Biochemical Systematics and Ecology 71 (2017) 212–218

Contents lists available at ScienceDirect

Biochemical Systematics and Ecology

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

Genome-wide identification and characterization of the WRKY gene family in potato (*Solanum tuberosum*)

Qiu-Ning Liu, Yu Liu, Zhao-Zhe Xin, Dai-Zhen Zhang, Bao-Ming Ge, Rui-Ping Yang, Zheng-Fei Wang, Li Yang, Bo-Ping Tang*, Chun-Lin Zhou**

Jiangsu Key Laboratory for Bioresources of Saline Soils, Jiangsu Synthetic Innovation Center for Coastal Bio-agriculture, Jiangsu Provincial Key Laboratory of Coastal Wetland Bioresources and Environmental Protection, School of Ocean and Biological Engineering, Yancheng Teachers University, Yancheng 224051, PR China

ARTICLE INFO

Article history:
Received 8 December 2016
Received in revised form 20 February 2017
Accepted 25 February 2017

Keywords:
Solanum tuberosum
WRKY
Transcription factor
Genome-wide analysis
Phylogenetic analysis

ABSTRACT

WRKY proteins are a large family of plant transcription factors, which play an important role in many biological processes such as plant development, metabolism, and responses to biotic and abiotic stresses. This study aimed to develop a complete overview of the WRKY family by using the potato genome. Based on the genome-wide analysis, a total of 82 StWRKY proteins were characterized and classified into three groups (I, II, and III) and five subgroups (IIa, IIb, IIc, IId, and IIe) according to the sequence similarity, motif varieties, and their phylogenetic relationships. By using multiple sequence alignment, a novel motif, WRKYGQR, was identified from StWRKY076 along with three variations of the WRKYGQK heptapeptide (WRKYGMK, WHKCGQK, and WRKYGKK) that might execute new biological functions. Chromosomal location analysis showed that the 82 StWRKY proteins were located in 11 of the 12 chromosomes, excluding chromosome 11. Phylogenetic analyses also showed that the 82 StWRKY proteins could be clustered into three large groups and five subgroups which shared typical characters of WRKY. The results provide fundamental information to inform further analysis and understanding of WRKY gene functions in potato.

1. Introduction

Throughout their long evolutionary development, plants have developed a series of adjustment mechanisms to avoid or cope with different types of stresses such as pests, fungal, bacterial, viral challenges, drought, cold, and salinity stress (Zhang et al., 1999). Transcription factors are a class of proteins that regulate gene expression to control plant physiological and biochemical processes. They bind to specific regions, known as *cis*-elements, in the promoters of genes and then activate or repress the expression of regulated genes in collaboration with other regulatory factors (Riechmann and Ratcliffe, 2000). Numerous transcription factor families are involved in these adjustment mechanisms by modulating gene expression patterns (Meshi and Iwabuchi, 1995; Zhang et al., 2015). The WRKY family of transcription factors is one of the largest and is widely distributed in plants (Rushton et al., 2012).

* Corresponding author.
** Corresponding author.
E-mail addresses: boptang@163.com (B.-P. Tang), zcl8199@163.com (C.-L. Zhou).

http://dx.doi.org/10.1016/j.bse.2017.02.010
0305-1978/

WRKY transcription factors contain one or two conserved DNA binding regions called the WRKY domain (Tripathi et al., 2014; Jiang et al., 2016). The WRKY domain, approximately 60 amino acid residues in size, contains a highly conserved 'WRKYGQK hexapeptide sequence' amino acid motif at its *N*-terminus and an atypical zinc-finger structure ($C-X_{4-6}-C-X_{22-23}-H-X_1-H$ or $C-X_7-C-X_{23}-H-X_1-C$) at its *C*-terminus (Rushton et al., 2010). Based on the number of WRKY domains present and the type of zinc-finger structure, WRKY proteins can be classified into three distinct groups (I, II, and III). Group I proteins include a C2H2 motif and two WRKY domains. Group II proteins typically contain a single WRKY domain and a C2H2 zinc-finger motif ($C-X_{4-6}-C-X_{22-23}-H-X_1-H$) and can be further divided into five subgroups (IIa-IIe). Group III proteins have only a single WRKY domain and a zinc-finger-like motif ($C-X_7-C-X_{23}-H-X_1-C$) (Zhang and Wang, 2005).

The WRKY gene was originally cloned and identified in sweet potato (*Ipomoea batatas*) (Ishiguro and Nakamura, 1994). To date, genome-wide WRKY analysis has been performed in many plant species including *Arabidopsis thaliana* (Eulgem et al., 2000), *Oryza sativa* (Jang et al., 2010), *Zea mays* (Wei et al., 2012), and *Solanum lycopersicum* (Huang et al., 2012). In plants, WRKY transcription factors are involved in various physiological processes, such as plant growth and development (Ay et al., 2009), signal transductions (Antoni et al., 2011), trichome formation (Johnson et al., 2002), seed development (Jiang and Yu, 2009), plant hormone signaling (Antoni et al., 2011), and leaf senescence (Miao and Zentgraf, 2010). Furthermore, WRKY proteins play a vital role in plant defenses against various biotic stresses and in responses to various abiotic stresses, including bacterial, fungal, and viral pathogens (Chen et al., 2013). However, little information is available regarding the regulation and genomic structure of WRKY transcription factors in potato.

Potato, *Solanum tuberosum*, the most important non-grain food crop and the fourth most important crop globally after wheat, rice, and maize, has been widely cultivated around the world for centuries (PGSC, 2011). Potato yield is frequently affected by different types of stress, leading to a decline in potato output. Therefore, the study of stress-related genes is becoming more important and urgent for potato breeding and production. The Potato Genome Sequencing Consortium completed the genome sequencing of doubled monoploid *S. tuberosum*, providing the possibility of identifying potato WRKY genes at the genome-wide level (PGSC, 2011). In this study, genome-wide identification of WRKY proteins was conducted by using the potato genome and bioinformatic analyses of the gene structures, chromosomal locations, conserved protein domains, and phylogenetic inferences. This data will provide a novel insight into future work on the function of WRKY proteins in potato.

2. Materials and methods

2.1. Identification of the WRKY proteins in potato

The whole-genome protein sequences of potato were downloaded from the Potato Genome Sequencing Consortium website (http://solanaceae.plantbiology.msu.edu/) and used to screen for potato WRKY proteins. To identify a complete list of potato WRKY genes, the Hidden Markov Model profile and consensus pattern of the WRKY domain (PF03106) were downloaded from the Pfam database (http://pfam.xfam.org/search) to obtain the conserved domain. We employed the WRKY domain as a query to identify all possible WRKY gene sequences in the potato genome database using the BLASTp program research in the NCBI protein database (http://blast.ncbi.nlm.nih.gov/blast.cgi) and Spud DB Potato Genomics Resources (http://solanaceae.plantbiology.msu.edu/) with an E-value of 0.001.

2.2. Sequence analysis of WRKY proteins in potato

The candidate genes were analyzed using the domain identification function of the Pfam database to remove the WRKYs without the conserved domain sequences. We located overlapping genes by aligning all of the candidate WRKY gene sequences using Clustal W. Only the nonoverlapping WRKY sequences were used for further analysis. The online ExPASy proteomics server (http://expasy.org/) was used to investigate the isoelectric point (pI) and molecular weight (MW) of the deduced StWRKY amino acid sequences. The presence of the WRKY domain was confirmed using the simple modular architecture research tool (SMART) program (http://smart.embl-heidelberg.de/).

2.3. Multiple sequence alignment and phylogenetic analysis

Based on the genome-wide analysis, a total of 82 StWRKY proteins were characterized and classified into three groups (I, II, and III) and five subgroups (IIa, IIb, IIc, IId, and IIe) according to sequence similarity and motif varieties. Multiple sequence alignments of potato WRKY proteins and domains were separately carried out using Clustal X software (Thompson et al., 1997). A phylogenetic tree was constructed using Molecular Evolutionary Genetics Analysis (MEGA) version 6.0 (Tamura et al., 2013). The data were analyzed using Poisson correction, and gaps were removed by complete deletion. The topological stability of the neighbor-joining trees was evaluated with 1000 bootstrapping replications.