Gene 576 (2016) 150–159

Contents lists available at ScienceDirect

Gene

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

Research paper

Transcriptome-wide analysis of basic helix-loop-helix transcription factors in *Isatis indigotica* and their methyl jasmonate responsive expression profiling

Lei Zhang [a,b,1], Junfeng Chen [a,1], Qing Li [a], Wansheng Chen [a,*]

[a] Department of Pharmacy, Changzheng Hospital, Second Military Medical University, Shanghai 200003, China
[b] Department of Pharmaceutical Botany, School of Pharmacy, Second Military Medical University, Shanghai 200433, China

ARTICLE INFO

Article history:
Received 27 February 2015
Received in revised form 6 September 2015
Accepted 30 September 2015
Available online 9 October 2015

Keywords:
Basic helix-loop-helix transcription factors
Isatis indigotica
Lignan
Methyl jasmonate
Radix isatidis
Transcriptome profiling

ABSTRACT

Jasmonates (JAs) act as conserved elicitors of plant secondary metabolism. JAs perception triggers extensive transcriptional reprogramming leading to activation of the entire metabolic pathways. The family of basic helix-loop-helix (bHLH) transcription factors (TFs) has essential roles in JA signaling; however, little is known about their roles in regulation of secondary metabolites in *Isatis indigotica*. In this study, we identified 78 putative IibHLH sequences using the annotation of *I. indigotica* transcriptome. The identified proteins were characterized based on phylogenetic and conserved motif analyses. Using RNA sequencing, 16 IibHLHs showed significant positive response to MeJA (methyl jasmonate) at 1 h, indicating their roles as early signaling events of JA-mediated transcriptional reprogramming. Ten IibHLHs presented co-expression pattern with biosynthetic pathway genes, suggesting their regulating role in secondary metabolite synthesis. These gene expression profiling data indicate that bHLHs can be used as candidate genes in molecular breeding programs to improve metabolite production in *I. indigotica*.

1. Introduction

Basic helix-loop-helix (bHLH) proteins are widely distributed in eukaryotic kingdoms (Riechmann and Ratcliffe, 2000; Ledent and Vervoort, 2001). Among the plant transcription factor (TF) families, bHLH proteins constitute the second largest class. bHLH proteins execute transcriptional regulation of wide range of biological processes in plants, including stress tolerance, organ development, hormone responses, and metabolite synthesis (Kazan and Manners, 2013). MYC-type proteins are the most thoroughly studied bHLH proteins in plants. These jasmonate ZIM-domain (JAZ) interaction domain (JID)-contained proteins are involved in the initial stage of the jasmonate (JA) signaling process (Dombrecht et al., 2007; Fernández-Calvo et al., 2011). JA-mediated reprogramming in plants is subjected to complex mechanisms integrated into cellular networks. JA triggers an extensive transcriptional reprogramming of metabolism by TFs. The primary signal transduction processes following JA perception converge on related bHLH TFs. Among these TFs, the best characterized and most multifunctional proteins are the MYC proteins. Recent studies have indicated the advanced functions executed by MY2 as the JA signaling switch (Dombrecht et al., 2007; Fernández-Calvo et al., 2011). JA-Ile is rapidly synthesized in response to pest and pathogen attack or wounding. The binding of JA-isoleucine (JA-Ile) to JAZ-Coronatine Insensitive 1 (JAZ-COI1) co-receptor complexes leads to removal of JAZ repressors by the 26S proteasome pathway (Chini et al., 2007; Sheard et al., 2010; Thines et al., 2007). The released MYC proteins bind to G-box *cis*-acting sequences in target promoters (Boter et al., 2004; Yadav et al., 2005) in a dimer form (Fernández-Calvo et al., 2011) to regulate the transcription of target genes.

JA signaling has been linked to essential roles in many plant processes, ranging from plant defense to growth and development (Wasternack, 2007). Some of these processes are mediated by reprogramming of secondary metabolites. Thus, JA signaling is of particular interest in plant cell engineering for producing bioactive compounds (De Geyter et al., 2012; Yamada et al., 2011). JA plays significant regulatory roles in biosynthesis of various classes of valuable secondary metabolites. Biosynthesis of bioactive terpenoids such as terpenoid indole alkaloids (TIA) (Rischer et al., 2006), taxol (Yukimune et al., 1996; Ketchum et al., 2003; Li et al., 2012), artemisinin (Caretto et al., 2011), and tanshinone (Gu et al., 2012) is induced by exogenous MeJA in plants. Accumulations of various phenylpropanoids are MeJA-inducible, such as anthocyanidin (Shan et al., 2009), flavonoids (Pauwels et al., 2008), and phenolic acids (Xiao

Abbreviations: bHLH, basic helix-loop-helix; DEG, different expression genes; HCL, hierarchical clustering; JA, jasmonate acid; JAZ, jasmonate ZIM-domain; JID, JAZ-interacting domain; MeJA, methyl jasmonate; ORF, open reading frames; RACE, rapid amplification of cDNA ends; TF, transcriptional factor; TIA, terpenoid indole alkaloids.

* Corresponding author.
E-mail addresses: zhanglei@smmu.edu.cn (L. Zhang), chenjf@smmu.edu.cn (J. Chen), qli@smmu.edu.cn (Q. Li), chenwanshengs@smmu.edu.cn (W. Chen).

[1] Equal contribution.

http://dx.doi.org/10.1016/j.gene.2015.09.083
0378-1119/

et al., 2009). Therefore, the JA-induced plant tissues provide an inducible system to identify and profile the transcripts and regulatory factors involved in secondary metabolite accumulation.

Isatis indigotica Fortune, which belongs to the *Brassicaceae* family, is a widely used Chinese herb with medicinal and economic significance. *Radix isatidis* (ban-lan-gen) is the root of *I. indigotica*, which is called 'Banlangen' in Chinese, has been used clinically to treat colds, fever, and influenza. To date, more than 70 components (mainly including indole alkaloids, flavonoids, and lignans) have been isolated and characterized from *I. indigotica*. As secondary metabolites, several active components are present with low natural abundance in *I. indigotica*. The content of lariciresinol is only 0.24‰ of dry weight (DW), lariciresinol-4′-bis-O-β-D-glucopyranoside is 0.4‰ DW, and hypoxanthine is 0.49‰ DW (QIN et al., 2009; Lin et al., 2005; Zhu et al., 2011). A large-scale transcriptome analysis project was performed to determine the biosynthetic processes of the bioactive compounds in *I. indigotica* to understand and enhance the biosynthesis of active metabolites (Chen et al., 2013). This process allows wide identification and characterization of genes involved in biosynthesis and regulation of active metabolites.

Previous studies have shown that transcripts of biosynthesis pathway and accumulation of lignans in *I. indigotica* hairy roots are both induced by JA treatment (Chen et al., 2013), suggesting the JA-induced *I. indigotica* cultures are ideal systems for investigating the synthetic characteristics and regulation of secondary metabolites. In this study, the transcriptional profiling of *I. indigotica* hairy root culture in response to MeJA treatment was detected by RNA sequencing (RNA-seq). We identified 78 bHLH genes form *I. indigotica* transcriptome to obtain an overall knowledge of bHLH family in *I. indigotica* and further investigation for JA-responsive bHLHs. Our goals were to (i) set up a data set with abundant transcript sequences of *I. indigotica* bHLH genes, (ii) establish a functional classification by protein structure and domain analysis, and (iii) utilize the power of expression analysis to discover candidate IibHLHs involved in JA-mediated regulation of secondary metabolism.

2. Results

2.1. Identification of bHLH transcriptional factors

In previous work, transcriptome sequencing and annotation of *I. indigotica*, which contains 36,367 unigenes, provide dataset for gene screening and transcriptional profiling in this un-model plant for which genomic sequence information is not yet available (Chen et al., 2013). Using Blastn search against *I. indigotica* transcriptome, 78 unigenes coding bHLH were identified from *I. indigotica* transcriptome assembly, making it the second largest family of transcription factors in *I. indigotica* (data not shown). No clear bHLH domain was found in isogene (Assembly number: comp1603); thus, this sequence was not further used in this study. By open reading frame (ORF) prediction, total 72 unigenes were identified to have full-length cDNA sequence. Six bHLHs, identified without full length cDNA sequence, were obtained by Rapid Amplification of cDNA Ends (RACE) method. All cDNA sequences were translated into proteins for further phylogenetic and protein structure analysis (Supplementary data S1). For simplicity, each putative protein was subjected to Basic Local Alignment Search Tool (BLAST) analysis separately and renamed based on the homologous gene in *Arabidopsis* (Table 1). Moreover, three bHLHs (Assembly number: comp1297, comp29429, and comp31022) had no significant similarity to any known bHLH in the available public database.

Large numbers of bHLH proteins were found in land plant species (embryophytes). *Arabidopsis* and *Oryza sativa* have more than 150 bHLH sequences in their genomes (Bailey et al., 2003; Li et al., 2006; Carretero-Paulet et al., 2010). In our study, we identified a smaller family of 78 bHLHs from *I. indigotica* transcriptome. This result may be due to its genome size (Satapathy et al., 2014). Conditionally expressed bHLHs may exist and are undetectable in the current transcriptome project of *I. indigotica*. Hence, a genome-wide analysis is expected to fully identify bHLH in *I. indigotica*.

2.2. Phylogenetic analysis

Based on phylogenetic relationships, DNA-binding motifs, and functional properties, animal bHLH proteins are divided into six major groups named as group A to F (Atchley and Fitch, 1997). The genomes of plants encode more bHLH sequences than those of animals. Previous phylogenetic studies proposed the classification of plant bHLHs into 15–25 subgroups (Bailey et al., 2003; Buck and Atchley, 2003; Heim et al., 2003). Recently, the diversity of bHLHs from multiple plant species was classified into 26 distinct subfamilies (Pires and Dolan, 2010). Using neighbor-joining (NJ) phylogenetic tree analysis and plant bHLH group nomenclature proposed by Nuno and Liam (Pires and Dolan, 2010), 78 IibHLHs were subdivided into 20 subfamilies (Fig. 1; Supplementary Table S3). Among all the subfamilies, subfamily XI was the largest clade, which included 11 bHLHs. IibHLH54, 38, and 144 were classified into subfamilies constituted by a single member based on the classification of their homologous proteins. Our result showed a similar match with previous classification of plant bHLHs. Moreover, members in certain subfamilies showed relative functional annotation, further supporting our subfamily definition. For example, IiMYC2, 3, 4, and two MYC-like proteins were clustered in subfamily III (d + e), which may be related to JA signaling (Dombrecht et al., 2007; Fernández-Calvo et al., 2011). Clade VII (a + b) contains two sub-clades, in which SPATULA (SPT) and ALCATRAZ (ALC) were clustered and may be required for gynoecium and fruit development (Groszmann et al., 2011). Phytochrome interacting factor (PIF) proteins,which are light response-associated regulators (Sentandreu et al., 2012), were clustered in another sub-clade. Moreover, BIM1 proteins were clustered in Clade VIIIb, suggesting their role in brassinosteroid signaling (Chandler et al., 2009).

2.3. Conserved motif of I. indigotica bHLH family

Key amino acid residues are highly conserved between plant bHLH motifs. The bHLH domain is highly conserved and defines the bHLH transcriptional family. This domain contains approximately 60 amino acids with two functionally distinct regions (Supplementary Data S4). The basic region determines the DNA binding ability, and is critical for function (Massari and Murre, 2000). IibHLHs were grouped based on DNA-binding prediction through alignment of bHLH domains (Fig. 2; Supplementary Fig. S5). Eight IibHLHs (IibHLH35, 135, 144, 155, comp29429, comp31019, and PIF6) were considered to lack DNA-binding ability because of the absence of certain conserved residues in the *N*-terminal of the bHLH domain. The remaining 70 IibHLHs were further grouped depending on the presence of residues Glu-9 and Arg-16 in the basic region (Li et al., 2006; Pires and Dolan, 2010). Among which, 61 were predicted as E-box-binding proteins with conserved Glu-9 and Arg-13. Seven IibHLHs that lack residues Glu-9 and Arg-13 were non-E-box binding proteins. Meanwhile, three residues in the basic region of the bHLH domain, namely, His/Lys-5, Glu-9, and Arg-13, constituted a classic G-box-binding region (Massari and Murre, 2000). Therefore, 47 IibHLHs, which contained His/Lys-5, Glu-9, and Arg-13 residues, were grouped into G-box-binding proteins. The result showed that most IibHLHs (60.25%) were G-box-binding proteins, which was similar to *Arabidopsis* (60.54%) and *O. sativa* (56.89%) (Li et al., 2006).

The helix region promotes the formation of homo- or hetero dimeric complexes between bHLH proteins. Certain highly conserved positions are essential for dimerization stability (Brownlie et al., 1997). A leucine (L) residue was present in sites 27 and 59 of IibHLHs, constituting 100% and 95%, respectively. Sites 50 and 57 had isoleucine (I), L, orvaline (V) in 99% and 89% of IibHLH. These highly conserved residues suggest their roles in dimerization or other functions of IibHLH proteins.