Gene 531 (2013) 44–52

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

BBX proteins in green plants: Insights into their evolution, structure, feature and functional diversification

Carlos D. Crocco [1], Javier F. Botto *

IFEVA, Facultad de Agronomía, Universidad de Buenos Aires y Consejo Nacional de Investigaciones Científicas y Técnicas, Av. San Martín 4453, C1417DSE, Ciudad de Buenos Aires, Argentina

ARTICLE INFO

Article history:
Accepted 10 August 2013
Available online 27 August 2013

Keywords:
BBX proteins
B-box domains
CCT domain
Nuclear localization signals (NLSs)
Plant evolution

ABSTRACT

The B-box domain is conserved in a large number of proteins involved in cell growth control, differentiation and transcriptional regulation among animal and plant species. In *Arabidopsis thaliana*, some works have found that B-box proteins (BBX) play central developmental functions in flowering, light and abiotic stress signaling. Despite the functional importance of this protein family, evolutionary and structural relationships of BBX proteins have not been extensively investigated in the plant kingdom. Using a phylogenetic approach, we conducted a comprehensive evolutionary analysis of the BBX protein family in twelve plant species (four green algae, one moss, one lycophyte, three monocots and three dicots). The analysis classified 214 BBX proteins into five structure groups, which evolved independently at early stages of green plant evolution. We showed that the B-box consensus sequences of each structure groups retained a common and conserved domain topology. Furthermore, we identified seven novel motifs specific to each structure group and a valine–proline (VP) pair conserved at the C-terminus domain in some BBX proteins suggesting that they are required for protein–protein interactions. As it has been documented in mammalian systems, we also found monopartite and bipartite amino acid sequences at the C-terminus domain that could function as nuclear localization signals (NLSs). The five BBX structure groups evolved constrained by the conservation of amino acid sequences in the two B-boxes, but radiating variation into NLSs and novel motifs of each structural group. We suggest that these features are the functional basis for the BBX protein diversity in green plants.

1. Introduction

The B-box domain is found in more than 1500 proteins of multicellular species and some unicellular eukaryotes (Meroni and Diez-Roux, 2005). In animals, the B-box is often found together with RING finger and coiled-coil domains, forming tripartite motif proteins named TRIM/RBCC. The TRIM/RBCC family includes a large number of proteins involved in diverse cellular processes like apoptosis, cell cycle regulation and viral response (Meroni and Diez-Roux, 2005). The B-box of some TRIM/RBCC proteins functions as a protein–protein interaction domain and is required for substrate recognition (Meroni and Diez-Roux, 2005; Torok and Etkin, 2001). Other functions of the B-box domain involve localization in nuclear bodies and transcriptional regulation (Beenders et al., 2007; Borden et al., 1996). This domain is present in the N-terminus of the proteins as a single B-box or tandem repeats designated as B-box1 (B1) and B-box2 (B2) (Massiah et al., 2006, 2007; Short and Cox, 2006). Despite the functional importance of TRIM/RBCC proteins in animals, they are absolutely absent in plants.

In plants, B-box (BBX) proteins regulate plant development. For example, CO/AtBBX1 is a central player in the photoperiod control of flowering in *Arabidopsis thaliana* plants (Onouchi et al., 2000; Samach et al., 2000). *co* mutants flower late only under long days, whereas *CO*-over-expressing plants of *A. thaliana* flower early in both long and short days (Putterill et al., 1995; Suarez-Lopez et al., 2001). Other AtBBX proteins with double B-box and CCT domains, such as COL3/AtBBX4 and COL9/AtBBX7, also regulate flowering time (Cheng and Wang, 2005; Datta et al., 2006). Besides, AtBBX proteins with only B-box domains are involved in the control of plant development by light and other abiotic stress factors. The characterization of *A. thaliana* mutants has revealed that DBB1b/AtBBX19, STO/AtBBX24 and STH/AtBBX25 are negative regulators (Datta et al., 2006; Gangappa et al., 2013, in press; Indorf et al., 2007; Kumagai et al., 2008), while DBB1a/AtBBX18, STH2/AtBBX21 and STH3/AtBBX22 act as positive regulators of seedling de-etiolation processes (Datta et al., 2007, 2008; Kumagai et al., 2008). Very recently, experiments in shaded environments have demonstrated that STH2/AtBBX21 is also involved in the fine-tuning of shade avoidance responses (Crocco et al., 2010, 2011). Additionally, STO/AtBBX24 is implicated in saline stress and UV-B photomorphogenesis responses (Indorf et al., 2007; Jiang et al., 2012;

Abbreviations: BBX, B-box proteins; VP, valine–proline; NLS, nuclear localization signal; CO, CONSTANS; COL, CO-like.

* Corresponding author. Tel.: +54 11 4524 8070; fax: +54 11 4514 8730.
E-mail address: botto@agro.uba.ar (J.F. Botto).
[1] Present address: Department of Botany and Plant Biology, University of Geneva, Sciences III, 30 Quai E. Ansermet, 1211 Geneva 4, Switzerland.

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.08.037

Nagaoka and Takano, 2003). Some BBX proteins can also act as helpers of other BBX proteins. For example, Holtan et al. (2012) have recently demonstrated that AtBBX32 is a transcriptional modulator of STH2/AtBBX21 action in the light signaling pathway.

The BBX proteins in plants are characterized by having one or two B-box domains in the N-terminus, and, in some cases, a CCT domain in the C-terminus (Griffiths et al., 2003; Robson et al., 2001). Recently, a phylogenetic analysis performed in *A. thaliana* has found that BBX proteins are encoded by a gene family of 32 members named from AtBBX1 to AtBBX32 (Khanna et al., 2009). This foundational work classified AtBBX proteins into five structure groups (I–V). The AtBBX members of structure group I (AtBBX1 to AtBBX6) contain two B-boxes in tandem (B1 and B2) and a CCT domain, being the B1 domain located N-terminal to the B2 domain and separated by 5 to 20 residues. The AtBBX members of structure group II (AtBBX7 to AtBBX13) contain B1, B2, and CCT domains, as those of structure group I, with some differences at their consensus sequences of the B2 domain (Chang et al., 2008). The AtBBX members of structure group III (AtBBX14 to AtBBX17) contain a single B-box domain in association with a CCT domain. The B-box proteins of structure group IV (AtBBX18 to AtBBX25) have the B1 and B2 domains, but not the CCT domain, whereas the protein members of structure group V (AtBBX26 to AtBBX32) carry a single B-box domain (Khanna et al., 2009). Clustering of BBX proteins into five structure groups was also supported by a recent study in rice (Huang et al., 2012).

Although the study of BBX proteins in plant development is a growing area of recent interest, the evolutionary relationships of the members of without "the" BBX family among different species have not yet been studied. Thus, in the present work, we carried out phylogenetic and structural analyses using the sequence information of 214 BBX proteins that belong to twelve green plant species (four green algae, one moss, one lycophyte, three monocot and three dicot species). Based on the evolutionary origins and structural protein changes, we discuss the functional diversity of BBX proteins of land plants.

2. Materials and methods

2.1. Identification of sequences and domains of BBX proteins across green plants

The KEGG (http://www.genome.ad.jp/kegg/) nucleotide and protein sequence databases for fully sequenced genomes were scanned for proteins related to AtBBX described in *Arabidopsis* (Khanna et al., 2009), using either BLASTp (protein databases) or tBLASTn (nucleotide databases). After initial sequence collection, the full BBX proteins of each species were used for the classification of protein sequences into paralog and ortholog clusters using sequence similarity profiles of KEGG/SSDB. This allowed enhancing the chances of finding sequences related to particular divergent BBX proteins and of better defining the range of organisms containing these proteins. The amino acid sequence data and the information of B-box domain locations were taken from KEGG, together with the annotations made by the original authors, SWISS-PROT, and KEGG itself. Each predicted BBX protein sequence was confirmed by Pfam search for the presence of a B-box signature (E-value < 0.05). Only 23 of the 30 OsBBX proteins described in rice by Huang et al. (2012) have a B-box signature with an E-value < 0.05, while the 32 AtBBX proteins described in *Arabidopsis* by Khanna et al. (2009) fulfill this criteria. The complete amino acid sequence of all BBX proteins can be found in Supplemental Data 1. The BBX proteins were renamed according to the result of the phylogenetic analysis for each species (Supplemental Data 2, 3).

2.2. B-box consensus sequences

The B-box motifs that belong to each structure group were detected with GLAM2 software (Frith et al., 2008), using the extracted B-box sequences of each BBX protein predicted by Pfam (E-value < 0.05). As some BBX proteins have a double B-box domain, we used the following criteria to designate them: the first B-box that appeared within the protein in the N-terminal position was called B1 and the second B-box termed B2.

2.3. Phylogenetic analysis

The alignment of the full length BBX protein and B-box domain sequences was performed in ClustalW (Thompson et al., 1997) using standard settings (Gonnet weight matrix, gap opening = 10 and gap extension = 0.2) and was adjusted by visual inspection. The full-length amino acid sequence alignment of BBX proteins is documented in Supplemental Data 4 and the B-box domain alignment can be found in Supplemental Data 6. The model of protein evolution that best fits the protein sequence data was selected using the program MEGA 5.01 (Tamura et al., 2011). The best-scoring model for the full-length BBX protein alignment was the Dayhoff probability model with rate variation among sites calculated as a gamma distribution (+G), and the best-scoring model for the B-box domain alignment was JTT probability model with rate variation among sites calculated as a gamma distribution (+G). Bayesian phylogenetic analyses on aligned full-length BBX sequences were performed with MrBayes v. 3.1.2 setting a MCMC algorithm (Ronquist and Huelsenbeck, 2003). Two independent runs were computed for 2,000,000 generations with a burn-in of 5000 trees in order to reach acceptable standard deviation of split frequencies. Trees were sampled from each chain every 100 generations. In addition, a neighbor-joining (NJ) tree was obtained using MEGA 5.01 (Tamura et al., 2011) (Supplemental Data 5). Maximum likelihood analyses on aligned B1 and B2 sequences were performed with MEGA 5.01 (Tamura et al., 2011). The evolutionary distances were computed using the JTT matrix-based method with a gamma distribution (shape parameter = 1). The bootstrap consensus tree inferred from 7000 replicates (Supplemental Data 7). In addition, a phylogenetic maximum likelihood tree of B1 and B2 sequences was made using the PhyML3.0 algorithm implemented through the web-based interface available at http://www.atgc-montpellier.fr/phyml/, using LG substitution model (Guindon et al., 2010) (Supplemental Data 7). Phylogenetic trees were visualized using the program MEGA 5.01 (Tamura et al., 2011).

2.4. Detection of conserved motifs

The MEME software (Bailey et al., 2009) was used to discover patterns in the complete amino acid sequences of plant BBX proteins. We performed a search between BBX proteins of the same structure group. Each motif was individually checked so that incorrect or insignificant matches were discarded.

3. Results

3.1. Identification and global phylogenetic analysis of BBX proteins in green plants

Previous phylogenetic analyses of BBX proteins were based on the genome sequences of *A. thaliana* (At) and *Oryza sativa* (Os) (Huang et al., 2012; Khanna et al., 2009). These foundational works generated a useful but a limited phylogenetic framework for the classification of BBX proteins in angiosperms. Our work includes twelve species and provides insight about the early diversity of this family in green plants. Besides *A. thaliana* and *O. sativa* species, we extended the phylogenetic analysis to the complete genomes of ten species including four algae [*Volvox carteri* (Vc), *Chlamydomonas reinhardtii* (Cr), *Ostreococcus tauri* (Ot) and *Ostreococcus lucimarinus* (Ol)], one moss [*Physcomitrella patens* (Pp)], one lycophyta [*Selaginella moellendorffii* (Sm)] and four additional angiosperms [*Zea mays* (Zm), *Brachypodium distachyon* (Bd), *Ricinus communis* (Rc) and *Populus trichocarpa* (Pt)]. In addition to 32 and 23 BBX proteins of *A. thaliana* and *O. sativa*, we identified 159 BBX proteins: