Plant Gene xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Plant Gene

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

Genome-wide identification and characterization of the homeodomain-leucine zipper I family of genes in cotton (*Gossypium* spp.)

Jian Zhang [a,b], Qian-Hao Zhu [a,*], Philippe Moncuquet [a], Danny Llewellyn [a], Iain Wilson [a,*]

[a] CSIRO Agriculture, Black Mountain Laboratories, GPO Box 1600, Canberra, ACT 2601, Australia
[b] College of Agronomy and Biotechnology, Southwest University, Chongqing 400716, China

ARTICLE INFO

Article history:
Received 13 April 2016
Received in revised form 31 May 2016
Accepted 31 May 2016
Available online xxxx

Keywords:
HD-Zip I transcription factor
Phylogenetic analysis
Gene conversion
Homoeologue
Verticillium dahliae
Cotton

ABSTRACT

Homeodomain-leucine zipper (HD-Zip) transcription factors are unique to the plant kingdom and are classified into four subfamilies, HD-Zip I to IV. This gene family has been extensively investigated in several plant species and many members have been shown to play important roles in plant development and in response to abiotic/biotic stresses. In cotton, several HD-Zip IV genes have been identified and their function investigated, but little is known about the HD-Zip I genes. Here, we performed a genome-wide survey and identified 72, 30 and 34 HD-Zip I genes in *Gossypium hirsutum*, *Gossypium arboreum* and *Gossypium raimondii*, respectively. Almost all *G. arboreum* and *G. raimondii* HD-Zip I genes were retained in allotetraploid *G. hirsutum*, and new HD-Zip I genes were evolved in *G. hirsutum* after polyploidization, probably through tandem and/or segmental duplication. Most HD-Zip I genes were under purifying selection although some could have undergone positive selection. Small indels and nonreciprocal homoeologous recombination (NRHR) events also played a role in shaping the HD-Zip I genes in *G. hirsutum*. Most HD-Zip I genes were preferentially expressed in certain tissues. Differential expression of homoeologues was observed but the differences were generally less than that between different genes. Three HD-Zip I genes were found to have a consistent response in *G. hirsutum* and *G. barbadense* cultivars resistant to *Verticillium dahliae* (*Vd*) following *Vd*-infection. Our results provided a comprehensive view of the cotton HD-Zip I genes and fundamental information for further research towards understanding the role of HD-Zip I genes in cotton.

1. Introduction

The homeodomain (HD) is a conserved 60-amino acid motif, which has a characteristic three-helix structure that is able to bind to specific DNA sequences (Ariel et al., 2007). HD-containing proteins play fundamental roles in a diverse range of developmental processes, from pattern formation to cell type specification, in all eukaryotic organisms (Gehring et al., 1994). Based on the distinguishing features of the HD-encoding sequence, its size and location, as well as other associated domains and gene structures, the HD-containing proteins are classified into six families, including homeodomain-leucine zipper (HD-Zip), Wuschel-related homeobox (WOX), plant homeodomain associated to a finger domain (PHD-finger) or a Bell domain (BELL), Knotted related homeobox (KNOX) and zinc finger-homeodomain (ZF-HD) (Ariel et al., 2007). The HD-Zip transcription factors are unique to the plant kingdom. They contain not only a highly conserved HD, but also a leucine-zipper (LZ) motif immediately downstream of the HD. The HD binds to specific DNA in its target genes while the leucine-zipper motif mediates dimerization (Schena and Davis, 1992). According to the conserved HD-Zip domain, additional conserved motifs, gene structural features and biological functions, the HD-Zip family are further divided into four subfamilies, *i.e.* HD-Zip I to IV (Ariel et al., 2007).

The HD-Zip I proteins contain only the HD domain and an adjacent LZ motif, whereas the HD-Zip II proteins contain five additional conserved amino acids at the C-terminus, known as "CPSCE", and sometimes an additional N-terminal motif (Tron et al., 2002). Members of the HD-Zip I and II subfamilies bind similar pseudo-palindromic *cis* elements, *i.e.* CAAT-NATTG (Sessa et al., 1993; Frank et al., 1998; Johannesson et al., 2001). Both the HD-Zip III and IV subfamily proteins can be distinguished from the HD-Zip I and II subfamilies by the presence of a StAR (steroidogenic acute regulatory protein)-related lipid-transfer (START) domain followed by a START-adjacent domain (Ponting and Aravind, 1999; Schrick et al., 2004); however, the HD-Zip III proteins contain an additional C-terminal MEKHLA domain, while the HD-Zip IV proteins lack this motif (Mukherjee and Bürglin,

Abbreviations: ABA, abscisic acid; dpa, days post anthesis; dpi, days post infection; FPKM, fragments per kilobase of exon per million fragments mapped; HD-Zip, homeodomain-leucine zipper; hpi, hours post infection; *LMI1*, *LATE MERISTEM IDENTITY 1*; NRHR, nonreciprocal homoeologous recombination; *RCO*, *REDUCED COMPLEXITY*; qRT-PCR, quantitative real-time polymerase chain reaction; *Vd*, *Verticillium dahliae*.

* Corresponding authors.
E-mail addresses: zhangjianswau@126.com (J. Zhang), Qianhao.Zhu@csiro.au (Q.-H. Zhu), Philippe.Moncuquet@csiro.au (P. Moncuquet), Danny.Llewellyn@csiro.au (D. Llewellyn), Iain.Wilson@csiro.au (I. Wilson).

http://dx.doi.org/10.1016/j.plgene.2016.05.002
2352-4073/

2006). The *Arabidopsis thaliana* genome has 17, 9, 5 and 16 genes encoding HD-Zip I, II, III and IV proteins, respectively (Ariel et al., 2007). Similarly, multiple members are present in each HD-Zip subfamily in all plant species investigated, such as maize (Zhao et al., 2011) and soybean (Chen et al., 2014).

The functions of some HD-Zip genes are known. HD-Zip II proteins, for example, have been shown to be involved in responses to illumination conditions, shade avoidance and auxin signaling (Morelli and Ruberti, 2000; Sawa et al., 2002; Rueda et al., 2005; Sessa et al., 2005). HD-Zip III proteins are required for apical meristem and vascular bundle development, embryogenesis, leaf polarity formation and lateral organ initiation (Mattsson et al., 2003; Prigge et al., 2005). HD-Zip IV proteins play crucial roles in anthocyanin accumulation, epidermal cell differentiation, trichome formation, and root and cuticle development (Nakamura et al., 2006; Ariel et al., 2007).

HD-Zip I proteins, on the other hand, have been shown to be important regulators in responses to abiotic/biotic stresses and in the regulation of plant organ growth and development (Ariel et al., 2007). Expression of the HD-Zip I genes has been found to be regulated by a wide range of abiotic stresses, such as drought, extreme temperatures, light conditions, osmosis and phytohormones (Himmelbach et al., 2002; Wang et al., 2003; Olsson et al., 2004; Dezar et al., 2005; Manavella et al., 2006, 2008; Ariel et al., 2010; Zhao et al., 2011; Cabello et al., 2012). For example, *ATHB6*, *ATHB7* and *ATHB12* are either up- or down-regulated by water-deficit conditions and/or externally applied abscisic acid (ABA) in *Arabidopsis*, implying roles in regulating plant responses to dehydration (Lee and Chun, 1998; Söderman et al., 1996, 1999). In *Craterostigma plantagineum*, two HD-Zip I members, *CpHB6* and *CpHB7*, were induced by both drought and ABA, whereas another two members, *CpHB4* and *CpHB5*, were down-regulated by dehydration and were not responsive to ABA treatment (Frank et al., 1998; Deng et al., 2002). *Oshox22*, a HD-Zip I gene in rice, has been shown to affect ABA biosynthesis and regulates drought and salt responses through ABA-mediated signal transduction pathways (Zhang et al., 2012). Similarly, *Zmhdz10*, a HD-Zip I gene in *Zea mays*, can positively regulate drought and salt tolerance through an ABA-dependent signaling pathway (Zhao et al., 2014). *Arabidopsis* plants overexpressing *AtHB13* were resistant to infections with downy mildew (*Hyaloperonospora arabidopsidis*) and green peach aphids (Gao et al., 2014). Regarding the role of HD-Zip I genes in plant development and morphogenesis, the six-rowed spike phenotype in barley (*Hordeum vulgare*) was found to be caused by loss-of-function mutations in *VRS1* (*Six-rowed spike 1*), which suppresses development of the rudimentary lateral spikelets as observed in the two-rowed wild-type barely (Komatsuda et al., 2007). In addition, *Arabidopsis LMI1* (*LATE MERISTEM IDENTITY 1*) and its *Cardamine hirsute* and *Cardamine grandiflora* homologue *RCO* (*REDUCED COMPLEXITY*) have been demonstrated to play a role in leaf serration and leaflet formation (Saddic et al., 2006; Sicard et al., 2014; Vlad et al., 2014). Overexpressing *C. hirsute RCO* in *Arabidopsis* converted simple leaves into complex leaves (Vlad et al., 2014). Our recent work showed that *GhOKRA* (*Gh_D01G2042*), a homologue of *LMI1* and *RCO*, is the gene underlying the okra leaf shape in *Gossypium hirsutum* (Zhu et al., 2016).

Cotton (*Gossypium* spp.) is an important economic crop and a model plant for the study of polyploidy, cell elongation and cell wall synthesis (Paterson et al., 2012). There are at least 50 species in the *Gossypium* genus (Fryxell, 1992). Four species, including *G. hirsutum* (AD_1), *Gossypium barbadense* (AD_2), *Gossypium arboreum* (A_2) and *Gossypium herbaceum* (A_1), are cultivated for their fibers. *G. hirsutum*, also known as Upland cotton, is widely planted in more than 70 countries with China, the United States of America, India and Pakistan being the leading producers, due to its wide adaptability and high production, and yields over 95% of the worldwide cotton fiber. *G. barbadense*, famous for its high fiber quality, produces about 2% of the world's cotton fiber, *G. arboreum* and *G. herbaceum* together contribute about 2% of the world's cotton production (Chen et al., 2007). The allotetraploid cottons (*G. hirsutum* and *G. barbadense*) were derived from a cross between a D-genome species similar to *G. raimondii* (D_5), as a pollen-providing parent, and an A-genome species similar to *G. arboreum* or *G. herbaceum*, as the maternal parent (Wendel and Cronn, 2003).

In addition to *GhOKRA* (Zhu et al., 2016), only a few other HD-Zip genes have been reported and their functions investigated in cotton. *GhHB1*, a HD-Zip I gene, may be involved in response to salt stress and ABA treatment (Ni et al., 2008). *GhHB2/3/4*, encoding HD-Zip II proteins, have been found to be preferentially expressed during the early developmental stages of cotton seedlings and may be involved in phytohormone signaling (Qin et al., 2010). Of the two HD-Zip IV genes identified in *G. arboreum*, *GaHOX1* is predominately expressed in the early fiber developmental stages, while *GaHOX2* is expressed in both fiber and other ovular tissues, including outer and inner integuments (Guan et al., 2008). Of the two HD-Zip IV genes characterized in *G. hirsutum*, *GhHD1* was reported to mainly be involved in trichome development and had only a mild effect on fiber cell development (Walford et al., 2012), whereas *GhHOX3* was reported to have a role in controlling cotton fiber elongation (Shan et al., 2014).

Genome-wide identification of the HD-Zip family genes have been performed in several plant species, including *A. thaliana* (Henriksson et al., 2005), *Oryza sativa* (Agalou et al., 2008), *Zea mays* (Zhao et al., 2011), *Populus trichocarpa* (Hu et al., 2012), *Cucumis sativus* (Liu et al., 2013), *Glycine max* (Chen et al., 2014), *Prunus persica* (Zhang et al., 2014) and *Pyrus betulifolia* (Wang et al., 2015), but no systematic analysis of the HD-Zip genes has been reported in cotton. In this study, we identified HD-Zip I genes in three cotton species (*G. hirsutum*, *G. arboreum* and *G. raimondii*), for which a genome sequence is currently available, with the aims i) to have a genome-wide overview of the HD-Zip I genes in these three cotton species; ii) to understand the evolutionary relationship of the HD-Zip I genes in cotton; and iii) to define the expression profiles of the HD-Zip I genes in different tissues and in response to *Verticillium dahliae* infection in *G. hirsutum*.

2. Materials and methods

2.1. Plant materials and Verticillium dahliae infection

Two *G. hirsutum* varieties (MCU-5 and Siokra 1–4) were used in this study. MCU-5 was used in expression analysis of the HD-Zip I genes in various tissues. MCU-5 (*Verticillium dahliae* resistant) and Siokra 1–4 (*V. dahliae* susceptible) were used in *V. dahliae* infection experiments. Cotton plants were grown in a glasshouse (Canberra, Australia) at 28 ± 2 °C or 22 ± 2 °C (for the *Vd*-infection experiment) with approximately 16 h day and 8 h night regime. Tissues used in gene expression analysis were roots and cotyledons (collected at the cotyledon stage), whole ovules collected at −1, 0, 1, 3 and 5 DPA (days post anthesis) and 15 DPA fiber. To investigate responses of HD-Zip I genes to *V. dahliae* infection, one-true-leaf stage of MCU-5 and Siokra 1–4 seedlings were inoculated with *V. dahliae* as previously described (Zhu et al., 2013). Seedlings treated with water (mock treatment) were used as controls. Leaf samples were collected at 1, 3 and 7 dpi (days post infection) from both *V. dahliae* and mock treated plants. Samples were immediately frozen in liquid nitrogen after collection and stored in −80 °C until RNA was extracted.

2.2. Identification of the HD-Zip I family genes in cotton

Protein sequences of the *Arabidopsis* HD-Zip I genes (17 in total) downloaded from the Arabidopsis Information Resource (TAIR, http://arabidopsis.org/) were used as queries to search against the annotated protein sequences (blastp, E value < 10^{-10}) of *G. raimondii* (Paterson et al., 2012), *G. arboreum* (Li et al., 2014) and *G. hirsutum* (Zhang et al., 2015). After building a list of unique protein sequence hits for each species, we analyzed all sequences to identify HD-Zip I genes by selecting those containing only the homeobox (PF00046) and the homeobox

Please cite this article as: Zhang, J., et al., Genome-wide identification and characterization of the homeodomain-leucine zipper I family of genes in cotton (*Gossypium* spp.), (2016), http://dx.doi.org/10.1016/j.plgene.2016.05.002