Computational Biology and Chemistry 67 (2017) 131–140

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Research Article

Comparative genome-wide phylogenetic and expression analysis of SBP genes from potato (*Solanum tuberosum*)

Musa Kavas*, Aslıhan Kurt Kızıldoğan, Büşra Abanoz

Ondokuz Mayıs University, Faculty of Agriculture, Department of Agricultural Biotechnology, Samsun, Turkey

ARTICLE INFO

Article history:
Received 23 May 2016
Received in revised form 27 July 2016
Accepted 3 January 2017
Available online 4 January 2017

Keywords:
SBP
Potato
Flowering genes
Transcription factors

ABSTRACT

Flowering time is a very important phase in transition to reproductive stage of life in higher plants. SQUAMOSA promoter-binding protein (SBP) gene family encodes plant-specific transcription factors that are involved in regulation of several developmental processes, especially flowering. Although *SBP*-box genes have been identified in different plants, there have been no study indicating the regulatory effect of SBP box in potato flowering. Here, we report for the first time the identification and characterization of SBP-box transcription factors as well as determination of expression level of *SBP*-box genes in *Solanum tuberosum* L. an important crop worldwide. Fifteen different SBP-box transcription factor genes were identified in potato genome. They were found to be distributed in nine chromosomes and eight of them included miR156 and miRNA157 target sites. Characterization of amino acid sequences were performed and protein interactions were predicted. In addition, expression levels of five *S. tuberosum SBP*-box genes were analysed by both *in silico* and qRT-PCR. All these results provide a better understanding of functional role of *SBP*-box gene family members in flowering time in potato.

1. Introduction

Solanum tuberosum is a member of Solanaceae that consists of 2000 species such as tomato, eggplant, petunia and tobacco. It was originated from Andean and Chilean landraces and spread all over the world. It is cultivated mostly for underground tubers, and it is consumed freshly, or as seed tubers for sowing as well as starch, potato crisps and frozen form (french fries) in industry. In addition, there are other areas of potato usage such as glue, animal feed, and fuel-grade ethanol production (Izmirlioglu and Demirci 2015; Slater et al., 2014). Potato is the third most grown crop behind wheat and rice worldwide. *S. tuberosum* is a polyploid crop in various chromosome numbers ranging from diploid (2n) to hexaploid (6n) and the cultivated ones are tetraploid (4n = 48). Instead of reproductive seeds, potato propagates clonally by tuber propagules (Slater et al., 2014). Flower meristems are crucial for fruit and seed formation which are required for efficient breeding and germplasm generation (Bamberg et al., 2014).

Light is a very important abiotic factor that affects plant's growth and development. There are different response ways of plants to light such as duration, intensity, direction and spectral quality. In higher plants, time of flowering is a very important phase in transition to reproductive stage of life. Spectral quality of light period and day length called as photoperiod are primary factors that affect initiation of flowering. The day length exerts its effect on floral transition in different ways. In *Arabidopsis*, flowering is promoted in long days (LD) in comparison to the short days (SD). In rice and Pharbitis nil, flowering is adversely affected in the LDs. However, in some species such as *Solanum lycopersicum*, flowering is independent of photoperiod and it is affected by mostly temperature and other stimuli (Bernier, 1988). In potato, tuber formation is under the control of short light period and cold temperature. Before winter, differentiation occurs in the vegetative propagation organs. The behavior under LDs is highly variable among *S. tuberosum* varieties. For example, *S. tuberosum* cv. Andigena cannot form tubers in LDs. In contrasts, Chilean potato cultivars are more adapted to LD conditions (Navarro et al., 2011). However, in Andigena, the perception of night length (Night break) with SD period causes repression of tuberization in comparison to SD response only (Gonzalez-Schain et al., 2012).

Two main flowering regulators namely CO (Constans) and FLC (FLOWERING LOCUS C) have antagonistic roles in floral induction pathways in Arabidopsis. These genes are activated in the presence of different light and temporal signals (Li et al., 2015). CO, a zinc finger motif containing transcriptional regulator, and acts as floral activator. Whereas the FLC gene encodes a floral repressor-MADS box protein mediating the autonomous and vernalization

* Corresponding author.
E-mail address: musa.kavas@omu.edu.tr (M. Kavas).

http://dx.doi.org/10.1016/j.compbiolchem.2017.01.001
1476-9271/

pathways. Both regulates the expression of FT (Flowering Locus T), a RAF kinase related protein, SUPPRESSOR OF OVEREXPRESSION OF CO 1 (SOC1), and LEAFY (LFY). Their expressions in flowering pathways precisely determine the time of flowering. Besides FLC, CDF1s (cycling dof factor 1), PIF3 (phytochrome-interacting bHLH transcription factor 3), TEM (TEMPRANILLO), SVP (SHORT VEGETATIVE PHASE) are involved in the negative regulation of FT (Kumar et al., 2012). Indeed, MYB transcription factor EFM (early flowering myb protein) represses the expression of FT in leaf, by interacting with factors involved in circadian clock. Therefore, EFM takes role in determination of floral initiation upon exposure to defined temperature and light conditions (Yan et al., 2014). For floral initiation, APETALA1 (AP1) is activated in the establishment of floral development at the shoot apical meristem (Lee and Lee, 2010).

As being a plant specific transcription factor family, SQUAMOSA promoter binding proteins (SBPs) are mainly related to flower development, as well. It is unique in plant kingdom and widespread from green algae to flowering plants. Plant Transcription Factor database (PlantTFDB) includes *SBP*-box genes belong to 65 different organisms (Zhang et al., 2015). SBPs possess a conserved SBP-box domain comprising 76 highly conserved amino acid residues (a GTAC core and gene-specific flanking regions) for DNA binding and nuclear localization, and two zinc binding sites as CysCys-His-Cys and Cys-Cys-Cys-His (Cardon et al., 1999; Song et al., 2016; Yamasaki et al., 2004). SBPs are also among the conserved miRNA targeted plant transcription factors especially for miR156/157 family members (Lakhotia et al., 2014). In Arabidopsis, SBP like (SPL) proteins assemble different signals from different physiological conditions such as photoperiod and age to promote flower formation within the inflorescence (Preston et al., 2016). Under SDs, SPL genes are both negatively and positively regulated by miR156 and SOC1, respectively. In contrast, LD conditions lead positive regulation of SPL by SOC1, FT and FD in leaves. Thereby, this results in activation of FUL, SOC1, AP1 and LFY in the shoot apex to promote flower production. AtSPL3 was shown to act in regulation of floral transition leading to early flowering when constitutively expressed (Cardon et al., 1997; Zhang et al., 2015).

Although several studies have focused on understanding molecular mechanisms of floral development and tuberization in potato, still it is unclear how potato enters flowering upon to light signals. This study aimed at investigation of the molecular mechanisms acting on onset of flowering in potato. Thus, a comprehensive genome-wide comparison has been carried out between early-flowering *S. tuberosum* cv. Atlantic and late-flowering *S. tuberosum* cv. Snowden. The resultant bioinformatic data was supported by qRT-PCR in which expression analysis of the selected flowering related *SBP*-box genes has been carried out. In addition, the characterization of genomic structures, chromosomal locations, promoter analysis, sequence homologies of flowering related transcription factor encoding *SBP*-box genes were performed.

2. Materials & methods

2.1. Sequence retrieval and analysis of SBP transcription factors in potato

The identification of genes coding SBP transcription factors in *Solanum tuberosum* was carried out by two different in silico approaches. In the first approach, SBP encoding amino acid sequences belong to four different plants (*Arabidopsis thaliana, Carica papaya, Oryza sativa subsp. japonica and Solanum lycopersicum)* were downloaded from plant transcription factor database 3.0 <planttfdb.cbi.pku.edu.cn> (Jin et al., 2014). The homologous peptides from potato was found with a BLASTP search at PHYTOZOME v11 database <www.phytozome.net> with default parameters (Goodstein et al., 2012). Secondly, the keyword search was performed with *SBP* in PHYTOZOME v11 database www.phytozome.net>. All identified amino acid sequences were compared with original sequences downloaded from plant transcription factor database 3.0 <planttfdb.cbi.pku.edu.cn> (Jin et al., 2014). The existence of SBP domain in identified amino acid sequences was finally checked with in the Pfam database <http://pfam.sanger.ac.uk> by using CLC Genomics Workbench 8.0 with PFAM module.

The final amino acid sequences of SBP proteins from *Arabidopsis thaliana, Carica papaya, Oryza sativa subsp. japonica* and *Solanum lycopersicum* were aligned in ClustalW by using CLC Genomics Workbench 8.0 software (CLC bio, Aarhus, Denmark) with default parameters. The phylogenetic relationship between these species in terms of SBP proteins was constructed using CLC Genomics Workbench Software via the Neighbor-Joining (NJ) method bootstrap (1000). ITOL v.3 (Letunic and Bork, 2007) was used for displaying and annotation of phylogenetic tree. The conserved motifs in full-length SBP proteins were also determined using the same software.

The online pI/Mw tool of ProtParam (http://web.expasy.org/protparam) was used to predict the molecular weights, stability and isoelectric points of the identified SBP proteins in potato. Subcellular localization of each gene was estimated by CELLO (http://cello.life.nctu.edu.tw/). The number of TM domains was determined using the CLC Genomics Workbench 8.0 software (CLC bio, Aarhus, Denmark). The RPSP program (http://biotech.ou.edu) was used to predict the solubility of recombinant proteins (Wilkinson and Harrison, 1991).

2.2. In silico expression analysis potato transcriptome

Transcriptome data obtained with RNA-seq analysis was downloaded from SRA database, following accession number SRR184103-SRR184104 for Snowden and SRR184099-SRR184100 reads for Atlantic (Hamilton et al., 2011) to evaluate the potato *SBP*-box gene expression patterns. SRA files was converted and splitted into paired end Fastaq reads by using NCBI SRA Toolkit's fastq-dump command. Evaluation of the quality of RNA-seq reads and trimming of low-quality reads (Phred quality (Q) score <20) were carried out (by using CLC Genomics Workbench 8.0). In order to calculate expression levels of *SBP*-box genes, the reads were mapped to *Solanum tuberosum* genome (v3.4), downloaded from PHYTOZOME 11 database (<http://www.phytozome.net>), by using CLC Genomics Workbench with default parameters.

Perfectly aligned sequence reads were used for the estimation of expression pattern of all genes. The Reads per kilobase of exon per million reads mapped (RPKM) was used as a units of expression to normalize our count data (Mortazavi et al., 2008). To identify differentially expressed *SBP*-box genes, a FDR-value $\leq$0.001, fold change (RPKM-tr/RPKM-cont) $\geq$2 and the absolute ratio of log2 (RPKM-tr/RPKM-cont) $\geq$1 were used as threshold values. Finally, the heat maps of hierarchical clustering were visualized with PermutMatrix (Caraux and Pinloche, 2005).

2.3. Prediction of chromosomal location and gene-structure of SBP genes

The exact positions of *SBP*-box genes on potato chromosomes were determined from PHYTOZOME v11 database <www.phytozome.net>. The genes were assigned separately onto twelve potato chromosomes based on their ascending order of physical position (bp), from the short-arm telomere to the long-arm telomere and finally displayed using MapChart v2.2 <www.wageningenur.nl/en/