Available online at www.sciencedirect.com

ScienceDirect

Genome-wide analysis of the *WRKY* transcription factor gene family in *Gossypium raimondii* and the expression of orthologs in cultivated tetraploid cotton

*Caiping Cai, Erli Niu, Hao Du, Liang Zhao, Yue Feng, Wangzhen Guo**

State Key Laboratory of Crop Genetics & Germplasm Enhancement, Hybrid Cotton R&D Engineering Research Center, MOE, Nanjing Agricultural University, Nanjing 210095, China

ARTICLE INFO

Article history:
Received 16 January 2014
Received in revised form
21 February 2014
Accepted 19 March 2014
Available online 27 March 2014

Keywords:
Expression pattern
Gossypium
Phylogenetic relationship
Stress
WRKY transcription factors

ABSTRACT

WRKY proteins are members of a family of transcription factors in higher plants that function in plant responses to various physiological processes. We identified 120 candidate *WRKY* genes from *Gossypium raimondii* with corresponding expressed sequence tags in at least one of four cotton species, *Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium arboreum*, and *G. raimondii*. These WRKY members were anchored on 13 chromosomes in *G. raimondii* with uneven distribution. Phylogenetic analysis showed that WRKY candidate genes can be classified into three groups, with 20 members in group I, 88 in group II, and 12 in group III. The 88 genes in group II were further classified into five subgroups, groups IIa–e, containing 7, 16, 37, 15, and 13 members, respectively. We characterized diversity in amino acid residues in the WRKY domain and/or other zinc finger motif regions in the WRKY proteins. The expression patterns of *WRKY* genes revealed their important roles in diverse functions in cotton developmental stages of vegetative and reproductive growth and stress response. Structural and expression analyses show that WRKY proteins are a class of important regulators of growth and development and play key roles in response to stresses in cotton.

© 2014 Crop Science Society of China and Institute of Crop Science, CAAS. Production and hosting by Elsevier B.V. All rights reserved.

1. Introduction

Transcription factors, which exist in all living organisms, are essential for the regulation of gene expression. WRKY transcription factors, a family of regulatory genes, were first identified in plants [1–3]. In WRKY family proteins, a 60 amino acid region is highly conserved among family members. It includes the conserved WRKYGQK sequence followed by one

Abbreviations: DPA, days post anthesis; EST, expressed sequence tag; ETI, effector-triggered immunity; MTI, MAMP-triggered immunity; ORF, open reading frame; PCR, polymerase chain reaction; PTI, PAMP-triggered immunity.

* *Corresponding author*. Tel.: +86 25 84396523.
E-mail address: moelab@njau.edu.cn (W. Guo)

Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.cj.2014.03.001

2214-5141/© 2014 Crop Science Society of China and Institute of Crop Science, CAAS. Production and hosting by Elsevier B.V. All rights reserved.

of the two types of zinc finger motifs, the C_2H_2 and C_2–HC types [4]. All known WRKY proteins can be divided into three groups (group I, II, and III) based on the number of WRKY domains and the types of zinc finger motif. Two WRKY domains can be found in group I proteins, whereas a single domain is present in group II and group III proteins. Generally, group I and group II proteins share the same C_2H_2-type zinc finger motif (C–X_{4-5}–C–X_{22-23}–H–X_1–H). In group III, WRKY domains contain a C_2–HC-type motif (C–X_7–C–X_{23}–H–X_1–C) [4]. Group II is further classified into several subgroups based on their phylogenetic clades [4–6].

In plants, WRKY proteins form a large family of transcription factors and are known to function in response to various physiological processes. WRKY transcription factors are important components of many aspects in the plant defense system, including MAMP-triggered immunity (MTI) or PAMP-triggered immunity (PTI), effector-triggered immunity (ETI), and systemic acquired resistance [7–14]. These transcription factors also play important regulatory roles in plant abiotic stress. For example, *Arabidopsis* plants that overexpress *GmWRKY21* are more cold-stress tolerant than wild-type plants, and plants overexpressing *GmWRKY54* exhibit increased salt and drought tolerance, whereas plants overexpressing *GmWRKY13* exhibit increased sensitivity to salt and mannitol stress [15]. In barley (*Hordeum vulgare*), *HvWRKY38* is involved in cold and drought responses [16]. The expression of *AtWRKY25* and *AtWRKY26* is induced upon treatment with high temperatures, whereas *AtWRKY33* expression is repressed in response to the same treatment [17]. In addition to functioning in biotic and abiotic stresses, WRKY transcription factors regulate developmental processes, such as trichome and seed coat development in *Arabidopsis* [18], sesquiterpene biosynthesis in cotton (*Gossypium hirsutum*) [19], seed development in barley, *Solanum chacoense*, and *Arabidopsis* [20–22], and senescence in *Arabidopsis* [23–25].

Since the release of a large number of publicly available sequences and even complete whole-genome sequences in some plants, genome-wide analyses of the *WRKY* gene family have been performed. There are at least 72 WRKY family members in *Arabidopsis* [4], more than 100 in rice (*Oryza sativa*) [5], 57 in *Cucumis sativus* [26], 104 in *Populus trichocarpa* [27], and 81 in *Solanum lycopersicum* [28]. Genome duplication events have been detected in this family [27], and the divergence of the monocots and dicots was verified based on the analysis of WRKY transcription factors [5,6].

The genus *Gossypium* has great economic and scientific importance. Cotton produces the most important natural textile fiber in the world and is also an important oilseed crop. Cotton fiber is an outstanding model for studying plant cell elongation and cell wall biosynthesis [29]. Tetraploid cotton is also an excellent model system for studying polyploidization and genome duplication. Despite the importance of *WRKY* genes in plant growth and developmental processes, to our knowledge only eight *WRKY* genes have previously been reported from different cotton species [13,19,30,31]. Genome-wide analysis of the WRKY transcription factor family in *Gossypium* will lay the foundation for elucidating their structure, evolution, and functional roles.

Currently 435,344 cotton EST sequences are available in the GenBank EST database (http://www.ncbi.nlm.nih.gov/dbEST/). Among them, 297,214 ESTs were identified in *G. hirsutum*, 63,577 in *Gossypium raimondii*, 41,781 in *Gossypium arboreum*, 32,525 in *Gossypium barbadense*, and 247 in *Gossypium herbaceum*. A pilot study for the whole-genome scaffold sequence of the diploid cotton *G. raimondii*, which is the putative contributor of D-subgenome to fiber-producing cotton species including *G. hirsutum* and *G. barbadense*, has been released by two research groups [32,33]. As an application, *G. raimondii* genome sequences have been of great advantage for assembling the tetraploid transcriptome and mining candidate genes of interest [34]. Information from the publicly available *Gossypium* database will serve as a foundation for identifying gene families such as *WRKY* genes.

The objective of the current study was to survey the *WRKY* genes and their phylogenetic relationship in *Gossypium* with a bioinformatic approach using information derived from the publicly available database from the two drafts of the D_5 genome (*G. raimondii*) and ESTs from NCBI (http://www.ncbi.nlm.nih.gov/dbEST/), combined with sequence data confirmation via cloning of cDNAs containing complete open reading frames (ORFs) from upland cotton. We further evaluated the expression patterns of *WRKY* genes in various developmental stages and under various stress conditions in tetraploid cultivated cotton species.

2. Materials and methods

2.1. Prediction of WRKY gene family

Genes and proteins annotated in *G. raimondii* were downloaded from http://www.phytozome.net/ and http://cgp.genomics.org.cn/. WRKY transcription factors were identified using HMMER software version 3.0 [35] and the PFAM protein family database using the WRKY domain (PF03106) as a query [36]. Expressed sequence tag (EST) sequences for four cotton species, *G. hirsutum* (Gh), *G. barbadense* (Gb), *G. arboreum* (Ga), and *G. raimondii* (Gr), were downloaded from the GenBank EST database (http://www.ncbi.nlm.nih.gov/dbEST/). WRKY protein sequences in *Arabidopsis* were obtained from The *Arabidopsis* Information Resource (TAIR: http://www.arabidopsis.org/).

2.2. Mapping and analysis of WRKY genes

Mapping of *WRKY* genes was performed using MapInspect (http://www.plantbreeding.wur.nl/UK/ software_mapinspect.html). Exons and introns were predicted by comparing the coding sequences with their genomic sequences using the online GSDS program [37]. Conserved motif prediction was performed using the MEME program [38]. The following parameters were used for analysis: maximum number of motifs, 10; minimum motif width, six; and maximum motif width, 70.

2.3. Sequence alignment and phylogenetic construction

Alignment of the amino acid sequences of the WRKY domain with approximately 60 amino acids was performed with ClustalX 1.83 [39]. The parameters used in the alignment were as follows: for pairwise parameters, gap opening: 10.00,

Download English Version:

https://daneshyari.com/en/article/2079520

Download Persian Version:

https://daneshyari.com/article/2079520

Daneshyari.com