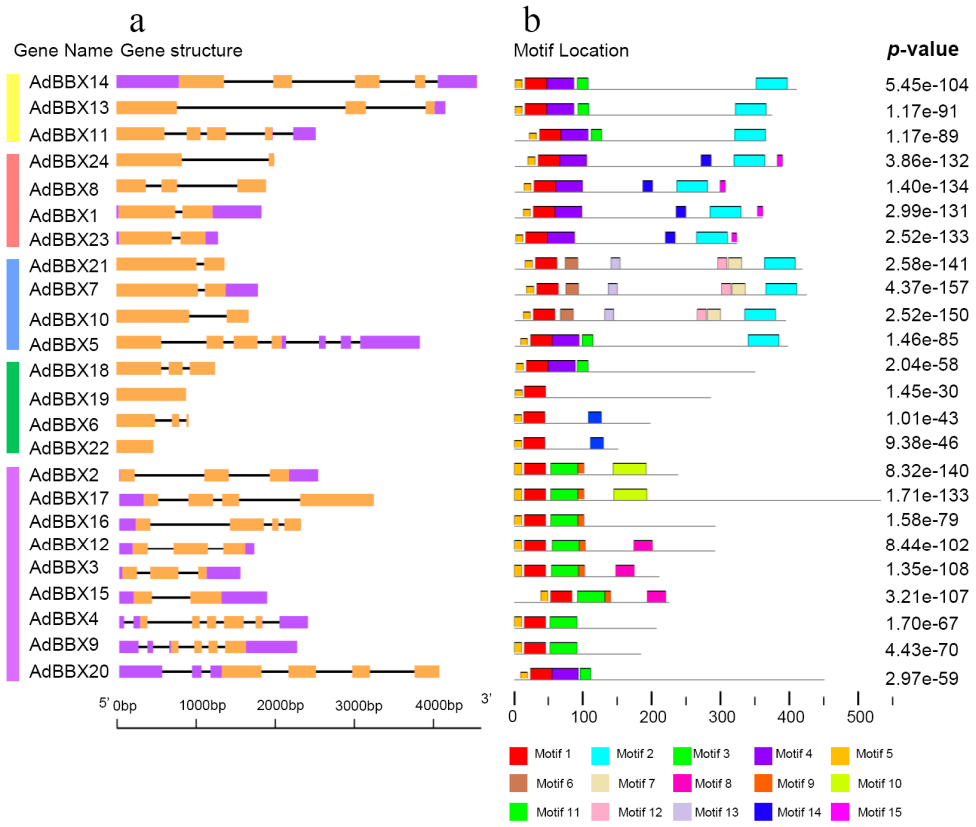


Figure 5. Analyses of exon–intron organization and conserved motifs of *AdBBX* genes. (**a**) The exon–intron organization of *AdBBX* genes. Purple, orange, and black boxes indicate untranslated regions (UTRs), exons, and introns, respectively. (**b**) The motifs of AdBBX proteins identified by MEME tools. Colored boxes 1–15 indicate different motifs.

3.5. Duplication Analysis of BBX Genes in A. duranensis

Polyploidy is a common feature of flowering plant evolution and produces many duplicated gene pairs. The wild peanut *A. duranensis* is thought to have experienced one round of whole genome duplication [33–37]. We investigated the duplication of *AdBBX* genes and found two interchromosomal duplicated gene pairs (*AdBBX2/AdBBX17* and *AdBBX8/AdBBX24*), but no tandem duplicated gene pairs (Figure 6). These duplicated genes were located on chromosomes 3, 4, 8, and 10, respectively. Moreover, the duplicated genes *AdBBX2/AdBBX17* were found to belong to group IV and *AdBBX8/AdBBX24* belonged to group II, and no duplicated gene pairs were found among the other groups. We also investigated duplication gene pairs in *A. ipaensis* and found two duplication events (*AraipAS7FB/AraipP77MW*, *AraipWH6UQ/AraipS98FB*), and all these genes were found to belong to group IV (Figure S6). Moreover, the allotetraploid *A. hypogaea* contained 16 duplication events, and these genes, group II, III, IV, and V contained 1, 10, 4, and 17 members, respectively (Figure S7).

Figure 6. Duplication analyses of *AdBBX* genes. The chromosomes are indicated by different colors, and the duplicated gene pairs are marked with lines.

3.6. Analysis of Cis-Acting Elements in AdBBX Promoter Regions

Cis-acting elements in promoter regions have critical roles in regulating plant gene expression. To further understand the expression responses of *AdBBX* genes, we identified cis-acting elements in *AdBBX* promoter regions 2 kb upstream of the initiation codon using PlantCARE [51]. Ninety-five kinds of cis-acting elements were identified, and 53 were predicted to have putative functions. These included seven development-related elements, five environmental stress-related elements, four site-binding-related elements, nine hormone-responsive elements, four promoter-related elements, and 24 light-responsive elements (Table S1). Binding sites in all 24 *AdBBX* genes related to development, including circadian control, metabolism regulation, stem expression, seed-specific regulation, cell differentiation, and cell cycle regulation (Figure 7a), environmental stress, including anaerobic conditions, drought, low temperature, and defense (Figure 7b), and hormones, including abscisic acid (ABA), gibberellic acid (GA), auxin, and jasmonic acid (MeJA) (Figure 7c), were identified in these promoters. The numbers and types of cis-acting elements varied among the *AdBBX* promoters, suggesting that the *AdBBX* genes have diverse roles in plant developmental regulation (Table 2). All the *AdBBX* promoters contained light-responsive elements, which were represented to be the most abundant element type in each of the *AdBBX* promoters, hormone-responsive elements, and promoter-related elements (Table 2, Supplementary Table S1), suggesting that the *AdBBX* genes share some common pathways in plant developmental regulation. The promoter-related elements CCAAT-box and TATA-box were found in all *AdBBX* promoter regions and likely constitute the basic components of the promoters. The light-responsive element, Box4, was identified in all *AdBBX* promoters except that of *AdBBX13* (Table S1), indicating that *AdBBX* genes play important roles in Box4-mediated light response pathways.

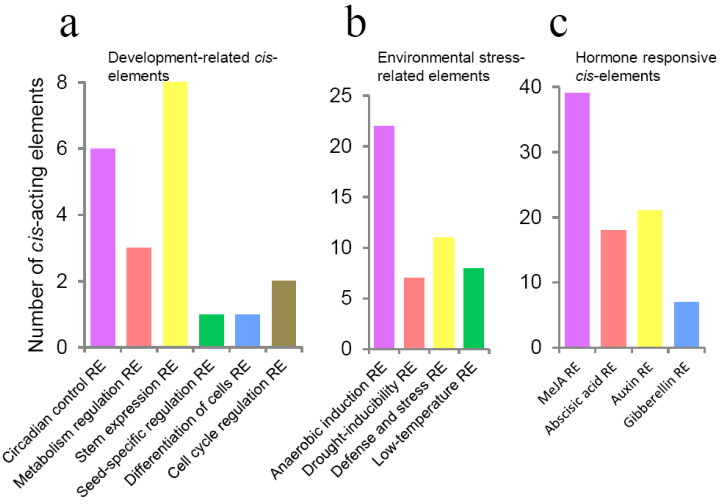


Figure 7. Cis-acting element analysis of *AdBBX* promoter regions. (**a**) Development-related cis-elements (circadian control, metabolism regulation, stem expression, seed-specific regulation, differentiation of cells, and cell cycle regulation) in *AdBBX* promoters. (**b**) Environmental stress–related cis-elements (anaerobic, drought, low temperature, and defense) in *AdBBX* promoter regions. (**c**) Hormone responsive cis-elements (abscisic acid, gibberellic acid, auxin, and jasmonic acid) in *AdBBX* promoter regions.

Table 2. Number and type of cis-acting elements in each *AdBBX* promoter region.

Gene Name	Development-Related Elements	Environmental Stress-Related Elements	Hormone-Responsive Elements	Light-Responsive Elements	Promoter-Related Elements	Site-Binding Related Elements	Others
AdBBX1	2	1	5	11	2	1	17
AdBBX2	3	2	3	6	3	1	15
AdBBX3	0	1	2	3	2	0	15
AdBBX4	2	2	3	3	2	1	12
AdBBX5	1	1	3	8	3	1	16
AdBBX6	0	0	3	3	2	1	15
AdBBX7	2	2	4	5	2	2	19
AdBBX8	1	2	1	8	2	1	13
AdBBX9	0	3	4	7	2	1	16
AdBBX10	1	2	4	4	2	0	18
AdBBX11	1	2	4	6	2	3	17
AdBBX12	0	1	4	5	2	0	9
AdBBX13	1	2	2	7	2	1	12
AdBBX14	0	2	2	11	2	0	13
AdBBX15	3	2	4	5	4	1	20
AdBBX16	1	2	3	10	2	0	19
AdBBX17	0	1	5	9	2	1	15
AdBBX18	0	1	3	7	2	0	16
AdBBX19	1	3	5	10	2	2	24
AdBBX20	0	3	3	6	2	0	15
AdBBX21	0	5	4	9	3	1	17
AdBBX22	2	3	3	8	2	0	11
AdBBX23	1	3	5	6	3	1	19
AdBBX24	0	2	6	7	2	0	15

3.7. Expression Patterns of AdBBX Genes in Different Tissues

To shed light on the potential functions of *AdBBX* genes during plant development, we investigated the expression levels of the 24 *AdBBX* genes in 22 different tissues (Figure 8). *AdBBX* genes showed distinct expression patterns among tissues, highlighting their functional diversity. For example, *AdBBX1* and *AdBBX23* were expressed at high levels in most tissues. In contrast, *AdBBX3*, *AdBBX5*, *AdBBX15*, *AdBBX18*, and *AdBBX24* showed low expression in all tissues, suggesting that they may

function only during specific stages of development in these tissues. *AdBBX21* was expressed at high levels in root nodules but showed low expression in other tissues, including roots, indicating that it may be involved in the formation of *A. duranensis* root nodules.

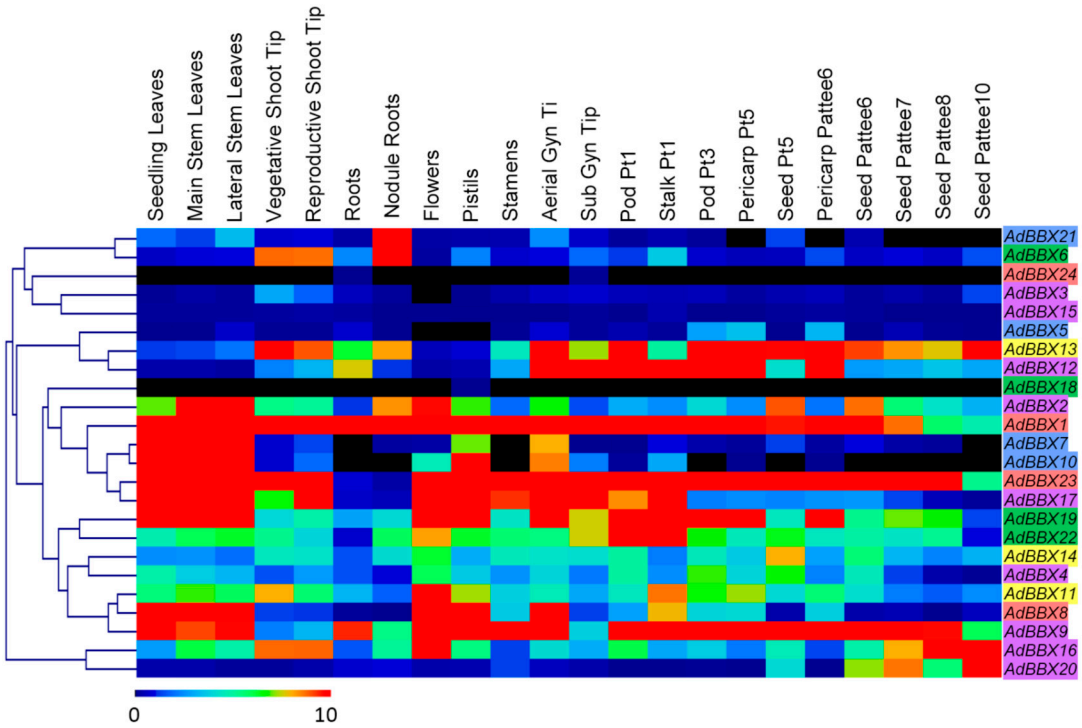


Figure 8. Expression patterns of *AdBBX* genes in 22 different tissues. Twenty-two peanut tissues were downloaded from the peanut genome database [53] and used for heat map analysis.

CO homologs are key factors in flowering time regulation in many species, and we therefore investigated the expression of *A. duranensis* orthologs of *Arabidopsis CO* and rice *Hd1* (i.e., *AdBBX8* and *AdBBX24*). *AdBBX8* was highly expressed in leaves, flowers, pistils, and Aerial Gyn Ti, while *AdBBX24* was expressed at low levels in all tissues. Some duplicated gene pairs, such as *AdBBX3/AdBBX15,* were expressed at similar levels in all tissues (Figure 8), suggesting functional conservation in the duplicated genes. In contrast, other duplicated pairs differed in their expression patterns. For example, *AdBBX9* was highly expressed in leaves and roots, while its duplicate *AdBBX4* exhibited low expression in the same tissues (Figure 8).

4. Discussion

In the past decades, the characterization of *BBX* genes, such as *Arabidopsis CO* and rice *Hd1*, has greatly increased our understanding of the molecular mechanisms involved in plant development. Peanut is an important crop around the world and provides essential oil for daily life, thus the investigation of peanut *BBX* genes is therefore useful for understanding and improving peanut development. In this study, we identified and characterized 24 BBX proteins from the wild peanut *A. duranensis* and carried out a comprehensive analysis of these genes.

BBX genes have changed during plant evolution, and the numbers and types of *BBXs* vary among different species [3,4,7,8,54]. For example, *A. duranensis*, *Arabidopsis*, rice, and pear contain 24, 32, 30, and 25 *BBX* members, respectively, and class IV contains the largest number of *BBX* genes in each species (Table S2). The genome sizes of diploid *A. duranensis* [33,55], *Arabidopsis* [56], rice [57], and pear [58] are 1.25 GB, 125 Mb, 466 Mb, and 512 Mb, respectively. Thus, genome size has no direct relationship to the number of *BBX* genes in these plants. Genes containing both BBX and CCT domains encode CO or CO-like (COL) proteins, and many *Arabidopsis CO*-like genes (*COL*) are thought to be

involved in circadian clock or flowering time regulation [4]. Approximately half of the BBX proteins are identified as *CO* or *COL* members (classes I, II, and III) (Figure 4) in *Arabidopsis* (53.13%), rice (56.67%), pear (44%), and *A. duranensis* (45.83%), suggesting that the evolution of *CO* and *COL* genes may be broadly conserved in these plants. In addition, the genome size of *A. hypogaea* (40 *BBX* members) is close to the sum of *A. duranensis* and *A. ipaensis* genomes, but the number of *BBX* genes in *A. hypogaea* is less than the sum of those in the *A. duranensis* (24 *BBX* members) and *A. ipaensis* (25 *BBX* members) genomes, indicating that *BBX* genes might be lost during evolution.

The cis-acting elements in promoter regions influence gene transcription, and differences in the type and number of cis-acting elements are responsible for differences in gene expression. *AtBBX* genes participate in the regulation of many pathways, including flowering time, the circadian clock, abiotic stress response, and photomorphogenesis [3,4]. Different numbers and types of cis-acting elements were found in the *AdBBX* promoter regions, underscoring the functional diversity of these genes. Many *BBX* genes in *Arabidopsis* are involved in light input signal pathways [4]. Light responsive elements were the most abundant elements in each of the *AdBBX* promoters, suggesting that *AdBBXs* may also be involved in light-dependent regulation pathways. Moreover, many cis-acting elements were also identified in the promoter regions of low expressed genes, including *AdBBX3*, *AdBBX5*, *AdBBX15*, *AdBBX18*, and *AdBBX24* (Figure 8). Many factors in addition to cis-acting elements affect gene expression in plants. For example, epigenetic modification and somatic genome variation also influence gene expression in many organisms [59]. Whether low-expressed *AdBBX* genes were influenced by these factors requires further investigation.

CO is an important factor involved in the regulation of flowering time in *Arabidopsis*, and it is highly expressed at the apex of seedlings and young leaves [60]. *CO* accelerates flowering time by activating the transcription of the Rafkinase inhibitor protein (RAF)-kinase-inhibitor-like protein, FT. *AdBBX8* and *AdBBX24* are close *A. duranensis* homologs of *CO* (Figure S4). Soybean *CO* orthologs, *GmCOL1a*, *GmCOL1b*, *GmCOL2a*, and *GmCOL2b*, are also involved in flowering time regulation [61]. Genes derived from the same common ancestor may have similar functions, and we therefore investigated synteny relationships among *CO* orthologs/homologs from soybean (*GmCOL1a*, *GmCOL1b*, *GmCOL2a*, and *GmCOL2b*) and *A. duranensis* (*AdBBX8* and *AdBBX24*) (Figure S8). Synteny analysis revealed that *AdBBX8* had a closer relationship to soybean *GmCOL1a* and *GmCOL1b* than did *AdBBX24*. In contrast, *AdBBX8* and *AdBBX24* showed similar close relationships to soybean *GmCOL2a* and *GmCOL2b*. *AdBBX8* was expressed highly in leaves, flowers, pistils, and aerial gyn Ti, but *AdBBX24* exhibited extremely low expression in all tissues (Figure 8). These results suggest that *AdBBX8* may play a similar role to *CO* in flowering time regulation, and that *AdBBX24* may be a redundant gene that has lost its functions during evolution. In addition, *CO* is regulated by the circadian clock and its expression changes during the day [62]. *AdBBX24* may therefore be expressed at a different time of day than the one at which the plants were sampled. Much work remains to be done to fully understand the functions of *AdBBX8* and *AdBBX24* in flowering time regulation.

Gene duplication produces new genes during evolution in many species. Some duplicated genes lose function, and some evolve new functions, compared to the original gene [63,64]. As a result, the allotetraploid *Arachis hypogaea* produces much more duplicated events than the sum of that in wild species *Arachis duranensis* and *Arachis ipaensis* (Figures S6 and S7). Two duplicated gene pairs were found in *A. duranensis*, and these duplication events occurred in class II and IV subfamilies, respectively, whose members contains two BBX domains (Figure 6). In addition, duplicated gene pairs had different exon–intron structures (Figure 5a), and their cis-acting promoter elements differed (Table 2), indicating functional differentiation of these gene pairs during evolution. The duplicated gene pairs contained similar motifs (Figure 5b), and some duplicated genes showed similar expression in specific tissues, such as *AdBBX2/AdBBX17* (Figure 8), indicating that these duplicated genes may have retained some original functions and may participate in common pathways.

5. Conclusions

In the present study, we identified and characterized 24 *BBX* genes from wild peanut, *A. duranensis*. We investigated their conserved domains, gene structures, phylogenetic relationships, chromosomal distributions, gene duplications, synteny relationships, cis-acting elements, and gene expression. Our results will not only be useful for understanding *AdBBX* genes but will also provide essential information for further functional analysis of these genes in *A. duranensis*.

Supplementary Materials: Supplementary Materials can be found at http://www.mdpi.com/2073-4395/10/1/23/s1. Figure S1. Phylogenetic analysis of AiBBX proteins. A phylogenetic tree was generated from protein sequences in MEGA 7 using the neighbor-joining method. The yellow, red, blue, purple, and green gene names indicate class I, II, III, IV, and V members, respectively. Figure S2. Phylogenetic analysis of AhBBX proteins. A phylogenetic tree was generated from protein sequences in MEGA 7 using the neighbor-joining method. The yellow, red, blue, purple, and green gene names indicate class I, II, III, IV, and V members, respectively. Figure S3. Phylogenetic analysis of AdBBX, AiBBX, and AhBBX proteins. A phylogenetic tree was generated from protein sequences in MEGA 7 using the neighbor-joining method. The yellow, red, blue, purple, and green gene names indicate class I, II, III, IV, and V members, respectively. Figure S4. Evolutionary relationship analysis of AdBBX proteins. Amino acid sequences of BBX proteins from *A. duranensis*, rice, and *Arabidopsis* were used to generate a phylogenetic tree with MEGA 7.0 using the neighbor-joining method. Figure S5. Sequence logos of 15 distinct motifs in 24 AdBBX proteins. Figure S6. Duplication analysis of *AiBBX* genes. The chromosomes are indicated by different colors, and the duplicated gene pairs are marked with lines. Figure S7. Duplication analysis of *AhBBX* genes. The chromosomes are indicated by different colors, and the duplicated gene pairs are marked with lines. Figure S8. Synteny analysis of *CO* orthologs/homologs in soybean and *A. duranensis*. The putative orthologous genes surrounding *CO* orthologs/homologs from soybean (*GmCOL1a*, *GmCOL1b*, *GmCOL2a*, and *GmCOL2b*) and *A. duranensis* (*AdBBX8* and *AdBBX24*) were identified by BLASTP search. Synteny between *AdBBX8*, *AdBBX24*, and (a) *GmCOL1a*, (b) *GmCOL1b*, (c) *GmCOL2a*, and (d) *GmCOL2b* are shown. The red boxes indicate our target genes, and the green boxes indicate genes surrounding the *CO* orthologs/homologs. Gm, *Glycine max*. Table S1. Function analysis of the cis-acting elements in *AdBBX* promoter regions. The classifications, names, and putative functions of related cis-acting elements are predicted and listed. Table S2. The numbers and types of BBX proteins in *Arabidopsis*, rice, pear, and *Arachis duranensis*.

Author Contributions: S.L. conceived and designed the experiments and wrote the manuscript. H.J. and M.X. conducted the experiments, bioinformatics work, and created the figures. C.C. contributed with valuable discussions and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant 31971898, 31601372), and the Qingdao Agricultural University Scientific Research Foundation (grant 6631115039, 6631119010).

Acknowledgments: The authors thank the reviewers for their valuable suggestions during the revision of the manuscripts.

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

Abbreviations

BBX	B-box
CCT	CONSTANS, CONSTANS-LIKE and TIMING OF CAB1
LD	Long day
SD	Short day
VP	valine-proline
CO	CONSTANS
CORE	CO-responsive elements
FT	FLOWERING LOCUS T
COL	CO-like
GSDS	Gene Structure Display Server program
ABA	Abscisic acid
GA	Gibberellic acid
Ad	Arachis duranensis
Gm	Glycine max
Os	Oryza sativa
pI	Isoelectric point
MW	Molecular weight

References

1. Riechmann, J.L.; Heard, J.; Martin, G.; Reuber, L.; Jiang, C.; Keddie, J.; Adam, L.; Pineda, O.; Ratcliffe, O.J.; Samaha, R.R.; et al. Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* **2000**, *290*, 2105–2110. [CrossRef] [PubMed]
2. Kielbowicz-Matuk, A. Involvement of plant C2H2-type zinc finger transcription factors in stress responses. *Plant Sci.* **2012**, *185–186*, 78–85. [CrossRef] [PubMed]
3. Khanna, R.; Kronmiller, B.; Maszle, D.R.; Coupland, G.; Holm, M.; Mizuno, T.; Wu, S.H. The Arabidopsis *B-Box* zinc finger family. *Plant Cell* **2009**, *21*, 3416–3420. [CrossRef] [PubMed]
4. Gangappa, S.N.; Botto, J.F. The *BBX* family of plant transcription factors. *Trends Plant Sci.* **2014**, *19*, 460–470. [CrossRef] [PubMed]
5. Zhang, W.; Yuan, J.; Cheng, T.; Tang, M.; Sun, K.; Song, S.; Xu, F.; Dai, C. Flowering-mediated root-fungus symbiosis loss is related to jasmonate-dependent root soluble sugar deprivation. *Plant Cell Environ.* **2019**, 1–19. [CrossRef] [PubMed]
6. Ronald, J.; Davis, S.J. Focusing on the nuclear and subnuclear dynamics of light and circadian signalling. *Plant Cell Environ.* **2019**, *42*, 2871–2884. [CrossRef]
7. Huang, J.; Zhao, X.; Weng, X.; Wang, L.; Xie, W. The rice *B-Box* zinc finger gene family: Genomic identification, characterization, expression profiling and diurnal analysis. *PLoS ONE* **2012**, *7*, e48242. [CrossRef]
8. Cao, Y.; Han, Y.; Meng, D.; Li, D.; Jiao, C.; Jin, Q.; Lin, Y.; Cai, Y. *B-BOX* genes: Genome-wide identification, evolution and their contribution to pollen growth in pear (*Pyrus bretschneideri* Rehd.). *BMC Plant Biol.* **2017**, *17*, 156. [CrossRef]
9. Shalmani, A.; Jing, X.Q.; Shi, Y.; Muhammad, I.; Zhou, M.R.; Wei, X.Y.; Chen, Q.Q.; Li, W.Q.; Liu, W.T.; Chen, K.M. Characterization of *B-BOX* gene family and their expression profiles under hormonal, abiotic and metal stresses in *Poaceae* plants. *BMC Genom.* **2019**, *20*, 27. [CrossRef]
10. Massiah, M.A.; Simmons, B.N.; Short, K.M.; Cox, T.C. Solution structure of the RBCC/TRIM B-box1 domain of human MID1: B-box with a RING. *J. Mol. Biol.* **2006**, *358*, 532–545. [CrossRef]
11. Massiah, M.A.; Matts, J.A.; Short, K.M.; Simmons, B.N.; Singireddy, S.; Yi, Z.; Cox, T.C. Solution structure of the MID1 B-box2 CHC(D/C)C2H2 Zinc-binding domain: Insights into an evolutionarily conserved RING fold. *J. Mol. Biol.* **2007**, *369*, 1–10. [CrossRef] [PubMed]
12. Crocco, C.D.; Botto, J.F. BBX proteins in green plants: Insights into their evolution, structure, feature and functional diversification. *Gene* **2013**, *531*, 44–52. [CrossRef] [PubMed]
13. Griffiths, S.; Dunford, R.P.; Coupland, G.; Laurie, D.A. The evolution of *CONSTANS*-like gene families in barley, rice, and Arabidopsis. *Plant Physiol.* **2003**, *131*, 1855–1867. [CrossRef] [PubMed]
14. Putterill, J.; Robson, F.; Lee, K.; Simon, R.; Coupland, G. The *CONSTANS* gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **1995**, *80*, 847–857. [CrossRef]
15. Samach, A.; Onouchi, H.; Gold, S.E.; Ditta, G.S.; Schwarz-Sommer, Z.; Yanofsky, M.F.; Coupland, G. Distinct roles of *CONSTANS* target genes in reproductive development of Arabidopsis. *Science* **2000**, *288*, 1613–1616. [CrossRef]
16. Valverde, F.; Mouradov, A.; Soppe, W.; Ravenscroft, D.; Samach, A.; Coupland, G. Photoreceptor regulation of *CONSTANS* protein in photoperiodic flowering. *Science* **2004**, *303*, 1003–1006. [CrossRef]
17. Tiwari, S.B.; Shen, Y.; Chang, H.C.; Hou, Y.; Harris, A.; Ma, S.F.; McPartland, M.; Hymus, G.J.; Adam, L.; Marion, C.; et al. The flowering time regulator *CONSTANS* is recruited to the *FLOWERING LOCUS T* promoter via a unique *cis*-element. *New Phytol.* **2010**, *187*, 57–66. [CrossRef]
18. Wenkel, S.; Turck, F.; Singer, K.; Gissot, L.; Le Gourrierec, J.; Samach, A.; Coupland, G. *CONSTANS* and the CCAAT Box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. *Plant Cell* **2006**, *18*, 2971–2984. [CrossRef]
19. Ben-Naim, O.; Eshed, R.; Parnis, A.; Teper-Bamnolker, P.; Shalit, A.; Coupland, G.; Samach, A.; Lifschitz, E. The CCAAT binding factor can mediate interactions between *CONSTANS*-like proteins and DNA. *Plant J.* **2006**, *46*, 462–476. [CrossRef]
20. Luccioni, L.; Krzymuski, M.; Sánchez-Lamas, M.; Karayekov, E.; Cerdán, P.D.; Casal, J.J. *CONSTANS* delays Arabidopsis flowering under short days. *Plant J.* **2019**, *97*, 923–932. [CrossRef]

21. Luo, X.; Gao, Z.; Wang, Y.; Chen, Z.; Zhang, W.; Huang, J.; Yu, H.; He, Y. The NUCLEAR FACTOR-*CONSTANS* complex antagonizes Polycomb repression to derepress *FLOWERING LOCUS T* expression in response to inductive long days in Arabidopsis. *Plant J.* **2018**, *95*, 17–29. [CrossRef] [PubMed]
22. Yano, M.; Katayose, Y.; Ashikari, M.; Yamanouchi, U.; Monna, L.; Fuse, T.; Baba, T.; Yamamoto, K.; Umehara, Y.; Nagamura, Y.; et al. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. *Plant Cell* **2000**, *12*, 2473–2483. [CrossRef] [PubMed]
23. Li, S.; Ying, Y.; Secco, D.; Wang, C.; Narsai, R.; Whelan, J.; Shou, H. Molecular interaction between PHO2 and GIGANTEA reveals a new crosstalk between flowering time and phosphate homeostasis in *Oryza sativa*. *Plant Cell Enviton.* **2017**, *40*, 1487–1499. [CrossRef] [PubMed]
24. Komiya, R.; Ikegami, A.; Tamaki, S.; Yokoi, S.; Shimamoto, K. *Hd3a* and *RFT1* are essential for flowering in rice. *Development* **2008**, *135*, 767–774. [CrossRef] [PubMed]
25. Komiya, R.; Yokoi, S.; Shimamoto, K. A gene network for long-day flowering activates *RFT1* encoding a mobile flowering signal in rice. *Development* **2009**, *136*, 3443–3450. [CrossRef]
26. Kochert, G.; Halward, T.; Branch, W.D.; Simpson, C.E. RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. *Theor. Appl. Genet.* **1991**, *81*, 565–570. [CrossRef]
27. Beinecke, F.A.; Grundmann, L.; Wiedmann, D.R.; Schmidt, F.J.; Caesar, A.S.; Zimmermann, M.; Lahme, M.; Twyman, R.M.; Prufer, D.; Noll, G.A. The FT/FD-dependent initiation of flowering under long-day conditions in the day-neutral species *Nicotiana* tabacum originates from the facultative short-day ancestor *Nicotiana* tomentosiformis. *Plant J.* **2018**, *96*, 329–342. [CrossRef]
28. Wu, F.; Sedivy, E.J.; Price, W.B.; Haider, W.; Hanzawa, Y. Evolutionary trajectories of duplicated *FT* homologues and their roles in soybean domestication. *Plant J.* **2017**, *90*, 941–953. [CrossRef]
29. Eom, H.; Park, S.J.; Kim, M.K.; Kim, H.; Kang, H.; Lee, I. *TAF15b*, involved in the autonomous pathway for flowering, represses transcription of FLOWERING LOCUS C. *Plant J.* **2018**, *93*, 79–91. [CrossRef]
30. Jing, Y.; Guo, Q.; Zha, P.; Lin, R. The chromatin-remodelling factor PICKLE interacts with *CONSTANS* to promote flowering in Arabidopsis. *Plant Cell Environ.* **2019**, *42*, 2495–2507. [CrossRef]
31. Jin, H.; Tang, X.; Xing, M.; Zhu, H.; Sui, J.; Cai, C.; Li, S. Molecular and transcriptional characterization of phosphatidyl ethanolamine-binding proteins in wild peanuts *Arachis duranensis* and *Arachis ipaensis*. *BMC Plant Biol.* **2019**, *19*, 484. [CrossRef] [PubMed]
32. Moretzsohn, M.C.; Gouvea, E.G.; Inglis, P.W.; Leal-Bertioli, S.C.; Valls, J.F.; Bertioli, D.J. A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Ann. Bot.* **2013**, *111*, 113–126. [CrossRef] [PubMed]
33. Bertioli, D.J.; Cannon, S.B.; Froenicke, L.; Huang, G.; Farmer, A.D.; Cannon, E.K.; Liu, X.; Gao, D.; Clevenger, J.; Dash, S.; et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* **2016**, *48*, 438–446. [CrossRef] [PubMed]
34. Chen, X.; Li, H.; Pandey, M.K.; Yang, Q.; Wang, X.; Garg, V.; Li, H.; Chi, X.; Doddamani, D.; Hong, Y.; et al. Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 6785–6790. [CrossRef] [PubMed]
35. Bertioli, D.J.; Jenkins, J.; Clevenger, J.; Dudchenko, O.; Gao, D.; Seijo, G.; Leal-Bertioli, S.C.M.; Ren, L.; Farmer, A.D.; Pandey, M.K.; et al. The genome sequence of segmental allotetraploid peanut *Arachi shypogaea*. *Nat. Genet.* **2019**, *51*, 877–884. [CrossRef]
36. Chen, X.; Lu, Q.; Liu, H.; Zhang, J.; Hong, Y.; Lan, H.; Li, H.; Wang, J.; Liu, H.; Li, S.; et al. Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. *Mol. Plant* **2019**, *12*, 920–934. [CrossRef]
37. Zhuang, W.; Chen, H.; Yang, M.; Wang, J.; Pandey, M.K.; Zhang, C.; Chang, W.C.; Zhang, L.; Zhang, X.; Tang, R.; et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* **2019**, *51*, 865–876. [CrossRef]
38. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.; Luciani, A.; Potter, S.; Qureshi, M.; Richardson, L.; Salazar, G.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, 427–432. [CrossRef]
39. Finn, R.; Attwood, T.; Babbitt, P.; Bateman, A.; Bork, P.; Bridge, A.; Chang, H.; Dosztányi, Z.; El-Gebali, S.; Fraser, M.; et al. InterPro in 2017–beyond protein family and domain annotations. *Nucleic Acids Res.* **2017**, *45*, 190–199. [CrossRef]

40. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and analysis tools in the ExPASy Server. In *The Proteomics Protocols Handbook*; John, M.W., Ed.; Humana Press: New York, NY, USA, 2005; pp. 571–607.
41. Oliver, T.; Schmidt, B.; Nathan, D.; Clemens, R.; Maskell, D. Using reconfigurable hardware to accelerate multiple sequence alignment with clustalw. *Bioinformatics* **2005**, *21*, 3431–3432. [CrossRef]
42. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]
43. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0. An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [CrossRef] [PubMed]
44. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. Meme suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef] [PubMed]
45. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logogenerator. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef] [PubMed]
46. Zhang, D.; Zhao, M.; Li, S.; Sun, L.; Wang, W.; Cai, C.; Dierking, E.C.; Ma, J. Plasticity and innovation of regulatory mechanisms underlying seed oil content mediated by duplicated genes in the palaeopolyploid soybean. *Plant J.* **2017**, *90*, 1120–1133. [CrossRef] [PubMed]
47. Li, S.; Ding, Y.; Zhang, D.; Wang, X.; Tang, X.; Dai, D.; Jin, H.; Lee, S.H.; Cai, C.; Ma, J. Parallel domestication with a broad mutational spectrum of determinate stem growth habit in leguminous crops. *Plant J.* **2018**, *96*, 761–771. [CrossRef]
48. Fischer, S.; Brunk, B.P.; Chen, F.; Gao, X.; Harb, O.S.; Iodice, J.B.; Shanmugam, D.; Roos, D.S.; Stoeckert, C.J. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr. Protoc. Bioinformatics* **2011**, *35*, 1–19.
49. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [CrossRef]
50. Hou, L.; Zhang, Z.; Dou, S.; Zhang, Y.; Pang, X.; Li, Y. Genome-wide identification, characterization, and expression analysis of the expansin gene family in Chinese jujube (*Zizi phusjujuba* Mill.). *Planta* **2019**, *249*, 815–829. [CrossRef]
51. Lescot, M.; Dehais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouze, P.; Rombauts, S. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [CrossRef]
52. Saeed, A.I.; Sharov, V.; White, J.; Li, J.; Liang, W.; Bhagabati, N.; Braisted, J.; Klapa, M.; Currier, T.; Thiagarajan, M.; et al. TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* **2003**, *34*, 374–378. [CrossRef] [PubMed]
53. Clevenger, J.; Chu, Y.; Scheffler, B.; Ozias-Akins, P. A developmental transcriptome map for allotetraploid *Arachis hypogaea*. *Front. Plant Sci.* **2016**, *7*, 1446. [CrossRef] [PubMed]
54. Liu, X.; Li, R.; Dai, Y.; Chen, X.; Wang, X. Genome-wide identification and expression analysis of the *B-box* gene family in the Apple (*Malus domestica* Borkh.) genome. *Mol. Genet. Genomics* **2018**, *293*, 303–315. [CrossRef] [PubMed]
55. Samoluk, S.S.; Robledo, G.; Podio, M.; Chalup, L.; Ortiz, J.P.; Pessino, S.C.; Seijo, J.G. First insight into divergence, representation and chromosome distribution of reverse transcriptase fragments from L1 retrotransposons in peanut and wild relative species. *Genetica* **2015**, *143*, 113–125. [CrossRef]
56. Initiative, A.G. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **2000**, *408*, 796–815. [CrossRef]
57. Yu, J.; Hu, S.; Wang, J.; Wong, G.K.-S.; Li, S.; Liu, B.; Deng, Y.; Dai, L.; Zhou, Y.; Zhang, X.; et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* **2002**, *296*, 79–92. [CrossRef]
58. Wu, J.; Wang, Z.; Shi, Z.; Zhang, S.; Ming, R.; Zhu, S.; Khan, M.A.; Tao, S.; Korban, S.S.; Wang, H.; et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* **2013**, *23*, 396–408. [CrossRef]
59. Li, X.Q.; Xing, T.; Du, D. Identification of top-ranked proteins withina directional protein interaction network using the pagerank algorithm: Applications in humans and plants. *Curr. Issues Mol. Biol.* **2016**, *20*, 13–28.
60. An, H.; Roussot, C.; Suárez-López, P.; Corbesier, L.; Vincent, C.; Piñeiro, M.; Hepworth, S.; Mouradov, A.; Justin, S.; Turnbull, C.; et al. *CONSTANS* acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. *Development* **2004**, *131*, 3615–3626. [CrossRef]

61. Wu, F.; Price, B.W.; Haider, W.; Seufferheld, G.; Nelson, R.; Hanzawa, Y. Functional and evolutionary characterization of the *CONSTANS* gene family in short-day photoperiodic flowering in soybean. *PLoS ONE* **2014**, *9*, e85754. [CrossRef]
62. Suárez-López, P.; Wheatley, K.; Robson, F.; Onouchi, H.; Valverde, F.; Coupland, G. *CONSTANS* mediates between the circadian clock and the control of flowering in Arabidopsis. *Nature* **2001**, *410*, 1116–1120. [CrossRef] [PubMed]
63. Kondrashov, F.A.; Rogozin, I.B.; Wolf, Y.I.; Koonin, E.V. Selection in the evolution of gene duplications. *Genome Biol.* **2002**, *3*, research0008. [CrossRef] [PubMed]
64. Wang, Z.; Zhou, Z.; Liu, Y.; Liu, T.; Li, Q.; Ji, Y.; Li, C.; Fang, C.; Wang, M.; Wu, M.; et al. Functional evolution of phosphatidyl ethanolamine binding proteins in soybean and Arabidopsis. *Plant Cell* **2015**, *27*, 323–336. [CrossRef] [PubMed]