Characterization of B-box proteins and their contribution to plant development in Arachis duranensis

Hanqi Jin
Qingdao Agricultural University
Mengge Xing
Qingdao Agricultural University
Chunmei Cai
Qingdao Agricultural University
Shuai Li (li2014shuai@qau.edu.cn)
Qingdao Agricultural University https://orcid.org/0000-0002-6905-2349

Research article

Keywords: Peanut, B-box (BBX), Arachis duranensis, flowering time, gene expression

Posted Date: July 6th, 2019

DOI: https://doi.org/10.21203/rs.2.11044/v1

Abstract

Background B-box (BBX) proteins are important factors involving in the regulation of plant growth and development, and have been identified in many plant species. However, the characteristics and transcription patterns of BBX genes in wild peanut are limited. Results In the present study, we identified and characterized 24 BBX genes in a wild peanut Arachis duranensis. The AdBBX members distributed on 9 of the 10 chromosomes and chromosome 3 contained the most AdBBX members, with 6 AdBBXs. 16 AdBBX proteins had two distinct BBX domains, 11 members contained one CCT domain, and 7 genes had both BBX and CCT domains. Protein structure analysis revealed that AdBBX were classified into five clades: I (3 genes), II (4 genes), III (4 genes), IV (9 genes) and V (4 genes), on the basis of the diversity of conserved BBX and CCT domains. Moreover, 15 distinct motifs were found in these 24 AdBBX proteins and motif 1 and 5 existed in all the AdBBX proteins. Duplication analysis revealed that 4 interchromosomal duplicated gene pairs were obtained and all of them belonged to group IV. In addition, 95 kinds of cis-acting elements were found in the promoter regions of AdBBXs and 53 types were predicted to have putative functions. The numbers and types of cis-acting elements varied in these AdBBX promoters, as a result, AdBBX genes exhibited distinct expression levels in different tissues. The transcription investigation combined with synteny analysis suggested AdBBX8 might be the key factor involving in flowering time regulation in Arachis duranensis. Conclusion Overall, this study provides a genome-wide identification of BBX genes in a wild peanut Arachis duranensis. Characteristic and transcription pattern analysis revealed their critical roles in plant growth and development. Our study will provide essential information for further functional characteristic investigation of AdBBX genes.

Abstract

Background: B-box (BBX) proteins are important factors involving in the regulation of plant growth and development, and have been identified in many plant species. However, the characteristics and transcription patterns of *BBX* genes in wild peanut are limited.

Results: In the present study, we identified and characterized 24 *BBX* genes in a wild peanut *Arachis duranensis*. The *AdBBX* members distributed on 9 of the 10 chromosomes and chromosome 3 contained the most *AdBBX* members, with 6 *AdBBXs*. 16 AdBBX proteins had two distinct BBX domains, 11 members contained one CCT domain, and 7 genes had both BBX and CCT domains. Protein structure analysis revealed that AdBBX were classified into five clades: I (3 genes), II (4 genes), III (4 genes), IV (9 genes) and V (4 genes), on the basis of the diversity of conserved BBX and CCT domains. Moreover, 15 distinct motifs were found in these 24 AdBBX proteins and motif 1 and 5 existed in all the AdBBX proteins. Duplication analysis revealed that 4 interchromosomal duplicated gene pairs were obtained and all of them belonged to group IV. In addition, 95 kinds of *cis*-acting elements were found in the promoter regions of *AdBBXs* and 53 types were predicted to have putative functions. The numbers and types of *cis*-acting elements varied in these *AdBBX* promoters, as a result, *AdBBX* genes exhibited distinct expression levels in different tissues. The transcription investigation combined with synteny analysis suggested *AdBBX8* might be the key factor involving in flowering time regulation in *Arachis duranensis*.

Conclusion: Overall, this study provides a genome-wide identification of *BBX* genes in a wild peanut *Arachis duranensis*. Characteristic and transcription pattern analysis revealed their critical roles in plant growth and development. Our study will provide essential information for further functional characteristic investigation of *AdBBX* genes.

Key words: Peanut, B-box (BBX), *Arachis duranensis*, flowering time, gene expression

Background

Transcription factors are essential elements participating in signal transduction pathways in plants. They often work as activators or suppressors to bind *cis*-acting elements in the target promoter regions to regulate downstream gene expressions [1, 2]. Various kinds of transcription factors have been found in plants and considered to be involved in different response pathways. Among these gene families, zinc-finger transcription factors, consisting of a conserved domain which can bind metal ions like zinc and interact with DNA, RNA or proteins, are a large transcription factor family and play critical roles in plant growth and development [3, 4]. Based on the diversification of protein sequences and structures, the 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 considered to be involved in protein-protein interactions, is designated as BBX family and highly conserved across all multicellular species [4-7].

Two types of BBX domains, B-box1 and B-box2, were found based on their consensus sequences and the spacing of zinc-binding residues [4, 8-10]. In Arabidopsis, 21 and 11 of the 32 BBX proteins are found to contain two and one BBX domains, respectively [3, 4]. In addition to the conserved BBX domain, some of the BBX members contain some other specific domains, such as CCT (for CONSTANS, CONSTANS-LIKE, TOC1) and VP (valine-proline) domains. 17 Arabidopsis *BBX* members contained a CCT domain close to their C termini of the protein. BBX proteins are grouped into five clades in Arabidopsis based on the differences of the types and numbers of BBX and CCT domains. Class I and II members have two BBX and one CCT domains and class III contained one BBX and one CCT domains, respectively. Class IV has two different BBX domains and group V only has one BBX domain [3, 4, 11].

The functions of BBX proteins have been revealed to be involved in the regulation many signal pathways in recent years, including flowering time, circadian clock, seedling photomorphogenesis and abiotic stress [4]. CONSTANS (CO/BBX1) is 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* shows significantly delayed flowering time under LD, while the flowering time of *co* is similar as wild type plants under SD [12-14]. CO protein directly binds to the *CO*-responsive elements (CORE) and CCAAT-box elements in the promoter region of *FLOWERING LOCUS T* (*FT*) to promote *FT* expression, which is responsible for the acceleration of flowering time [15-17]. The *CO-FT* flowering time regulatory pathway is conserved in some other species [18-22]. For example, *Heading date 1* (*Hd1*), the *CO* ortholog in rice, participates in the regulation of *FT* orthologs in rice, *Heading date 3a (OsHd3a)* and *Flowering Locus T1 (OsRFT1)* [20, 21]. *Hd1* promotes flowering time under SD but delayed flowering time under LD [18]. In

addition, some CO-like (COL) proteins in Arabidopsis have also been shown to be involved in flowering time or circadian clock regulation, such as *BBX2*, *BBX3*, *BBX4*, *BBX6* and *BBX7*. Moreover, *BBX18, BBX24* and *BBX32* are considered to be involved in abiotic and biotic stress responses [4].

Peanut is an important oil crop and widely cultivated in many countries. The allotetraploid *Arachis hypogaea* (AABB, 2*n*=4*x*=40) is considered to be derived from the hybridization and polyploidization of two diploid peanuts, *Arachis duranensis* (AA genome) and *Arachis ipaensis* (BB genome) [22-24]. The emergence of peanut genome database in recent years allows the investigation of peanut gene functions [24-28]. Although much knowledge on *BBX* functions has been advanced in many species, there is less research on their roles in peanut development. In this study, we identified and characterized 24 *BBX* proteins from one of the wild peanut species, *Arachis duranensis*. We investigated many characteristics of these *BBX* genes, including gene structures, conserved motifs, chromosome localizations, phylogenetic relationships, gene duplications, *cis*-acting elements in promoter regions and tissue expression profiles. Our study will provide essential information for further functional characterization of *BBX* proteins in peanut.

Methods

Identification of *BBX* members in *Arachis duranensis*

The amino acid sequences of the BBX conserved domain (PF00643) and Arabidopsis BBX proteins, downloaded from TAIR (https://www.arabidopsis.org/), were used as blast queries against peanut genome database to search for *Arachis duranensis BBX* genes (https://www.peanutbase.org/) [24]. All the output candidate genes were analyzed by Pfam database (http://pfam.xfam.org/search) and Pro Scan program (https://www.ebi.ac.uk/interpro/) to confirm the presence of BBX domains and remove genes without conserved BBX domains. The positions of the BBX and CCT domains in each AdBBX protein were analyzed by Pro Scan program. The genomic lengths, CDS lengths and amino acid numbers of these *AdBBX* genes were obtained from peanut genome database. The GC contents were analyzed using DNASTAR. Moreover, the chemical features of AdBBX proteins, such as molecular weight and theoretical iso-electric points, were investigated by ProtParam (https://web.expasy.org/protparam/).

Sequence alignment and phylogenetic relationship analyses

The amino acid sequences of BBX and CCT domains were aligned by Clustal-X2 and the results were used to generate the alignment map using DNAMAN software (Version 5.2.2, LynnonBiosoft, Canada). To analyze the evolution relationships of *AdBBX* genes with the well-studied *BBX* genes, the full lengths of BBX amino acid sequences from *Arachis duranensis*, rice and Arabidopsis were aligned by Clustal-X2, and the alignment results were used to construct the phylogenetic tree by MEGA7.0 with Neighbor-Joining method [29].

Gene structure, conserved motif and sequence logo analyses

The genomic and CDS sequences of *AdBBX* genes were downloaded from peanut genome database [24] and used for the analysis of gene structures using Gene Structure Display Server program (GSDS) (http://gsds.cbi.pku.edu.cn/) [30]. To investigate the conserved motifs in these proteins, the full lengths of AdBBX amino acid sequences were obtained from peanut genome database and analyzed using MEME tools (http://meme-suite.org/). In addition, the sequence logos of BBX (B-box1 and B-box2) and CCT conserved domains were investigated using online WebLogo platform (http://weblogo.berkeley.edu/logo.cgi) [31].

Chromosomal localization, synteny and gene duplication analyses

AdBBX genes were mapped to peanut genome to identify the chromosomal localizations and physical positions, and chromosomal distribution map was generated using MapInspect software (http://www.mybiosoftware.com/mapinspect-compare-dis

play-linkage-maps.html). To investigate the synteny relationships of the related genome regions in different species, the putative orthologous genes surrounding *CO* orthologs/homologs from soybean and *Arachis duranensis* were identified by BLASTP search as described [32, 33], and the blast results were used for the generation of the synteny map. To analyze the duplicated gene pairs, we clustered *AdBBX* genes using OrthoMCL software (https://orthomcl.org/orthomcl/), and the duplicated gene pairs were drawn by Circos software as described [34, 35].

C*is*-acting element analysis

The promoter sequences, 2kb upstream of the initiation code of each gene, were obtained from peanut genome database [24], and used for further *cis*-acting element analysis using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/htm

l/) [36]. Then the *cis*-acting elements were classified based on their putative functions.

Analysis of gene expression in different tissues

To investigate the expression of *AdBBX* genes in different *Arachis duranensis* tissues, RNA-seq datasets were downloaded from peanut genome database (https://peanutbase.org/gene_expression/atlas) and used for transcription analysis. Expressions of *AdBBX* genes were investigated in these 22 different tissues and the related data was used for the construction of the heatmap using MeV 4.9.0 (Multiple Experiment Viewer) [37].

Results

Identification of *BBX* genes in *Arachis duranensis*

To identify *BBX* genes in *Arachis duranensis*, we used the amino acid sequences of the BBX conserved domain (PF00643) and 32 Arabidopsis BBXs as blast queries against peanut genome database, and then

we used Pfam and Pro Scan program to confirm the conserved BBX domains in these candidate genes. A total of 24 *BBX* genes were found in wild peanut *Arachis duranensis*. The characteristics of these *AdBBX* genes were investigated and the detailed information was listed (Table 1). *AdBBX* genes exhibited diversities in genomic lengths, CDS lengths, amino acid numbers, isoelectric points, molecular weights and GC contents (Table 1). The genomic lengths of *AdBBX* genes changed from 467 (*AraduF08JS*) to 4563bp (*AraduBV95P*), the CDS ranged from 468 (*AraduF08JS*) to 1641bp (*AraduVV0J1*), and the numbers of deduced amino acids varied from 155 to 546 (Table 1). The GC contents represent stability of the genes at some degree, thus we investigated GC contents in these *AdBBX* genes. The GC contents in *AdBBX* genes ranged from 32.13% (*Aradu1V7PF*) to 52.72% (*Aradu28KTI*). In addition, the isoelectric points of the AdBBX proteins were predicted and varied from 4.16 (AraduF08JS) to 8.85 (Aradu28KTI), and the molecular weight changed from 17020.4 (AraduF08JS) to 61012.69 (AraduVV0J1).

Chromosomal distribution analysis of *AdBBXs*

To investigate the chromosomal localizations of *AdBBXs*, we mapped these genes to peanut genome database to obtain their related physical positions. The *AdBBX* genes were named from *AdBBX1* to *AdBBX24* based on the chromosomal distributions (Fig. 1, Table 1). Among these 24 *AdBBX* genes, 13 members were located on the plus strand and 11 were located on minus strand (Table 1). 9 of the 10 chromosomes were found to contain *AdBBX* genes, except for chromosome 2 (Fig. 1 and Table 1). Chromosome 3 contained the most *AdBBX* genes, with 6 *AdBBX* members, followed by chromosome 6, 7, 9 and 10, with 3 *AdBBXs* on each, respectively. The *AdBBX2, AdBBX3* and *AdBBX4* were located closely in chromosome 3, as well as *AdBBX5* and *AdBBX6*. In addition, *AdBBX20* and *AdBBX21* also distributed closely together on chromosome 9. Among these *AdBBX* genes, most of them were located in the chromosome arms, while only three genes *AdBBX14, AdBBX17* and *AdBBX18* distributed in the middle part of related chromosomes.

Protein sequence and classification analysis of *AdBBX* genes

The BBX proteins were classified into five subgroups on the basis of the conserved domains they contained, including the types and numbers of BBX and CCT domains [3, 4]. We then analyzed the conserved domains in these 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 of these conserved domains, the logos of *Arachis duranensis* B-box1 ($CX_2CX_8CX_2DXAXLCX_2CDX_2VHX_2NXLX_3H$, 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 analyzed using Weblogo [31] (Fig. 2). In addition, the amino acid sequences of the B-box1, B-box2 and CCT domains were also aligned to analyze the correspondence positions of the conserved amino acid sequences (Fig. 3).

To further investigate the evolutionary relationship of these *AdBBXs*, we created a phylogenetic tree using their amino acid sequences (Fig. 4a). Conserved domain analysis revealed that 16 of the 24 AdBBX proteins had two BBX domains, B-box1 and B-box2, 11 members contained one CCT domain, and 7 genes

had both BBX and CCT domains (Fig. 4b). The AdBBX proteins were also grouped into five subfamilies based on the diversity of the conserved domains (Fig. 4c). Group I and II, the difference between which was the diverse of B-box2 domain, had a B-box1, a B-box2 and a CCT domain, and contained 3 and 4 members in *Arachis duranensis*, respectively. Class III subfamily had a B-box1 and a CCT domain and contained 4 members. The group IV subclass was considered to contain two B-box domains, a B-box1 and a B-box2, and had the most members, 9 *AdBBX* genes. Most of the *AdBBX* genes in the same clade were clustered together, except for *AdBBX5* (*AraduJ5IAH*), which had a distinct relationship with other class III genes, but had a closer relationship with class V subfamily members (Fig.4a). To obtain some information from the well-studied *BBX* genes in other species, we analyzed the evolutionary relationship using *BBX* genes from Arabidopsis, rice and *Arachis duranensis* (Additional file 1: Figure S1). Phylogenetic analysis revealed that *AdBBX8* and *AdBBX24* were clustered together with the well-studied flowering time genes *At5g15840* (*CO*) and *Os06g16370* (*Hd1*) in Arabidopsis and rice, respectively [4], indicating both or either of these two genes may play important roles in flowering time regulation in *Arachis duranensis*.

Gene structures and conserved motifs of *AdBBX* genes

To investigate the exon-intron organizations, the genomic and CDS sequences of *AdBBXs* obtained from peanut genome database were analyzed using Gene Structure Display Server program [30]. The exon numbers of *AdBBX* genes changed from 1 (*AdBBX19* and *AdBBX22*) to 5 (*AdBBX4*), and the intron numbers varied from 0 (*AdBBX19* and *AdBBX22*) to 6 (*AdBBX4* and *AdBBX5*). 9 *AdBBX* genes had both 5' and 3' UTRs, 4 members only contained 3' UTRs, 3 genes had only 5' UTRs and 8 members had no UTR (Fig. 5a). To further investigate the conservation and diversity of AdBBX protein structures, the putative motifs of these genes were predicted using MEME tools. A total of 15 distinct motifs were found in all the AdBBX proteins (Fig. 5b and Additional file 2: Figure S2). Among these motifs, motif 1 and 5 were found in all the AdBBX proteins. The conservation of AdBBX protein structures was observed in genes clustered into the same clades, for example, all the members in subclass I shared 5 motifs, including motifs 1, 2, 3, 4 and 5, and subgroup II members shared 6 motifs, including motifs 1, 2, 4, 5, 14 and 15. In addition, structure diversity was also found among these AdBBX proteins. The motif numbers in AdBBX proteins varied from 2 (AdBBX19) to 7 (AdBBX7, AdBBX10 and AdBBX21), and some motifs were only found in specific AdBBX proteins, for example, motif 2 is specific to subgroups I, II and III members and was considered to be the CCT domain, and motif 15 was only found in subclass II members.

Duplication analysis of *BBX* genes in *Arachis duranensis*

Polyploidy is common in flowering plants during evolution and produces many duplicated gene pairs. The wild peanut *Arachis duranensis* was considered to experience one round of duplication [24-28]. We then investigated the duplication events of *AdBBX* genes in *Arachis duranensis* and found 4 interchromosomal, but no tandem, duplicated gene pairs (*AdBBX2/AdBBX16*, *AdBBX2/dBBX17*, *AdBBX3/AdBBX15* and *AdBBX4/AdBBX9*) (Fig. 6). Chromosome 3 had the most duplicated genes, with 3 members, followed by chromosome 7, with 2 duplicated members. In contrast, *AdBBX* genes located on

chromosome 1, 4, 6, 9 and 10 have no duplication events. Moreover, all the duplicated gene pairs were found to belong to group IV, while no duplicated gene pairs were found in the other 4 groups.

Analysis of *cis*-acting elements in *AdBBX* promoter regions

Cis-acting elements in promoter regions display critical roles in regulating gene expressions in plants. To further understand the expression responses of *AdBBX* genes, we analyzed the *cis*-acting elements in *AdBBX* promoter regions 2kb upstream of initiation codon using PlantCARE [36]. A total of 95 kinds of *cis*-acting elements were obtained and 53 types were predicted to have putative functions, including 7 development related elements, 5 environmental stress related elements, 4 site-binding related elements, 9 hormone-responsive elements, 4 promoter related elements and 24 light-responsive elements (Additional file 3: Table S1). The binding sites related to development, including circadian control, metabolism regulation, stem expression, seed-specific regulation, differentiation of cells and cell cycle regulation (Fig. 7a), environmental stress, such as anaerobic, drought, low temperature, and defense and stress related *cis*-acting elements (Fig. 7b), and hormones, containing abscisic acid (ABA), gibberellic acid (GA), auxin and jasmonic acid (MeJA) related elements (Fig. 7c), were obtained in these promoters. In addition, the numbers and types of *cis*-acting elements varied in these *AdBBX* promoters, indicating their functional diversities in plant development regulation (Table 2). Among these putative functional elements, all the *AdBBX* genes contained light-responsive elements, which were represented to be the most abundant type in each of the *AdBBX* promoters, hormone-responsive elements and promoter related elements (Table 2 and Additional file 3: Table S1), suggesting these genes shared some common pathways involving in plant development regulation. The promoter-related elements CAAT-box and TATA-box were found in all the *AdBBX* promoter regions, which might be the basic components for the promoters. Moreover, the light-responsive element Box4 was obtained in 23 *AdBBX* promoters, except for *AdBBX13*, indicating that *AdBBX* genes play important roles in Box4-mediated light response regulation pathways.

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 these 24 *AdBBX* genes in 22 different tissues (Fig. 8). *AdBBX* genes showed distinct transcription patterns in the tested tissues, indicating the functional diversity of these genes. For example, *AdBBX1* and *AdBBX23* were expressed at high levels in most of the tissues. In contrast, *AdBBX3*, *AdBBX5*, *AdBBX15*, *AdBBX18* and *AdBBX24* showed low expression levels in all these tissues, which might be consistent with their functions during plant development in these tissues. Moreover, *AdBBX21* was expressed at high levels specifically in nodule roots, but showed low abundance in other tissues, including root, indicating *AdBBX21* might be involved in the formation of nodule in *Arachis duranensis*. *CO* homologous genes are key factors in regulating flowering time in many species, thus the expression levels of *Arachis duranensis* orthologues of Arabidopsis *CO* and rice *Hd1* genes, *AdBBX8* and *AdBBX24*, were investigated in these tissues. *AdBBX8* showed high expression level in leaves, flowers, pistils and aerial gyn Ti, while *AdBBX24* exhibited low levels in all the tissues. In addition, expressions of duplicated gene pairs were also analyzed in these tissues. Some duplication events showed similar expression

patterns in these tissues. For example, the duplicated gene pair *AdBBX3/AdBBX15* was expressed almost the same abundance in all these tissues (Fig. 8), indicating the functional conservation of the duplicated genes. In contrast, some duplication events showed distinct expression levels in some tissues. For example, *AdBBX9* showed high expression abundance in leaves and roots, while its duplicated gene pair *AdBBX4* exhibited low expression level in these tissues (Fig. 8).

Discussion

In the past decades, the characterization of BBX genes, such as Arabidopsis *CO* and rice *Hd1*, from many species have greatly increased our knowledge about the molecular mechanisms involving in plant development. Peanut is an important crop in the world and provide essential oil for our daily life, thus the investigation of peanut *BBX* genes is of great help for understanding and improvement of peanut development. In this study, we identified and characterized 24 BBX proteins from the wild peanut *Arachis duranensis* and carried out comprehensive analysis of these genes.

BBX genes changed during plant evolution and the numbers and types of *BBXs* varied in different species [3-6, 38]. For example, wild peanut *Arachis duranensis*, Arabidopsis, rice and pear contained 24, 32, 30 and 25 *BBX* members, respectively, and group IV contained the most *BBX* genes among these five subclasses in each of these species (Additional file 4: Table S2). The genome sizes of diploid wild peanut *Arachis duranensis* [24, 39], Arabidopsis [40], rice [41] and pear [42] were 1.25 GB, 125Mb, 403Mb and 512 Mb, respectively. Thus the genome sizes have no directly relationship with the numbers of *BBX* members in these plants. In addition, the genes containing both BBX and CCT domains were designated as CO or CO-like (COL) proteins, and many *CO*-like genes (*COL*) were considered to be involved in circadian clock or flowering time regulation in Arabidopsis [4]. Approximately half of the BBX proteins were clustered to be *CO* or *COL* members (Group I, II and III members) in plants (Fig. 4), such as Arabidopsis (53.13%), rice (56.67%), pear (44%) and wild peanut *Arachis duranensis* (45.83%), indicating the evolution of *CO* and *COL* genes is conserved in these plants.

The c*is*-acting elements in promoter regions are responsible for the transcription of genes and the variety of the types and numbers of *cis*-acting elements in the promoter regions results in the difference of gene responses. *AtBBX* genes have been reported to participate in the regulation of many pathways, such as flowering time, circadian clock, abiotic stress and photomorphogenesis [3, 4]. Different numbers and types of *cis*-acting elements were found in these *AdBBX* promoter regions, indicating the functional diversity of these genes. Many *BBX* genes in Arabidopsis were found to involve in light input signal pathways [4], and the light responsive elements were also found to be the most abundance one in each of these *AdBBX* promoters, indicating these *AdBBXs* might work in response to light-dependent regulation pathways to contribute to plants' survival and adaption. Moreover, many *cis*-acting elements were also obtained from promoter regions of the low expressed genes, including *AdBBX3*, *AdBBX5*, *AdBBX15*, *AdBBX18* and *AdBBX24* (Fig. 8). There are many factors, but not only *cis*-acting elements, affecting gene expression in plants. For example, the epigenetic modification and somatic genome variations were

considered to change gene expression in many organisms [43]. Whether the low expressed genes were affected by these factors still need further investigation.

CO is an important factor involving in the regulation of flowering time in Arabidopsis and expressed highly at the apex of the seedlings and young leaves [44]. *CO* accelerates flowering time via activating the transcription of a RAF-kinase-inhibitor-like protein FT. *AdBBX8* and *AdBBX24* were the close homologous genes of *CO* in *Arachis duranensis* (Additional file 1: Figure S1). Soybean *CO* ortholog, *GmCOL1a*, *GmCOL1b*, *GmCOL2a* and *GmCOL2b*, were shown to be involved in flowering time regulation [45]. Genes evolved from the same origin might have similar functions, thus we investigated synteny relationships of *CO* orthologs/homologs from soybean (*GmCOL1a*, *GmCOL1b*, *GmCOL2a* and *GmCOL2b*) and *Arachis duranensis* (*AdBBX8* and *AdBBX24*), respectively (Additional file 5: Figure S3). Synteny analysis revealed that *AdBBX8* had closer relationships with soybean *GmCOL1a* and *GmCOL1b* than *AdBBX24*. In contrast, *AdBBX8* and *AdBBX24* showed similar close relationships with soybean *GmCOL2a* and *GmCOL2b*. Moreover, *AdBBX8* was expressed highly in leaves, flowers, pistils and aerial gyn Ti, however, *AdBBX24* exhibited extremely low expression levels in all the tissues (Fig. 8), indicating *AdBBX8* might be the key factor acting a similar role as *CO* in flowering time regulation and *AdBBX24* might be a redundant gene and lost functions during evolution. In addition, *CO* is regulated by circadian clock and its expression changed during the day [46], and *AdBBX24* might be expressed at other time of the day rather than the tested time. Much work still need to do to investigate how *AdBBX8* and *AdBBX24* work in flowering time regulation.

Gene duplication produced new genes during evolution in many species. Some duplicated genes lost functions and some duplication events evolved new functions during gene duplication, compared to their origin genes [47, 48]. 4 duplicated gene pairs were found in *Arachis duranensis* and all these duplication events belonged to group IV subfamily, which contained only two BBX domains (Fig. 6), making group IV the largest subfamily in these groups. In addition, duplication events showed different exon-intron structures for each of these duplicated gene pairs (Fig. 5a) and *cis*-acting elements varied in the promoter regions of each of these duplicated gene pairs (Table 2), indicating the functional differentiation of these gene pairs during evolution. Moreover, the duplicated gene pairs *AdBBX2/dBBX17*, *AdBBX3/AdBBX15* and *AdBBX4/AdBBX9* contained similar motifs (Fig. 5b), and the expression of some duplication events showed similar levels in some tissues (Fig. 8), such as *AdBBX3/AdBBX15*, and thus they might remained some origin functions and participate in some common pathways.

Conclusions

In our present study, we identified and characterized 24 *BBX* genes from a wild peanut *Arachis duranensis*. The characteristics, such as conserved domains, gene structures, phylogenetic relationships, chromosomal distributions, gene duplications, synteny relationships, *cis*-acting elements and gene expressions, were investigated. *AdBBX* members distributed on 9 of the 10 chromosomes and were classified into five clades based on the diversity of the conserved BBX and CCT domains. 4 interchromosomal duplicated gene pairs were found, all of which belonged to group IV. Moreover, a total

of 95 kinds of *cis*-acting elements were obtained and 53 types were predicted to have putative functions. The potential functions of *AdBBX* genes in various tissues were predicted based on their expressions and the results indicated *AdBBX8* might be a key factor involving flowering time regulation. Our results will not only be useful for the understanding of *AdBBX* genes, but also provide essential information for further functional analysis of these members.

Abbreviations

BBX, B-box; CCT, CONSTANS, CONSTANS-LIKE and TIMING OF CAB1; LD, Long day; SD, Short day; DNA, Deoxyribonucleic acid; RNA, Ribonucleic acid; CDS, Coding domain sequence; UTR, Untranslated Regions; AA, Amino acid; 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; MeJA, Auxin and jasmonic acid; RE, Responsive elements; Ad, Arachis duranensis; Gm, Glycine max; Os, Oryza sativa; pI, Isoelectric point; MW, Molecular weight.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Not applicable.

Funding

This work was supported by the National Natural Science Foundation of China (grant 31601372), and the Qingdao Agricultural University Scientific Research Foundation (grant 6631115039, 6631119010). The funders have no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Author contributions

SL conceived and designed the research. HJ, MX and CC conducted the experiments and analyzed the data. SL wrote the manuscript. All authors read and approved the manuscript.

References

1. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science. 2000; 290 (5499):2105-2110.

2. Kielbowicz-Matuk, A. Involvement of plant C2H2-type zinc finger transcription factors in stress responses. Plant Sci. 2012; 185-186:78-85.

3. Khanna R, Kronmiller B, Maszle DR, Coupland G, Holm M, Mizuno T, Wu SH. The Arabidopsis *B-Box* zinc finger family. Plant Cell. 2009; 21(11):3416-3420.

4. Gangappa SN, Botto JF. The *BBX* family of plant transcription factors. Trends Plant Sci. 2014; 19(7):460-470.

5. 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(10):e48242.

6. 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(1):156.

7. Shalmani A, Jing XQ, Shi Y, Muhammad I, Zhou MR, Wei XY, Chen QQ, Li WQ, Liu WT, Chen KM. Characterization of *B-BOX* gene family and their expression profiles under hormonal, abiotic and metal stresses in *Poaceae* plants. BMC Genomics. 2019; 20(1):27.

8. Massiah MA, Simmons BN, Short KM, Cox TC. Solution structure of the RBCC/TRIM B-box1 domain of human MID1: B-box with a RING. J Mol Biol. 2006; 358(2):532-545.

9. Massiah MA, Matts JA, Short KM, Simmons BN, Singireddy S, Yi Z, Cox TC. 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):1-10.

10. Crocco CD, Botto JF. BBX proteins in green plants: Insights into their evolution, structure, feature and functional diversification. Gene. 2013; 531(1):44-52.

11. Griffiths S, Dunford RP, Coupland G, Laurie DA. The evolution of *CONSTANS*-like gene families in barley, rice, and Arabidopsis. Plant Physiol. 2003; 131(4):1855-1867.

12. 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(6):847-857.

13. Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G. Distinct roles of *CONSTANS* target genes in reproductive development of Arabidopsis. Science. 2000; 288(5471):1613-1616.

14. Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science. 2004; 303(5660):1003-1006.

15. Tiwari SB, Shen Y, Chang HC, Hou Y, Harris A, Ma SF, McPartland M, Hymus GJ, 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(1):57-66.

16. 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(11):2971-2984.

17. 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(3):462-476.

18. 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(12):2473-2483.

19. 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(8):1487-1499.

20. Komiya R, Ikegami A, Tamaki S, Yokoi S, Shimamoto K. *Hd3a* and *RFT1* are essential for flowering in rice. Development. 2008; 135(4):767-774.

21. 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(20):3443-3450.

22. Kochert G, Halward T, BranchWD, Simpson CE. RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor Appl Genet. 1991; 81(5):565-570.

23. Moretzsohn MC, Gouvea EG, Inglis PW, Leal-Bertioli SC, Valls JF, Bertioli DJ. 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.

24. Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EK, 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(4):438-446.

25. Chen X, Li H, Pandey MK, 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. P Natl Acad Sci USAS. 2016; 113(24):6785-6790.

26. Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, Leal-Bertioli SCM, Ren L, Farmer AD, Pandey MK, et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nat Genet. 2019; 51(5):877-884.

27. 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. 2019; https://doi.org/10.1016/j.molp.2019.03.005.

28. Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang WC, 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(5):865-876.

29. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; 33(7):1870-1874.

30. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0. an upgraded gene feature visualization server. Bioinformatics. 2015; 31(8):1296-1297.

31. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004; 14(6):1188-1190.

32. Zhang D, Zhao M, Li S, Sun L, Wang W, Cai C, Dierking EC, Ma J. Plasticity and innovation of regulatory mechanisms underlying seed oil content mediated by duplicated genes in the palaeopolyploid soybean. Plant J. 2017; 90(6):1120-1133.

33. Li S, Ding Y, Zhang D, Wang X, Tang X, Dai D, Jin H, Lee SH, Cai C, Ma J. Parallel domestication with a broad mutational spectrum of determinate stem growth habit in leguminous crops. Plant J. 2018; 96(4):761-771.

34. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19(9):1639-1645.

35. 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 (*Ziziphus jujuba* Mill.). Planta. 2019; 249(3):815-829.

36. 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(1):325-327.

37. Saeed AI, 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(2):374-8.

38. 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(2):303-315.

39. Samoluk SS, Robledo G, Podio M, Chalup L, Ortiz JP, Pessino SC, Seijo JG. First insight into divergence, representation and chromosome distribution of reverse transcriptase fragments from L1 retrotransposons in peanut and wild relative species. Genetica. 2015; 143(1):113-125.

40. Initiative AG. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 2000; 408(6814):796-815.

41. Yu J, Hu S, Wang J, Wong GK-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(5565):79-92.

42. Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H, et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). Genome Res. 2013; 23(2):396-408.

43. Li XQ, Xing T, Du D. Identification of top-ranked proteins within a directional protein interaction network using the pagerank algorithm: applications in humans and plants. Curr Issues Mol Biol. 2016; 20:13-28.

44. 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.

45. Wu F, Price BW, 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(1):e85754.

46. 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.

47. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. Selection in the evolution of gene duplications. Genome Biol. 2002; 3(2):RESEARCH0008.

48. 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(2):323-336.

Figures, Tables And Supplementary Materials

Figure legends

Fig. 1 Chromosomal distribution of *AdBBX* genes. Chromosome number and length are represented.

Fig. 2 The motif logos of B-box1, B-box2 and CCT domains. The x-axis indicates the conserved sequences of these domains and the height of each letter indicates the conservation of each residue.

Fig. 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 the identical and similar amino acids were shaded.

Fig. 4 Phylogenetic and structure analyses *AdBBX* genes. (a) The Phylogenetic tree was generated by MEGA 7 using Neighbor-Joining methods. The yellow, red, blue, purple and green colored gene names indicate group I, II, III, IV and V members, respectively. (b) The positions of conserved domains in each AdBBX proteins were represented, the green, pale brown and cyan boxes indicate B-box1, B-box2 and CCT domains, respectively. (c) The classification of *AdBBX* members and positions of each domain.

Fig. 5 Analyses of exon-intron organizations and conserved motifs of AdBBX proteins. (a) The exon-intron organizations of *AdBBX* genes, the purple, orange and black boxes indicate the UTRs, exons and introns, respectively. (b) The motifs of AdBBX proteins identified by MEME tools. The colored boxes 1-15 indicate different motifs.

Fig. 6 Duplication analysis of *AdBBX* genes. The chromosomes are indicated by different colors and the duplicated gene pairs were marked with green lines.

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

Fig. 8 Transcription patterns of *AdBBX* genes in 22 different tissues. The tissues including seedling leaves (seedling leaf 10 day after emergence), main stem leaves, lateral stem leaves, vegetative shoot tip (vegetative shoot tip from the main stem), reproductive shoot tip (reproductive shoot tip from the first lateral leaf), roots (10 day-old roots), nodule roots (25-day old nodules), flowers, pistils, stamens, aerial gyn tip (aerial gynophore tip), sub gyn tip (subterranean gynophore tip), podpt1 (pattee stage 1 pod), stalkpt1 (pattee stage 1 stalk), podpt3 (pattee stage 3 pod), pericarp pattee 5 (pattee stage 5 pericarp), seed pattee 5 (pattee stage 5 seed), pericarp pattee 6 (pattee stage 6 pericarp), seed pattee 6 (pattee stage 6 seed), seed pattee 7 (pattee stage 7 seed), seed pattee 8 (pattee stage 8 seed), seed pattee10

(pattee stage 10 seed) were used for gene expression analysis. The yellow, red, blue, purple and green colored gene names indicate group I, II, III, IV and V members, respectively.

Tables

Due to technical limitations, tables 1 and 2 are only available as downloads in the supplemental files section.

Supplementary material

Additional file 1: Figure S1. Evolution relationship analysis of AdBBX proteins. Amino acid sequences of BBX proteins from *Arachis duranensis*, rice and Arabidopsis were used to generate the phylogenetic tree with MEGA 7.0 using Neighbor-Joining method.

Additional file 2: Figure S2. Sequence logos of 15 distinct motifs in 24 AdBBX proteins.

Additional file 3: Table S1. Function analysis of *cis*-acting elements in *AdBBX* promoter regions. The classifications, names and putative functions of related *cis*-acting elements are predicted and listed.

Additional file 4: Table S2. The numbers and types of BBX proteins in Arabidopsis, rice, pear and wild peanut *Arachis duranensis*.

Additional file 5: Figure S3. Synteny analysis of *CO* orthologs/homologs in soybean and *Arachis duranensis*. The putative orthologous genes surrounding *CO* orthologs/homologs from soybean (*GmCOL1a*, *GmCOL1b*, *GmCOL2a* and *GmCOL2b*) and *Arachis duranensis* (*AdBBX8* and *AdBBX24*) were identified by BLASTP search. Synteny between *AdBBX8*, *AdBBX24* and *GmCOL1a* (a), *GmCOL1b* (b), *GmCOL2a* (c), *GmCOL2b* (d) were shown, respectively. The red boxes indicate our target genes and the green boxes indicate genes surrounding *CO* orthologs/homologs. Gm, *Glycine max*.

Supplementary Files

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

- Figures.pdf
- Additionalfile4TableS2.xlsx
- Table2.jpg
- Additionalfile5FigureS3.pdf
- Additionalfile1FigureS1.pptx
- Table1.jpg
- Additionalfile2FigureS2.pptx
- Additionalfile3TableS1.xlsx