Biochemical and Biophysical Research Communications 468 (2015) 653–659

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

Genome-wide identification and expression profiling analysis of trihelix gene family in tomato

Chuying Yu [a, 1], Xiaofeng Cai [b, 1], Zhibiao Ye [a], Hanxia Li [a, *]

[a] Key Laboratory of Horticultural Plant Biology, Ministry of Education, Huazhong Agricultural University, Wuhan, 430070, China
[b] Development and Collaborative Innovation Center of Plant Germplasm Resources, College of Life and Environment Science, Shanghai Normal University, Shanghai, 200234, China

ARTICLE INFO

Article history:
Received 28 October 2015
Accepted 3 November 2015
Available online 25 November 2015

Keywords:
Tomato
Trihelix
GT factor
Phylogenetic
Gene expression
Abiotic stresses

ABSTRACT

The trihelix family, classified as GT factors due to their binding specificity for GT elements, constitutes a plant-specific transcription factor family with a conserved trihelix DNA binding domain. In the present study, the comprehensive analysis of 36 putative GT factors was performed in tomato. SlGT members can be classified into six subgroups (GT-1, GT-2, SH4, SIP1, GT-γ and GT-δ). Expression analysis of *SlGT* gene transcripts showed the distinct expression patterns of *SlGT* genes in various tomato organs. All the *SlGT* genes were regulated in response to various abiotic stresses and hormone treatments by the quantitative real-time PCR (qRT-PCR) analysis. Several *SlGT* genes, including *SlGT-27* and *SlGT-34*, were highly regulated by multiple abiotic stresses and phytohormone treatments. Taken together, our results presented here would be providing a useful platform for molecular clone and functional identification of *SlGT* genes in tomato.

1. Introduction

Transcriptional regulation of gene expression plays a pivotal role in influencing or controlling many important biological processes, such as cellular morphogenesis, signaling transduction and environmental stress responses [1]. Transcription factors (TFs) are important regulators to regulate gene expression by binding to plant-specific *cis*-regulatory elements in the promoter region, thereby activating or repressing the transcriptional rates of their target genes [2]. The tomato genome encodes at least 998 TFs of 62 different families [3], accounting for about 2.87% of its estimated a total number of 34, 727 genes [4].

The trihelix family, one of the first TFs discovered in plants, is classified as GT factors due to their binding specificity for GT elements. The DNA-binding domain of GT factor features a typical trihelix (helix-loop-helix-loop-helix) structure that determines the specific binding of GT elements [5,6]. Trihelix TF contain one or two trihelix DNA binding domains, and were divided into five clades, GT-1, GT-2, SH4, GTγ and SIP1 [5]. The diverse function of trihelix TFs were characterized due to the extension of the family members, functions including regulation of light-dependent expression [6–8], responses to abiotic and abiotic stresses [9–14], and to roles in a range of developmental processes involving morphogenesis control of manifold flower organs and leaves [15–17], trichomes [18] and embryos development [19,20].

Tomato is an important fruit plant that serves as a model system for carrying out functional genomics and investigating epigenetic regulation. Despite of the crucial roles of trihelix TFs in plant, the functional of trihelix TFs and systematic analysis of the tomato trihelix family have not been reported yet. In this study, a systemic analysis of 36 trihelix proteins in tomato was performed. The comparison phylogenetic tree was constructed to evaluate the evolutionary relationships of trihelix proteins in tomato and other species. In addition, expression profiling analysis of trihelix family was determined in order to further discern gene functions of this family in normal growth conditions and response to various abiotic stress and hormone treatments using quantitative qRT-PCR.

2. Materials and methods

2.1. Identification of trihelix family members in tomato

The *Arabidopsis* At1g13450 protein was subjected to BLASTN searches in SOL Genomics Network (SGN) (http://solgenomics.net/

* Corresponding author.
E-mail address: hxli@mail.hzau.edu.cn (H. Li).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.bbrc.2015.11.010
0006-291X/

search/loci) to obtain trihelix sequences in database. The conserved domains of tomato trihelix members were confirmed by Pfam (http://pfam.sanger.ac.uk/) and InterProScan (http://www.ebi.ac.uk/Tools/InterProScan/). If there was more than one allele, the longest allele was chosen as representative. The theoretical isoelectric point (pI) and molecular weight (Mw) of tomato trihelix proteins were identified by ProtParam tool (http://web.expasy.org/protparam/).

The exon/intron structure of tomato trihelix genes was generated using GSDS (http://gsds.cbi.pku.edu.cn/) by aligning the cDNA sequences with the corresponding genomic sequences. The distribution of tomato trihelix genes on tomato chromosomes was conducted as described previously [21].

2.2. Identification of conserved motifs and gene duplication in tomato trihelix members

Protein motifs of the tomato trihelix members were identified statistically using MEME program (http://meme.nbcr.net/meme/) [22]. The functional annotation of the identified motifs was implemented by InterProScan. To find potential duplicated tomato trihelix genes, MicroSyn and Circos software was used. The synteny of the *SlGT* genes was calculated by MicroSyn software (http://fcsb.njau.edu.cn/microsyn/) [23]. The result was displayed using Circos software [24].

2.3. Sequence alignment and phylogenetic tree

A multiple alignment analysis was performed with multiple sequence alignment [25] and ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) program, and similarity analysis of the tomato trihelix sequences at amino acid level was performed using ClustalW. The phylogenetic tree were created by MEGA4 program using the neighbor-joining (NJ) method [26]. Images of the phylogenetic trees were also drawn using MEGA4.

2.4. Analysis of expression profile of trihelix genes in tomato various tissues

The expression profile was determined through analyzing the RNA-seq datas based on Locus gene name. The RNA-seq datas were downloaded from Tomato Functional Genomics Database (http://ted.bti.cornell.edu/cgi-bin/TFGD/digital/home.cgi). Only genes with at least one average RPKM value from all 11 tissues $\geq$2 in this study were considered to be expressed. A hierarchical cluster was created using the Cluster 3.0 and heat map generated using TreeView Version 1.60 software [27].

2.5. Plant growth conditions and real-time qRT-PCR analysis

Tomato seeds (*S. habrochaites* LA1777) were germinated and grown in a greenhouse under a photosynthetic photon flux density of approximately 120 μmol photons/m^2/s with a 12-h light/dark photoperiod. The temperature was maintained at 25 °C ± 2 °C, and the relative humidity was controlled at approximately 60%. For abiotic stress and hormone treatments, six weeks old seedlings were treated with drought, salinity (400 mM NaCl), MV (100 μM), cold (4 °C), heat (40 °C), ABA (100 μM), GA (100 μM), Ethephon (200 μM), JA (100 μM), SA (100 μM) or IAA (100 μM), respectively. Leaves of five or more seedlings were harvested at 0, 1, 3, 6, 12 and 24 h after initiation of the treatments, frozen immediately in liquid nitrogen, and then stored at −80 °C.

Total RNA was extracted with TRIzol reagent (Invitrogen, USA) and treated with RNase-free DNase I (Invitrogen, Gaithersburg, MD, USA). First strand cDNA was synthesized from 3 μg total RNA from each sample using M-MLV reverse transcriptase (Toyobo, Osaka, Japan) according to the supplier's protocols. Quantitative RT-PCR (qRT-PCR) was then conducted as described previously [28]. A total of 14 gene specific primer pairs for tomato trihelix genes were designed. The *β-actin* gene was employed as an internal control (Supplementary Table S1). Real-time PCR data were analyzed using the $2^{-\Delta\Delta CT}$ method.

3. Results

3.1. The trihelix gene family in tomato

36 non-redundant tomato trihelix genes were identified and manually verified their uniqueness, and were listed in Table 1. The 36 trihelix genes were subsequently renamed from *SlGT1* to *SlGT36* according to their order on the chromosomes from chromosomes 1 to 12, respectively (Table 1). All of the 36 trihelix proteins contained the myb/SANT-like domains confirmed by Pfam and InterProScan, and all the trihelixs contained a conserved tryptophan (W), a typical feature of the trihelix domain (Fig. S1). The identified nucleotide and amino acid sequences are presented in Supplementary S1.

The open reading frame (ORF) lengths of trihelix identified in this study ranged from 321 bp to 2934 bp, encoding peptides ranges from 106 to 977 amino acids (aa), with a few exceptionally smaller or longer proteins. The molecular weights of these deduced trihelix proteins ranged from 12.86 kDa to 109.94 kDa, and the isoelectric points ranged from 4.71 to 9.76 (Table 1).

The coding sequences of all the trihelix genes are disrupted by introns, and the intron numbers ranged from 1 to 16, except for six trihelix genes, which had no intron (Table 1 and Fig. S2). All of the trihelix genes were distributed on 12 chromosomes throughout the tomato genome with different densities, the number of *SlGT* genes per chromosome varied from 1 to 7. A maximum number of seven genes were present on chromosome 1 (Fig. S3 and Table 1).

3.2. Phylogenetic analysis of trihelix family in tomato, Arabidopsis and rice

To investigate the molecular evolution and phylogenetic relationships among trihelix in tomato and other species, 96 trihelix protein sequences from tomato, *Arabidopsis* and rice were aligned using ClustalX and were used to construct the phylogenetic tree in MEGA 4 program. According to the phylogenetic analysis (Fig. S4), 96 trihelix proteins could be divided into six subfamilies, designated as clade GT-1, GT-2, SH4, SIP1, GT-γ and GT-δ (bootstrap values > 50%), containing 10, 18, 18, 34, 9 and 7 members, respectively (Fig. S4). This classification was largely similar to previous phylogenetic analyses of trihelixs in *Arabidopsis* and rice [5,12], and the 36 tomato trihelix genes were distributed over all clades. Consequently, tomato trihelixs could be grouped into 6 clades, and the number of each clade was 3, 6, 8, 12, 2 and 5, respectively.

3.3. Gene duplication of SlGT genes in tomato

Gene families are generated through either tandem duplication or large-scale segmental duplication during evolution [29]. Therefore, we constructed comparative syntenic maps of tomato associated with *Arabidopsis*. 14 pairs of orthologous *GT* genes were found between *SlGTs* and *AtGTs*. And there were 8 and 15 paralogs located in genome of tomato and *Arabidopsis*, respectively (Fig. 1). Among *GT* genes in tomato, 14 were found to be segmentally duplicated, which are located on duplicated segments on chromosomes 1, 2, 3, 4, 7, 8, 9, 11 and 12 (Fig. 1).