Gene 523 (2013) 126–136

Contents lists available at SciVerse ScienceDirect

Gene

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

Superfamily of ankyrin repeat proteins in tomato

Xiaowei Yuan [a,1], Shizhong Zhang [a,b,1], Xiaohe Qing [a], Meihong Sun [a], Shiyang Liu [a], Hongyan Su [c], Huairui Shu [a,b], Xinzheng Li [a,*]

[a] State Key Laboratory of Crop Biology, Shandong Agricultural University, Tai-An, Shandong 271018, China
[b] National Research Center for Apple Engineering and Technology, Shandong Agricultural University, Tai-An, Shandong 271018, China
[c] Ludong University, Yantai, Shandong 264025, China

ARTICLE INFO

Article history:
Accepted 27 March 2013
Available online 13 April 2013

Keywords:
Genome-wide analysis
Ankyrin repeat
Protein domain
Tomato

ABSTRACT

The ankyrin repeat (ANK) protein family plays a crucial role in plant growth and development and in response to biotic and abiotic stresses. However, no detailed information concerning this family is available for tomato (*Solanum lycopersicum*) due to the limited information on whole genome sequences. In this study, we identified a total of 130 *ANK* genes in tomato genome (*SlANK*), and these genes were distributed across all 12 chromosomes at various densities. And chromosomal localizations of *SlANK* genes indicated 25 *SlANK* genes were involved in tandem duplications. Based on their domain composition, all of the SlANK proteins were grouped into 13 subgroups. A combined phylogenetic tree was constructed with the aligned SlANK protein sequences. This tree revealed that the SlANK proteins comprise five major groups. An analysis of the expression profiles of *SlANK* genes in tomato in different tissues and in response to stresses showed that the SlANK proteins play roles in plant growth, development and stress responses. To our knowledge, this is the first report of a genome-wide analysis of the tomato *ANK* gene family. This study provides valuable information regarding the classification and putative functions of *SlANK* genes in tomato.

1. Introduction

Tomato (*Solanum lycopersicum*) is one of the most economically important fruits and vegetables in the world (Hartz et al., 1997). Tomato belongs to the *Solanaceae* family, which also includes species such as potato (*Solanum*), eggplant (*Solanum*), tobacco (*Nicotiana*), pepper (*Capsicum*), mandrake (*Mandragora*) and Petunia (Knapp, 2002). Tomato research impacts agricultural and public health fields because tomato contains a large number of secondary metabolites that benefit human health and nutrition (Agarwal and Rao, 2000; Canene-Adams et al., 2005). Despite its importance, genome-wide research resources are limited for tomato. Recently, the completion of the sequencing of the tomato and the generation of a wealth of microarray data has facilitated the detailed characterisation and functional analysis of tomato genes, and these advances represent a breakthrough for basic and applied plant science (Sato et al., 2012). An extensive bioinformatic analysis of the completed tomato genome could identify numerous known or novel gene families that are associated with growth, development, gene regulation, metabolism and stress response.

The ankyrin repeat (ANK) is one of the most common protein domains and is widely distributed in organisms ranging from viruses to plants (Sedgwick and Smerdon, 1999). This domain was initially discovered in the two yeast cell-cycle regulators Swi6/Cdc10 (Breeden and Nasmyth, 1987a) and in the *Drosophila* signalling protein Notch (Breeden and Nasmyth, 1987b), and it was named after 24 copies of this sequence were discovered in the cytoskeletal protein ankyrin (Lux et al., 1990). The primary structure of ANK consists of 33 residues that are repeated in tandem to build specific secondary and tertiary structures (Mosavi et al., 2002). Only a few of the amino acids are invariant, and these residues correspond to hydrophobic positions

Abbreviations: ANK, ankyrin repeat; XB3, XA21-binding protein 3; Pkinase, protein kinase domain; IQ, calmodulin-binding motif; BTB, BTB/POZ domain; TPR, tetratricopeptide-like repeats; PPR, pentatricopeptide repeat; RF, C3HC4-type RING finger domain; CCCH, CCCH-type zinc finger domain; DHHC, DHHC-type zinc finger domain; TSC22, TSC22 domain; AAA, ATPase family; RCC1, regulator of chromosome condensation repeat; Pkinase_Tyr, Tyrosine kinase; BAR, Bin-Amphiphysin-Rvs domain; PH, pleckstrin homology domain; Chromo, chromatin organisation modifier; HA2, helicase-associated domain; Helicase, RNA helicase domain; ArfGap, putative GTPase activating protein for Arf; Ion_trans, transmembrane ion channel family; cNMP_binding, cyclic nucleotide-binding domain; SBP, squamosa-promoter binding protein; GPCR_chapero_1, GPCR_chapero_1 domain; Motile_Sperm, major sperm protein domain; AMP-binding, AMP-binding domain; Herpes_DNAp, Herpes_DNAp domain; Methyltransf, Methyltransferase domain; UBN2_3, gag-polypeptide of LTR copia-type domain; UCH37_bd, ubiquitin C-terminal hydrolase 37 receptor binding site; BLAST, the Basic Local Alignment Search Tool; SMART, a Simple Modular Architecture Research Tool; GSDS, Gene Structure Display Sever; HMM, Hidden Markov Model; NJ, the neighbor-joining; MEGA, Molecular Evolutionary Genetics Analysis.

* Corresponding author at: College of Life Sciences, Shandong Agricultural University, Tai-An, Shandong 271018, China. Tel.: +86 538 8246205.
E-mail address: lxz@sdau.edu.cn (X. Li).

[1] These authors contributed equally to this work.

0378-1119/$ – see front matter. Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.03.122

that are required to maintain the secondary structure (Bork, 1993; Mosavi et al., 2002; Rohde and Bork, 1993).

ANK, which is known to mediate protein–protein interactions (Michaely and Bennett, 1992), has been identified in numerous proteins with diverse functions (Bork, 1993; Mosavi et al., 2004). In animals and yeast, some ANK proteins play important roles in cell-cycle control, transcriptional regulation, cytoskeleton integrity and signal transduction (Sedgwick and Smerdon, 1999). In plants, ANK proteins have been demonstrated to be involved in a number of important physiological processes (Becerra et al., 2004; Huang et al., 2009). AKR, which was the first reported ANK protein in *Arabidopsis*, is regulated by light and plays a regulatory role in cell differentiation and development (Zhang et al., 1992). *BOP1*, an *Arabidopsis ANK* gene, is required for leaf morphogenesis (Ha et al., 2004). EMB506, which contains five ANK repeats organised in tandem within the C-terminal moiety, is essential for *Arabidopsis* embryogenesis (Albert et al., 1999). Further study revealed that EMB506 interacts with AKRP (encoded by *AKR*) through their ANK domains (Garcion et al., 2006). *TIP1*, an *Arabidopsis* ANK gene that encodes an *S*-acyl transferase, affected cell growth throughout the plant life cycle (Hemsley et al., 2005). XBAT32, which contains five ANK repeats, positively regulates lateral root development via the degradation of the ethylene biosynthetic enzyme 1-aminocyclopropane-1-carboxylate synthase 7 (Lyzenga et al., 2012); this process therefore down-regulates ethylene biosynthesis (Nodzon et al., 2004; Prasad et al., 2010). In addition, XBAT35 is a novel player in ethylene signalling that is involved in negatively regulating apical hook curvature (Carvalho et al., 2012). *LIANK*, a *Lily ANK* gene, is essential for pollen germination and pollen tube growth (Huang et al., 2006). LKT1, a K^+ uptake channel in tomato root hairs, may serve as a low-affinity influx pathway for K^+ into root hair cells (Hartje et al., 2000).

ANK proteins have also been found to play important roles in responses to biotic and abiotic stresses in plants. The *Arabidopsis* ANK protein AKR2 may be involved in regulating antioxidant metabolism during disease resistance and stress responses (Yan et al., 2002). In pepper, CaKR1 was found to play roles in both biotic and abiotic stress responses (Seong et al., 2007a, 2007b). ACD6, acting as a plasma membrane-localised positive regulator of salicylic acid signalling, controls defence responses against virulent bacteria (Lu et al., 2003; Lu et al., 2005). *ZFAR1* encodes a putative zinc-finger protein containing ANK domains, and the mutant confers increased local susceptibility to *Botrytis* and sensitivity in the presence of abscisic acid in germination (AbuQamar et al., 2006). *ITN1*, which encodes an ANK-transmembrane protein, has been implicated in diverse cellular processes (Sakamoto et al., 2008). NPR1, a positive regulator of acquired resistance responses, is a central activator of SA-regulated gene expression (Cao et al., 1997). XB3 (XA21-binding protein 3) is a substrate for the XA21 serine and threonine kinase and is required for the full accumulation of the XA21 protein and for Xa21-mediated immunity in rice (Wang et al., 2006).

In *Arabidopsis*, 105 ANK proteins have been identified and classified into 12 subgroups (Becerra et al., 2004), and in rice, 175 *ANK* genes have identified and classified into 10 subfamilies (Huang et al., 2009). However, no detailed information concerning this family is available for tomato (*S. lycopersicum*). The draft tomato genome sequences, which were reported earlier this year (Sato et al., 2012), offer the opportunity to investigate the ANK-containing protein family in this species. This study aimed to identify the complete set of *ANK* genes in the sequenced tomato genome using a bioinformatics approach. We first identified the putative *ANK* genes in tomato. These genes were classified according to their domain compositions, and the chromosomal locations of these *ANK* genes were studied. Finally, we examined the expression profiles of these genes under normal and stress conditions. As a first step towards genome-wide analyses of the tomato *ANK* genes, our results constitute a foundation to understand the classification and further functionally analyse each ANK protein family member in tomato.

2. Materials and methods

2.1. The identification of the ANK genes in tomato

Two different approaches were utilised to identify the members of the *ANK* gene family in tomato. First, all of the known *Arabidopsis* and rice *ANK* gene sequences were used as query sequences to perform multiple database searches against the proteome and genome files downloaded from the plantGDB database (plant genome database and analysis tools: http://www.plantgdb.org/) (Duvick et al., 2008). Stand-alone versions of BLAST (Basic Local Alignment Search Tool: http://blast.ncbi.nlm.nih.gov) (Mount, 2007), which are available from NCBI, were used with an e-value cut-off of 1e-3. All of the protein sequences that were derived from the collected candidate *SlANK* genes were examined using the domain analysis programs PFAM (Protein family: http://pfam.sanger.ac.uk/) and SMART (Simple Modular Architecture Research Toole: http://smart.embl-heidelberg.de/) with the default cut-off parameters (Schultz et al., 1998; Xu and Dunbrack, 2012). We next analysed the domains of all of the tomato peptide sequences using a hidden Markov model (HMM) analysis with Pfam searching and SMART tools (Johnson et al., 2010). We then obtained the sequences with the PF00023 Pfam number and SMART accession number SM00248, which contained typical ANK domains, from the tomato genome sequences using a Perl-based script. Finally, all of the protein sequences were compared with known ANK sequences using the ClustalX software program (http://www.clustal.org/) to confirm that the sequences were candidate *ANK* genes (Thompson et al., 2002).

The proteins' isoelectric points and molecular weights were predicted using the proteomics and sequence analysis tools on the ExPASy Proteomics Server (http://expasy.org/) (Gasteiger et al., 2003). The chromosomal locations were predicted using the plantGDB database with a Perl-based program.

2.2. The chromosomal locations and gene structures of the ANK genes

The chromosomal locations and gene structures were retrieved from the tomato genome data that were downloaded from the plantGDB database (http://www.plantgdb.org/). The remaining genes were mapped onto the chromosomes using the MapDraw software program (Liu and Meng, 2003), and the gene structures of the *SlANK* genes were generated using the GSDS (Gene Structure Display Sever: http://gsds.cbi.pku.edu.cn/) (Guo et al., 2007).

2.3. Sequence alignment and phylogenetic analysis of SlANK

The complete SlANK protein sequences were aligned using the ClustalX program, and BLOSUM30 was used as the protein-weight matrix. The MUSCLE software program (version 3.52) was also used to perform multiple sequence alignments to confirm the ClustalX results (http://www.clustal.org/). Phylogenetic trees of the SlANK protein sequences were constructed using the neighbour-joining (NJ) method of the MEGA5 software program (molecular evolutionary genetics analysis: http://www.megasoftware.net/) using the p-distance and complete deletion option parameters (Tamura et al., 2011). The reliability of the obtained trees was tested using a bootstrapping method with 1000 replicates. The images of the phylogenetic trees were drawn using MEGA5.

2.4. Expression analyses of the SlANK genes

Microarray expression data from various datasets were obtained using Genevestigator (https://www.genevestigator.com/gv/) with the tomato Gene Chip platform. The tomato whole-genome sequences were used as query sequences to perform a BLAST against all of the gene probe sequences from the Affymetrix Gene Chip (http://www.