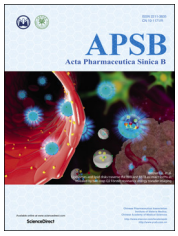
Acta Pharmaceutica Sinica B 2018;8(2):295–305

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



ORIGINAL ARTICLE

Genomic survey of bZIP transcription factor genes related to tanshinone biosynthesis in *Salvia miltiorrhiza*

Yu Zhang, Zhichao Xu, Aijia Ji, Hongmei Luo, Jingyuan Song*

Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of the People's Republic of China, Institute of Medicinal Plant Development, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100193, China

Received 27 May 2017; received in revised form 24 July 2017; accepted 6 August 2017

KEY WORDS

bZIP genes;
Salvia miltiorrhiza;
Phylogenetic analysis;
Expression pattern analysis;
Tanshinone biosynthesis

Abstract Tanshinones are a class of bioactive components in the traditional Chinese medicine *Salvia miltiorrhiza*, and their biosynthesis and regulation have been widely studied. Current studies show that basic leucine zipper (bZIP) proteins regulate plant secondary metabolism, growth and developmental processes. However, the bZIP transcription factors involved in tanshinone biosynthesis are unknown. Here, we conducted the first genome-wide survey of the bZIP gene family and analyzed the phylogeny, gene structure, additional conserved motifs and alternative splicing events in *S. miltiorrhiza.* A total of 70 SmbZIP transcription factors were identified and categorized into 11 subgroups based on their phylogenetic relationships with those in *Arabidopsis*. Moreover, seventeen *SmbZIP* genes underwent alternative splicing events. According to the transcriptomic data, the *SmbZIP* genes that were highly expressed in the Danshen root and periderm were selected. Based on the prediction of bZIP binding sites in the promoters and the co-expression analysis and co-induction patterns in response to Ag^{+} treatment *via* quantitative real-time polymerase chain reaction (qRT-PCR), we concluded that *SmbZIP7* and *SmbZIP20* potentially participate in the regulation of tanshinone biosynthesis. These results provide a foundation for further functional characterization of the candidate *SmbZIP* genes, which have the potential to increase tanshinone production.

*Corresponding author.
E-mail address: jysong@implad.ac.cn (Jingyuan Song).

Peer review under responsibility of Institute of Materia Medica, Chinese Academy of Medical Sciences and Chinese Pharmaceutical Association.

http://dx.doi.org/10.1016/j.apsb.2017.09.002

1. Introduction

Coronary heart disease (CHD) is one of the leading causes of death worldwide, and the number of CHD patients is projected to reach 82 million by 2020[1,2]. *Salvia miltiorrhiza* Bunge (Danshen) is one of the most important traditional Chinese medicines, and it has been widely applied to treat CHD[3]. Tanshinones represent the major bioactive constituents of *S. miltiorrhiza*, and they exhibit strong anti-atherosclerosis, antioxidant, anti-cancer and anti-inflammatory activities[4–7]. More than 40 lipophilic diterpenoids (*e.g.*, tanshinone I, tanshinone IIA, cryptotanshinone and dihydrotanshinone) have been isolated and identified from *S. miltiorrhiza*[8,9]. Tanshinone is synthesized and accumulates in the root periderm of *S. miltiorrhiza*[10]. Many genes encoding key enzymes in the tanshinone biosynthetic pathway have been cloned and analyzed, including *SmGGPPS*, *SmCPS*, *SmKSL*, *SmCYP76AH1*, *CYP76AH3* and *CYP76AK1*[11]. Biotic and abiotic elicitors have been reported to increase the accumulation of tanshinone in *S. miltiorrhiza* hairy roots[12–15]. In addition, a total of 15 *SmSPL*s[16], 110 *SmMYB*s[17], 61 *SmWRKY*s[18], 127 *SmbHLH*s[19], 170 *SmAP2/ERF*s[20] and 25 *SmARF*s[21] from *S. miltiorrhiza* have been identified and characterized. SmMYC2a and SmMYC2b are two SmbHLH transcription factors that have been reported to positively regulate genes in the tanshinone biosynthetic pathway in *S. miltiorrhiza*[22]. Although tanshinone biosynthesis has attracted widespread attention, studies on the basic leucine zipper (bZIP) transcription factors (TFs) involved in the regulation of tanshinone biosynthesis are limited.

The bZIP TF family is one of the largest and most conserved gene families among eukaryotic organisms. The bZIP TFs have a highly conserved bZIP domain composed of a basic region that binds DNA and a leucine zipper region that confers dimerization specificity[23]. The basic region preferentially binds to DNA sequences with an ACGT core, particularly G-box (CACGTG), C-box (GACGTC) and A-box sequences (TACGTA)[24–26]. The TGACG motif is also bound by bZIP TFs[27]. Importantly, bZIP TFs have been demonstrated to participate in the transcriptional regulation of secondary metabolism[28,29]. bZIP TFs have been identified extensively in the plant kingdom based on the availability of whole genome sequences[30–36]. The complete genomic sequence of *S. miltiorrhiza* provides the opportunity to identify the sequences and analyze the features of the bZIP gene family[37].

In this study, we performed the first investigation and analysis of the members of the bZIP TF family. Based on the transcriptome data as well as the results of quantitative real-time polymerase chain reaction (qRT-PCR) analysis and predictions of *cis*-regulatory elements in the promoters, 2 *SmbZIP* genes (*SmbZIP7* and *SmbZIP20*) related to the regulation of tanshinone biosynthesis were selected. Our results are crucial for analyzing the putative functions and regulatory mechanisms of *SmbZIP* genes in *S. miltiorrhiza*.

2. Materials and methods

2.1. Plant materials and Ag^+ treatment

S. miltiorrhiza Bunge (line 99–3) was cultivated in a field nursery at the Institute of Medicinal Plant Development in Beijing. The roots, stems, leaves and flowers were collected in May 2016, and certain root samples were divided into three parts (periderm, phloem and xylem). Three biological replicates of different organs and tissues were frozen for further use.

S. miltiorrhiza hairy roots were cultured in 250 mL beaker flasks with 100 mL 6, 7-V liquid medium on an orbital shaker. Induction was started 18 days after 0.2 g (fresh weight) of the hairy roots was inoculated in each beaker flask. The cultures were treated with Ag^+ at a final concentration of 30 μmol/L and harvested at 0, 0.5, 1.5, 3, 6, 12, and 24 h post-treatment. Each treatment was replicated three times. The methods for Ag^+ preparation were described in a previous study[38]. All samples were stored at –80 °C until further study.

2.2. Identification of SmbZIP genes

A hidden Markov model (HMM: PF00170) of the bZIP domain was used to search all putative *bZIP* genes from the *S. miltiorrhiza* genome database (SRA accession: SRP051524). The presence of a bZIP domain in the selected bZIP proteins was further verified using the online program SMART (http://smart.embl-heidelberg.de/). The sequences of conserved SmbZIP domains were obtained using an NCBI program (http://ncbi.nlm.nih.gov/). In addition, the theoretical isoelectric point (pI) and molecular weight (MW) of the SmbZIP proteins were calculated using the ExPASy program (http://web.expasy.org/).

2.3. Bioinformatic analysis

The AtbZIP protein sequences were downloaded from the Arabidopsis Information Resource (http://Arabidopsis.org/). Multiple alignments of amino acid sequences of the bZIP domain of SmbZIP and AtbZIP proteins were performed by ClustalX 1.83, and an unrooted neighbor-joining tree was generated using MEGA 5.0 with 1000 bootstrap replicates. Conserved motifs outside the bZIP domain were identified by the MEME program (http://meme-suite.org/tools/meme) with the following parameters: any number of repetitions, maximum number of motifs less than 50, an optimum width of 10–200 amino acids, and expected E-values less than 1×10^{-48}. The intron/exon structure of the cDNAs and corresponding genomic sequence of *S. miltiorrhiza* were displayed using the online Gene Structure Display Server analysis tool (http://gsds.cbi.pku.edu.cn/index.php). The alternative splicing isoforms were analyzed using IGV 2.3.34 software (http://www.broadinstitute.org/software/igv/). Additionally, cis-elements for bZIP TFs were searched by analyzing the 1500-bp promoter sequences of 72 genes encoding key enzymes involved in tanshinone biosynthesis using the PLACE database[19] (http://www.dna.affrc.go.jp/PLACE/signalscan.html).

2.4. Analysis of RNA-Seq expression data

A total of 17 sequence libraries were constructed for the expression analyses of the *SmbZIP* genes, including four organs (root, stem, leaf, and flower) and three root tissues (periderm, phloem, and xylem)[10]. RNA-Seq reads were obtained with the Illumina HiSeq. 2000 and 2500 platforms (SRA accessions: SRR1640458, SRP051564, and SRP028388). The RPKM (reads per kilobase per million) values were calculated based on the RNA-Seq reads[10]. For the organ-specific and tissue-specific gene expression profiles of the SmbZIP gene family, heat maps were generated with the R statistical package.

2.5. *RNA extraction and qRT-PCR*

Total RNA was extracted from each sample using TRIzol reagent (Tiangen Biotech, Beijing, China) according to the manufacturer's procedure. The quality and integrity of the total RNA were tested by agarose gel electrophoresis and a NanoDrop 2000C spectrophotometer, respectively. The first-strand cDNA was synthesized from 1.0 μg of total RNA using PrimeScript Reverse Transcriptase (TAKARA) for qRT-PCR based on the manufacturer's instructions. Gene-specific primers for qRT-PCR were designed using Primer Premier 6.0 software with the amplicon length ranging from 130 to 180 bp and an optimal Tm of 54 ± 1 °C (Supplementary Information Table S1). Primer specificity was checked by 2% agarose gel electrophoresis. The qPCR reactions were performed using the SYBR premix Ex Taq kit (TaKaRa, China) and the Applied Biosystems 7500 Real-Time PCR system (Life Technologies, USA). The reaction conditions were as follows: 95 °C for 30 s, 40 cycles of 95 °C for 3 s, 60 °C for 30 s. *SmActin* was tested as the reference gene, and two key enzyme-coding genes (*SmKSL1* and *SmCYP76AH1*) downstream in the tanshinone biosynthetic pathway were used as positive controls. The gene expression levels of different samples were estimated by the Ct value using the $2^{-\Delta\Delta C_T}$ method[39]. To evaluate the differences in the expressions of the candidate genes among the various organs and tissues, one-way ANOVA was performed with SPSS 20 software. $P < 0.01$ was considered statistically significant. Co-expression analysis of candidate genes was performed using Pearson's correlation test, and coefficients greater than 0.8 indicated co-expression[40].

3. Results

3.1. *Genome-wide identification of the SmbZIP gene family and phylogenetic analysis*

A total of 77 *bZIP* genes were selected from the *S. miltiorrhiza* genome database on the basis of the HMM profile of the bZIP domain (PF00170; http://pfam.sanger.ac.uk/). Based on the survey of the presence of the bZIP domain using the SMART program with an E-value < 0.1, 70 genes encoding bZIP proteins were identified for the first time, and 7 other bZIP protein sequences were identified that did not have a complete bZIP domain. The 70 *SmbZIP* genes were named *SmbZIP1* to *SmbZIP70*. The length of the SmbZIPs varied from 107 to 851 amino acids, with an average of 347.7 amino acids. The deduced molecular weights of 70 SmbZIP proteins ranged from 13.19 kDa to 93.79 kDa, and the pI ranged from 5.16 to 11.83. The details of the SmbZIP protein sequences are summarized in Supplementary Information Table S2.

To evaluate the evolutionary relationships between the SmbZIP and AtbZIP proteins, an unrooted phylogenetic tree was constructed based on the alignment of the amino acid sequences of the bZIP domain with MEGA 5.0 (Fig. 1 and Supplementary Information Table S3). The SmbZIP TF family was classified into 11 subgroups (A, B, C, D, E, F, G, H, I, S, and U; Fig. 1). Subgroup S had the

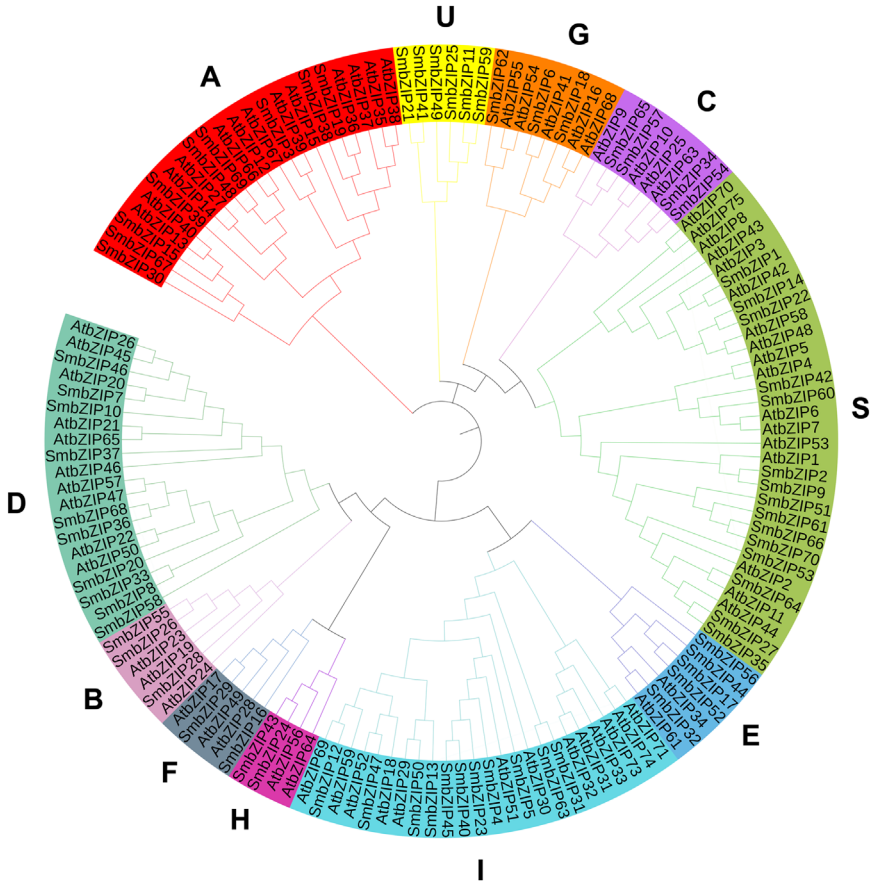


Figure 1 Phylogenetic tree of *S. miltiorrhiza* and *A. thaliana* bZIP proteins. The bZIP domain protein sequences of *S. miltiorrhiza* and *Arabidopsis* were aligned with ClustalX 1.83, and the phylogenetic tree was constructed using MEGA 5.0 based on the neighbor-joining (NJ) method. The bootstrap value was 1000 replicates. The proteins were classified into 11 distinct clusters. The colored branches indicate the different subgroups.

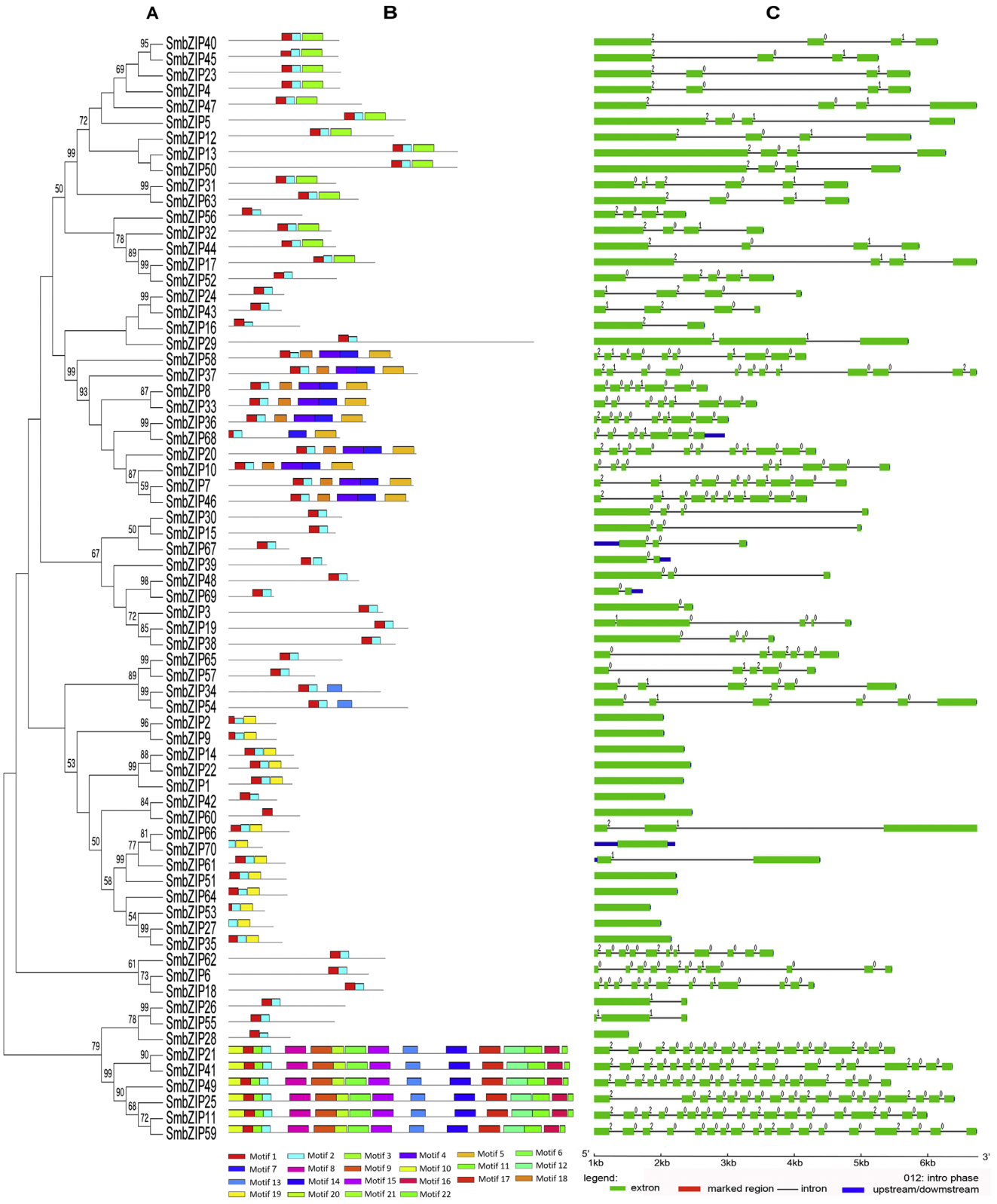


Figure 2 Characterization of the *SmbZIP* genes. (A) Phylogenetic relationships of SmbZIP proteins. (B) Conserved motifs of SmbZIP proteins. Different motifs are highlighted with different colored boxes with numbers 1 to 22. (C) Exon/intron organization of SmbZIP proteins. The exons and introns are represented by green boxes and black lines, respectively. The numbers "0" and "2" denote different splicing phases, with "0" indicating that splicing occurred after the third nucleotide of the codon and "2" indicating that splicing occurred after the second nucleotide.

most members (15 proteins), followed by subgroup I (11 proteins), whereas subgroups F and H had the fewest members (2 proteins). Moreover, most of the SmbZIP proteins clustered together with AtbZIP proteins in each clade. However, 6 SmbZIP proteins (SmbZIP11, SmbZIP21, SmbZIP25, SmbZIP41, SmbZIP49, and SmbZIP59) formed one small unique clade (indicated by a yellow branch in Fig. 1), which was named subgroup U according to the nomenclatures of the legume bZIP proteins[41].

3.2. Detection of conserved motifs, gene structure and alternative splicing events

A total of 22 conserved motifs (E-value cutoff $<1\times10^{-48}$), including the bZIP domain, were identified in the SmbZIP proteins, and the multilevel consensus amino acid sequences of the motifs are listed in Supplementary Information Fig. S1. The motif distribution and its correspondence to the phylogenetic tree of the SmbZIP gene family are shown in Figs. 2 A and B. Motif 1 corresponded to the basic region and the hinge of the bZIP domain, and motifs 2 and 21 occurred in the leucine zipper region. All subgroups at least presented either motif 1 or motif 2. However, motif 21 appeared exclusively in subgroup U. Outside the bZIP domain region, the bZIP proteins usually contained additional conserved motifs. The SmbZIP proteins within each subgroup shared similar additional motif compositions. Certain motifs were shared by several subgroups; for example, motif 3 occurred in subgroups I and E, and motif 13 occurred in subgroups C and U. Nonetheless, most of the conserved motifs appeared to be subgroup specific. For example, motifs 4, 5, 7 and 18 were detected in subgroup D; motif 19 was found exclusively in subgroup S; and motifs 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20 and 22 were observed in subgroup U. Certain additional conserved motifs had potential functional sites and might contribute to functional specialization. For instance, motif 7, which includes a proline-rich domain, might mediate protein–protein interactions, and motif 14, a glutamine-rich domain, might be associated with transcriptional activation.

The exon/intron organizations of the SmbZIP gene family are shown in Fig. 2 C, including the number, distribution and splicing phase of the exons. Interestingly, 13 (18.5%) of the 70 *SmbZIP* genes lacked introns, all of which clustered in subgroup S. Among other genes containing introns, the number of introns varied from 1 to 17, which is similar to the highest number of introns within the ORFs reported in sorghum (14)[33] and maize (14)[42]. The position of exons in each subgroup was diverse. However, a similar number and splicing phase of exons were present in most of the subgroups. For example, the number and splicing phase of exons were relatively conserved in subgroups A, B, C, E, F, G, H, I and U, whereas the number and splicing phase of exons varied significantly in subgroups D and S.

Based on the transcriptomic data, 24% of *SmbZIP* genes have alternative splice isoforms (Supplementary Information Fig. S2). Eight genes (*SmbZIP13*, *SmbZIP17*, *SmbZIP21*, *SmbZIP33*, *SmbZIP34*, *SmbZIP37*, *SmbZIP46*, and *SmbZIP50*) each expressed two isoforms, five genes (*SmbZIP3*, *SmbZIP12*, *SmbZIP52*, *SmbZIP54*, and *SmbZIP65*) each expressed three isoforms, *SmbZIP11* expressed four isoforms, *SmbZIP10* and *SmbZIP18* each expressed five isoforms, while *SmbZIP59* expressed seven isoforms (Supplementary Information Table S4). Among their respective alternative splicing events, the gene loci of *SmbZIP10*, *SmbZIP12*, *SmbZIP13*, *SmbZIP33*, *SmbZIP37*, *SmbZIP50*, *SmbZIP17* and *SmbZIP65* significantly expressed one isoform, while the isoforms of *SmbZIP3*, *SmbZIP11*, *SmbZIP18*, *SmbZIP21*, *SmbZIP34*, *SmbZIP46*, *SmbZIP52*, *SmbZIP54* and *SmbZIP59* exhibited similar expressions (Supplementary Information Table S4).

3.3. Analysis of bZIP binding sites in the promoters of pathway genes

The promoter sequences (1500 bp) of 72 enzyme-coding genes in the tanshinone biosynthetic pathways were analyzed for the presence of bZIP binding sites (A-box, G-box, C-box and TGACG motifs)[24–27]. Importantly, the bZIP binding sites were located in the promoter regions of 46 enzyme-coding genes, including 4 functionally characterized genes (*SmHMGR2*, *SmCPS1*, *SmKSL1* and *SmCYP76AH1*, Supplementary Information Table S5). The A-box was predicted to occur in the promoter sequences of 16 enzyme-coding genes, the C-box was predicted to occur in the promoter sequences of 4 enzyme-coding genes, the G-box was predicted to occur in the promoter sequences of 4 enzyme-coding genes, and the TGACG motif was predicted to occur in the promoter sequences of 35 enzyme-coding genes. Interestingly, 12 pathway genes had at least two bZIP binding sites. These results indicated that SmbZIP TFs might be involved in the regulation of these genes.

3.4. Differential expression of SmbZIP genes among different organs and tissues

The differential expression profiles of the *SmbZIP* genes among various organs and root tissues were analyzed based on RNA-Seq data (Fig. 3 and Supplementary Information Table S6). Among the 70 *SmbZIP* genes, the transcripts for 66 genes were found in at least one of four organs and largely varied among the organs (RPKM values higher than 0; Fig. 3 A). Eight genes (*SmbZIP8*, *SmbZIP20*, *SmbZIP36*, *SmbZIP50*, *SmbZIP51*, *SmbZIP60*, *SmbZIP64* and *SmbZIP68*) were predominantly expressed in the root (at least 2-fold higher than in other organs, with RPKM values greater than 1), 3 genes (*SmbZIP11*, *SmbZIP25* and *SmbZIP47*) were predominantly expressed in the stem, only one gene (*SmbZIP38*) was predominantly expressed in the leaf, and 8 genes (*SmbZIP15*, *SmbZIP31*, *SmbZIP32*, *SmbZIP34*, *SmbZIP44*, *SmbZIP52*, *SmbZIP57* and *SmbZIP63*) were predominantly expressed in the flower. Of the 8 genes with high expression in the root, 3 (*SmbZIP8*, *SmbZIP20* and *SmbZIP60*) were highly expressed in the periderm (approximately 2-fold higher than in the other organs, with RPKM values greater than 1), 1 (*SmbZIP68*) was highly expressed in the phloem and xylem, and 1 (*SmbZIP36*) was highly expressed in the phloem. Three other genes (*SmbZIP50*, *SmbZIP51* and *SmbZIP64*) were found to have similar expression profiles in various root tissues.

As illustrated in Fig. 3B, the expression levels of 66 *SmbZIP* genes were detected in at least one of the three root tissues, although their expression levels varied among the different tissues. Twelve *SmbZIP* genes (*SmbZIP5*, *SmbZIP7*, *SmbZIP8*, *SmbZIP20*, *SmbZIP23*, *SmbZIP28*, *SmbZIP30*, *SmbZIP34*, *SmbZIP37*, *SmbZIP38*, *SmbZIP52* and *SmbZIP60*) had high expression in the periderm, whereas 6 *SmbZIP* genes (*SmbZIP1*, *SmbZIP21*, *SmbZIP25*, *SmbZIP49*, *SmbZIP53* and *SmbZIP68*) were highly expressed in the phloem and xylem. We further investigated the expression of the *SmbZIP* genes with high expression in the periderm. Seven genes (*SmbZIP5*, *SmbZIP7*, *SmbZIP8*, *SmbZIP20*,

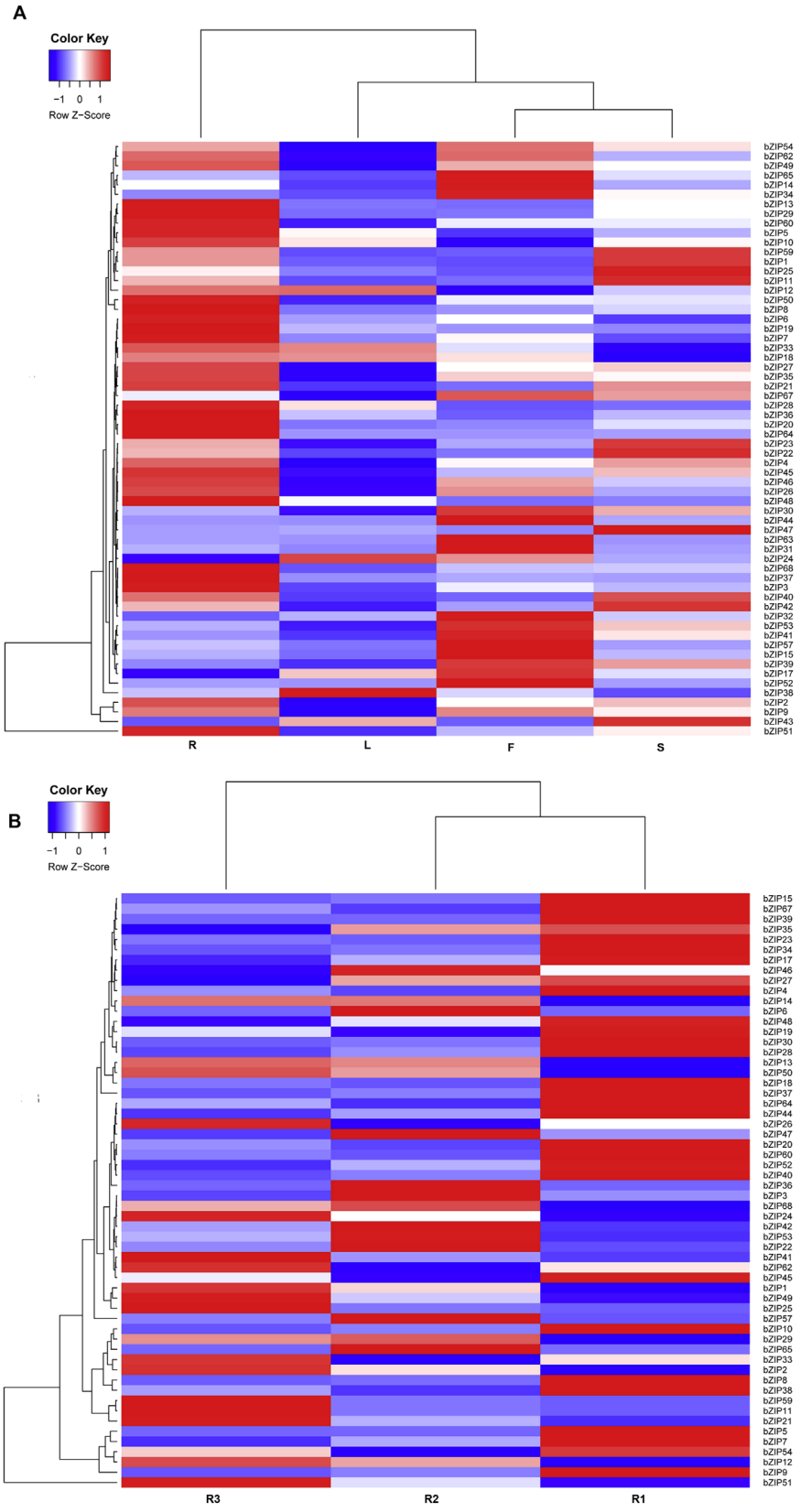


Figure 3 Heatmaps of the differentially expressed *SmbZIP* genes in various organs and root tissues. (A) Expression profiles of the *SmbZIP* genes in the flower (F), leaf (L), root (R) and stem (S). (B) Expression profiles of the *SmbZIP* genes in the periderm (R1), phloem (R2) and xylem (R3). Expression values are mapped to a color gradient from low (blue) to high (red) expression.

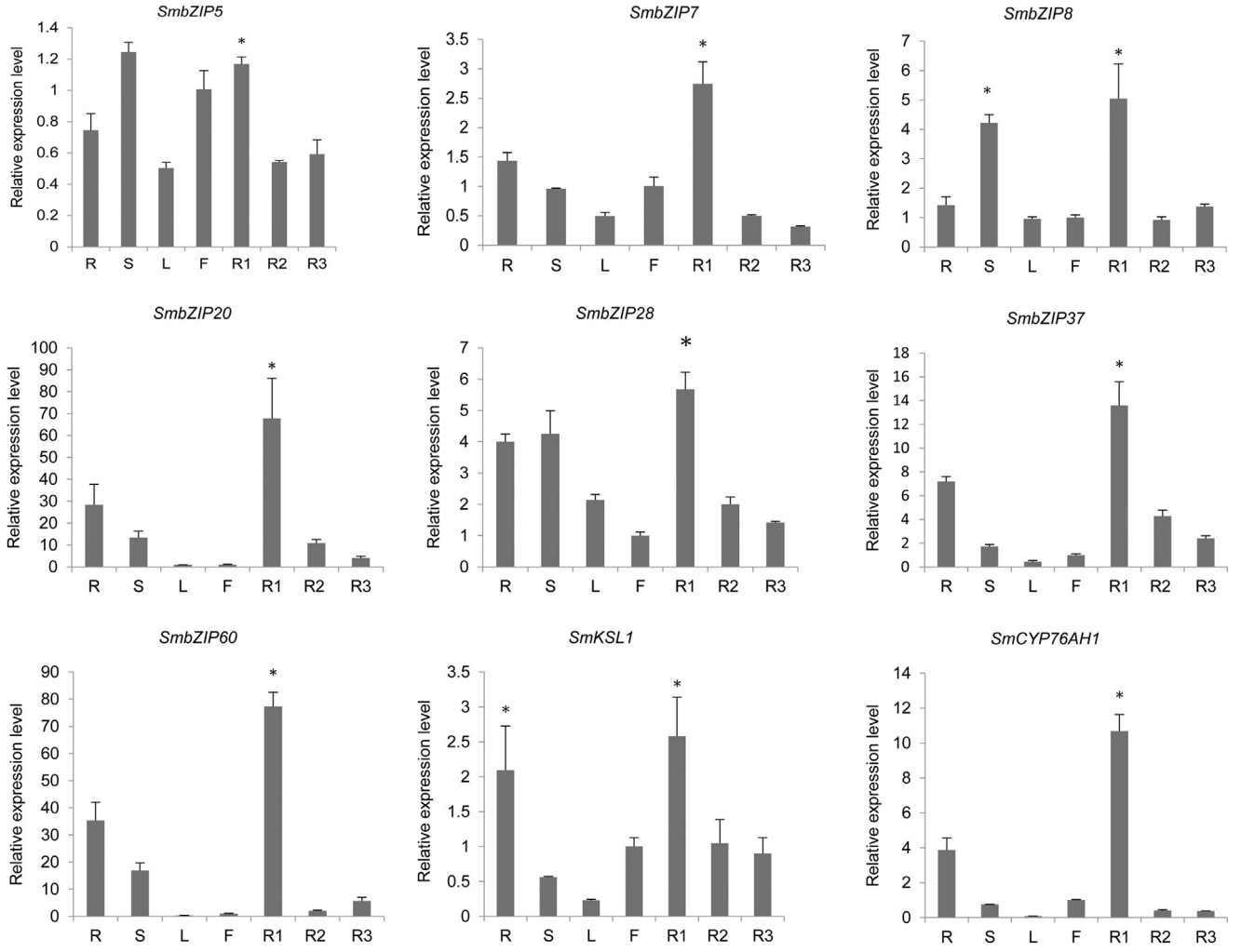


Figure 4 Expression profiles of the selected *SmbZIP* genes in the flower (F), leaf (L), root (R), stem (S), periderm (R1), phloem (R2) and xylem (R3). The relative expressions of 7 selected *SmbZIP* genes were determined using quantitative real-time RT-PCR (qRT-PCR). *SmkSL1* and *SmCYP76AH1* were tested as positive controls. Asterisks represent significant differences. $P < 0.01$ was considered highly significant.

SmbZIP28, *SmbZIP37* and *SmbZIP60*) exhibited high expression in the root, and three genes (*SmbZIP30*, *SmbZIP34* and *SmbZIP52*) were highly expressed in the flower, whereas two genes (*SmbZIP23* and *SmbZIP38)* were mainly expressed in the stem and leaf, respectively.

The expression profiles of the *SmbZIP* genes from the transcriptome data showed that 7 *SmbZIP* genes were highly expressed in the root and periderm, which is consistent with the site of tanshinone biosynthesis and accumulation.

3.5. *Co-expression analysis of SmbZIP genes*

To confirm the RNA-Seq observations, changes in the transcript levels of 7 *SmbZIP* genes with high expression in the root and periderm were examined in different organs and tissues by qRT-PCR (Fig. 4, and Supplementary Information Table S7). All of the tested *SmbZIP* genes had higher expression levels in the periderm than in the phloem and xylem. Additionally, 3 (*SmbZIP5*, *SmbZIP8* and *SmbZIP28*) of the 7 genes were highly expressed in the stem, and their expression patterns did not match those obtained using RNA-Seq. However, another 4 *SmbZIP* genes (*SmbZIP7*, *SmbZIP20*, *SmbZIP37* and *SmbZIP60*) were markedly expressed in the root, and their expression patterns were consistent with those obtained using RNA-Seq. Moreover, the Pearson coefficients of these four genes and two pathway genes of tanshinone biosynthesis (*SmKSL1* and *SmCYP76AH1*) were greater than 0.8, respectively (Supplementary Information Table S8). Statistical analysis revealed that *SmbZIP7*, *SmbZIP20*, *SmbZIP37* and *SmbZIP60* were co-expressed with *SmKSL1* and *SmCYP76AH1*, implying that these genes may be associated with tanshinone biosynthesis.

3.6. *Identification of SmbZIP TFs related to tanshinone biosynthesis*

To further investigate the function of 4 *SmbZIP* genes in *S. miltiorrhiza,* qRT-PCR expression analysis was performed for

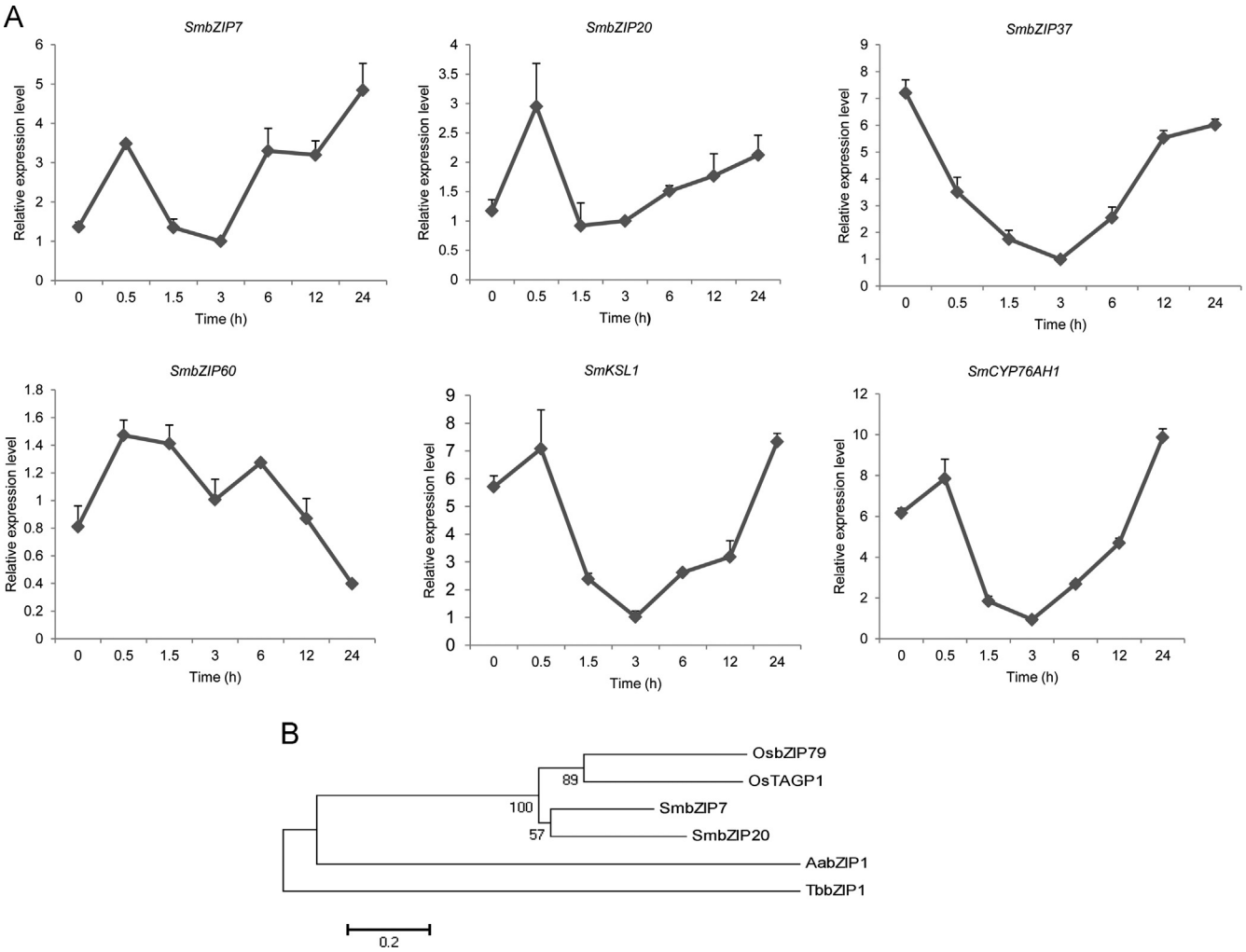


Figure 5 (A) Expression patterns of 4 selected *SmbZIP* genes in response to Ag^{+}. Relative expression levels of the genes in hairy roots upon 0 (Negative control), 0.5, 1.5, 3, 6, 12, and 24 h of Ag^{+} treatment are shown. *SmKSL1* and *SmCYP76AH1* were tested as positive controls. (B) Phylogenetic analysis of SmbZIP7, SmbZIP20 and other functional TFs that regulate terpenoid biosynthesis.

the Ag^{+}-treated hairy cultures (Supplementary Information Table S9). As illustrated in Fig. 5 A, *SmbZIP7*, *SmbZIP20* and *SmbZIP60* were up-regulated in response to Ag^{+} within 0.5 h of treatment compared with the untreated hairy roots, and their expression sharply decreased until 1.5 or 3 h. Subsequently, the transcript levels of *SmbZIP7* and *SmbZIP20* increased, whereas the expression level of *SmbZIP60* decreased gradually at 6, 12, and 24 h after Ag^{+} treatment. Importantly, the expression of *SmKSL1* and *SmCYP76AH1* was correlated with the expression of *SmbZIP7* and *SmbZIP20*. However, *SmbZIP37* was down-regulated for up to 3 h and up-regulated at 6, 12, and 24 h following induction. Generally, the results of the quantitative PCR analysis showed strong co-induction of *SmbZIP7*, *SmbZIP20, SmKSL1* and *SmCYP76AH1* after Ag^{+} treatment.

To identify the SmbZIP TFs that are potentially involved in the regulation of tanshinone, we analyzed the phylogenetic relationships between the candidate SmbZIP TFs and functional bZIP TFs that regulate terpenoid biosynthesis (*i.e.*, OsbZIP79, OsTAGP1, AabZIP1, TbbZIP1[43], Table S10). These TFs were grouped into three clades in the neighbour-joining tree (Fig. 5B). The SmbZIP7 and SmbZIP20 proteins clustered with OsbZIP79 and OsTAGP1 in clade 1, whereas AabZIP1 and TbbZIP1 belonged to clade 2 and 3, respectively. Consistent with the phylogenetic analysis, OsTGAP1, OsbZIP79, SmbZIP7 and SmbZIP20 are TGA-type TFs, which belong to subgroup D. Therefore, we concluded that members of subgroup D most likely participate in regulating tanshinone biosynthesis. Based on the expression patterns and phylogeny, *SmbZIP7* and *SmbZIP20* are implicated as having a functional role in tanshinone biosynthesis.

4. Discussion

Increasing evidence suggests that bZIP proteins play crucial roles in regulating secondary metabolism in plants[28,29]. Much of the research into the genome-wide analyses of *bZIP* genes has focused on plants such as *Arabidopsis*[30], rice[32], grapevine[35] and cucumber[44]. However, this TF family is poorly understood in medicinal plants. *S. miltiorrhiza* is considered a model medicinal plant in traditional Chinese medical research, and its dried roots and rhizomes are highly valued in the treatment of CHD[11]. This study combines genomics and transcriptomics to allow a genome-wide characterization of bZIP TFs and the *SmbZIP* genes involved in tanshinone biosynthesis in *S. miltiorrhiza*. Herbgenomics provides

an effective platform to analyze secondary metabolism in medicinal plants. Genomic and transcriptomic information together with proteomic and metabolomic data have facilitated the study of secondary metabolite biosynthetic pathways and their regulation[45]. These strategies provide references for further research on the potential regulatory roles of SmbZIP TFs in *S. miltiorrhiza*.

4.1. *Characterization of bZIP TFs*

In this study, the *bZIP* gene family was comprehensively identified in *S. miltiorrhiza* for the first time. Compared with other plants, the size of the SmbZIP family (70) is similar to that of *Arabidopsis* (75)[30], tomato (69)[46] and cucumber (64)[44], but smaller than rice (89)[32], maize (125)[42], sorghum (92)[33], barley (141)[47] and *B. distachyon* (96)[48]. This result is consistent with the notion that monocot plants harbor a relatively larger bZIP family than dicots. All of the bZIP proteins from *A. thaliana* and *S. miltiorrhiza* were subdivided into 11 subgroups (A–I, S, U; Fig. 1). Ten subgroups contained the bZIP proteins from *A. thaliana* and *S. miltiorrhiza*, whereas one other subgroup (U) was specific to *S. miltiorrhiza*. This result indicated that most of the *bZIP* genes of *Arabidopsis* and Danshen are closely related, whereas subgroup U possibly had evolutionary trajectories independent of the other groups. In addition, members of subgroup U had more introns and conserved domains than any other subgroup, implying that subgroup U might have unique functions. Although the function of subgroup U is still unknown, it is worth studying.

4.2. *Possible regulatory function of bZIP TFs in tanshinone biosynthesis*

In this study, we examined the expression levels of all *SmbZIP* genes based on RNA-Seq data. The expression patterns of the *SmbZIP* genes were further confirmed using qRT-PCR. Four *SmbZIP* genes exhibited high expression levels in the root and periderm, and these genes were co-expressed with *SmKSL1* and *SmCYP76AH1*. Furthermore, 2 (*SmbZIP7* and *SmbZIP20*) of the 4 genes were responsive to Ag^{+}, and the variation trends were similar to changes in *SmKSL1* and *SmCYP76AH1*. Considering that tanshinone is produced and accumulates in the root periderm of *S. miltiorrhiza*[10], the SmbZIP TFs expressed in the root periderm most likely participate in regulating tanshinone biosynthesis. Additionally, previous studies have shown that the expression profiles of TFs are usually consistent with the expression profiles of the enzyme genes they regulate[20]. Thus, *SmbZIP7* and *SmbZIP20* are predicted to be responsible for regulating tanshinone biosynthesis. bZIP regulators involved in terpenoid biosynthesis have been reported in previous studies. For instance, AabZIP1 and TbbZIP1 are involved in regulating artemisinin and rubber biosynthesis, respectively[49,50]. In rice, OsTGAP1 and OsbZIP79 were identified as regulators of the production of diterpenoid phytoalexins[51,52]. Phylogenetic analysis showed that the SmbZIP7 and SmbZIP20 proteins clustered with OsbZIP79 and OsTAGP1 in one clade (Fig. 5B). Importantly, tanshinones and phytoalexins are both diterpenoid compounds, and the biosynthetic pathways of these two compounds are similar. This similarity suggests that *SmbZIP7* and *SmbZIP20* might function in tanshinone biosynthesis. Further studies will focus on the functional verification of the candidate *SmbZIP* genes, such as through an analysis of transgenic *S. miltiorrhiza* plants with over-expressed or silenced *SmbZIPs*..

4.3. *Possible Regulatory Mechanism of bZIP TFs*

In the present study, 2 candidate SmbZIP TFs (SmbZIP7 and SmbZIP20) were predicted to participate in regulating tanshinone biosynthesis. The promoters of 46 enzyme-coding genes involved in tanshinone biosynthesis were found to contain bZIP binding sites (A-box, G-box, C-box or TGACG motifs). Previous studies have shown that OsTGAP1 regulates momilactone biosynthesis through the positive control of *OsKSL4* expression, which apparently occurs *via* direct binding to a *cis*-element in the *OsKSL4* promoter[51]. The TGACG motif is known as a *cis*-element for TGA TFs[27]. Because SmbZIP7 and SmbZIP20 are TGA TFs, the TGACG motif may also be their *cis*-element. In fact, the TGACG motif was found in the promoters of 35 tanshinone biosynthetic genes, including three functionally characterized genes (*SmHMGR2*, *SmKSL1* and *SmCYP76AH1*). Thus, we concluded that SmbZIP7 and SmbZIP20 may influence the expression of these genes by binding to the TGACG motif. Confirmation of SmbZIP binding sites on the tanshinone biosynthetic genes is the next step. Electronic mobility shift assays (EMSAs) were performed to test SmbHLH binding behavior in *S. miltiorrhiza*[22]. In addition, to examine whether AabZIP1 binds to ABRE, a Yeast One-Hybrid Assay was performed in *Artemisia annua*[49]. However, additional methods are needed to analyze the possible regulatory mechanisms of SmbZIP TFs.

5. Conclusions

We provided a first overview of the SmbZIP TF family in *S. miltiorrhiza*. We also investigated *SmbZIP* gene expression profiles under Ag^{+} treatment and in different organs and root tissues of *S. miltiorrhiza*. *SmbZIP7* and *SmbZIP20* are predicted to be potential regulatory factors related to tanshinone biosynthesis in *S. miltiorrhiza*. Our findings provide a foundation for further studies of the function of these genes *via* genetic engineering combined with metabolite profiling in *S. miltiorrhiza*.

Acknowledgments

This work was supported by the National Natural Science Foundation (Grant No. 81573398) and Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS, 2016-I2M-3–016).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.apsb.2017.09.002.

References

1 Leigh JA, Alvarez M, Rodriguez CJ. Ethnic minorities and coronary heart disease: an update and future directions. *Curr Atheroscler Rep* 2016;**18**:9.

2 Kanu JS, Gu Y, Zhi S, Yu M, Lu Y, Cong Y, et al. Single nucleotide polymorphism rs3774261 in the AdipoQ gene is associated with the risk of coronary heart disease (CHD) in Northeast Han Chinese population: a case-control study. *Lipids Health Dis* 2016;**15**:6.

3 Zhou L, Zuo Z, Chow MS. Danshen: an overview of its chemistry, pharmacology, pharmacokinetics, and clinical use. *J Clin Pharmacol* 2005;**45**:1345–59.
4 Chang CC, Chu CF, Wang CN, Wu HT, Bi KW, Pang JH, et al. The anti-atherosclerotic effect of tanshinone IIA is associated with the inhibition of TNF-α-induced VCAM-1, ICAM-1 and CX3CL1 expression. *Phytomedicine* 2014;**21**:207–16.
5 Tao S, Zheng Y, Lau A, Jaramillo MC, Chau BT, Lantz RC, et al. Tanshinone I activates the Nrf2-dependent antioxidant response and protects against as (III)-induced lung inflammation *in vitro* and *in vivo*. *Antioxid Redox Signal* 2013;**19**:1647–61.
6 Zhang Y, Jiang P, Ye M, Kim SH, Jiang C, Lü J, et al. Tanshinones: sources, pharmacokinetics and anti-cancer activities. *Int J Mol Sci* 2012;**13**:13621–66.
7 Liu X, Guo CY, Ma XJ, Wu CF, Zhang Y, Sun MY, et al. Anti-inflammatory effects of tanshinone IIA on atherosclerostic vessels of ovariectomized ApoE mice are mediated by estrogen receptor activation and through the ERK signaling pathway. *Cell Physiol Biochem* 2015;**35**:1744–55.
8 Wang X, Morris-Natschke SL, Lee KH. New developments in the chemistry and biology of the bioactive constituents of Tanshen. *Med Res Rev* 2007;**27**:133–48.
9 Dong Y, Morris-Natschke SL, Lee KH. Biosynthesis, total syntheses, and antitumor activity of tanshinones and their analogs as potential therapeutic agents. *Nat Prod Rep* 2011;**28**:529–42.
10 Xu Z, Peters RJ, Weirather J, Luo H, Liao B, Zhang X, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J* 2015;**82**:951–61.
11 Xu ZC, Ji AJ, Zhang X, Song JY, Chen SL. Biosynthesis and regulation of active compounds in medicinal model plant *Salvia miltiorrhiza*. *Chin Herb Med* 2016;**8**:3–11.
12 Kai G, Liao P, Xu H, Wang J, Zhou C, Zhou W, et al. Molecular mechanism of elicitor-induced tanshinone accumulation in *Salvia miltiorrhiza*, hairy root cultures. *Acta Physiol Plant* 2012;**34**:1421–33.
13 Luo H, Zhu Y, Song J, Xu L, Sun C, Zhang X, et al. Transcriptional data mining of *Salvia miltiorrhiza* in response to methyl jasmonate to examine the mechanism of bioactive compound biosynthesis and regulation. *Physiol Plant* 2014;**152**:241–55.
14 Gao W, Sun HX, Xiao H, Cui G, Hillwig ML, Jackson A, et al. Combining metabolomics and transcriptomics to characterize tanshinone biosynthesis in *Salvia miltiorrhiza*. *BMC Genom* 2014;**15**:73.
15 Wang Y, Shen Y, Shen Z, Zhao L, Ning D, Jiang C, et al. Comparative proteomic analysis of the response to silver ions and yeast extract in *Salvia miltiorrhiza* hairy root cultures. *Plant Physiol Biochem* 2016;**107**:364–73.
16 Zhang L, Wu B, Zhao D, Li C, Shao F, Lu S, et al. Genome-wide analysis and molecular dissection of the *SPL* gene family in *Salvia miltiorrhiza*. *J Integr Plant Biol* 2014;**56**:38–50.
17 Li C, Lu S. Genome-wide characterization and comparative analysis of R2R3-MYB transcription factors shows the complexity of MYB-associated regulatory networks in *Salvia miltiorrhiza*. *BMC Genom* 2014;**15**:277.
18 Li C, Li D, Shao F, Lu S. Molecular cloning and expression analysis of WRKY transcription factor genes in *Salvia miltiorrhiza*. *BMC Genom* 2015;**16**:200.
19 Zhang X, Luo H, Xu Z, Zhu Y, Ji A, Song J, et al. Genome-wide characterisation and analysis of bHLH transcription factors related to tanshinone biosynthesis in *Salvia miltiorrhiza*. *Sci Rep* 2015;**5**:11244.
20 Ji A, Luo H, Xu Z, Zhang X, Zhu Y, Liao B, et al. Genome-wide identification of the *AP2/ERF* gene family involved in active constituent biosynthesis in *Salvia miltiorrhiza*. *Plant Genome* 2016;**9**:1–11.
21 Xu Z, Ji A, Song J, Chen S. Genome-wide analysis of auxin response factor gene family members in medicinal model plant *Salvia miltiorrhiza*. *Biol Open* 2016;**5**:848–57.
22 Zhou Y, Sun W, Chen J, Tan H, Xiao Y, Li Q, et al. SmMYC2a and SmMYC2b played similar but irreplaceable roles in regulating the biosynthesis of tanshinones and phenolic acids in *Salvia miltiorrhiza*. *Sci Rep* 2016;**6**:22852.
23 Hurst HC. Transcription factors 1: bZIP proteins. *Protein Profile* 1995;**2**:105–68.
24 Izawa T, Foster R, Chua NH. Plant bZIP protein DNA binding specificity. *J Mol Biol* 1993;**230**:1131–44.
25 Foster R, Izawa T, Chua NH. Plant bZIP proteins gather at ACGT elements. *FASEB J* 1994;**8**:192–200.
26 Sibéril Y, Benhamron S, Memelink J, Giglioli-Guivarc'h N, Thiersault M, Boisson B, et al. *Catharanthus roseus* G-box binding factors 1 and 2 act as repressors of strictosidine synthase gene expression in cell cultures. *Plant Mol Biol* 2001;**45**:477–88.
27 Espín FM, Peraza-Echeverria S, Fuentes G, Santamaría JM. *In silico* cloning and characterization of the TGA (TGACG motif-binding factor) transcription factors subfamily in *Carica papaya*. *Plant Physiol Biochem* 2012;**54**:113–22.
28 Akagi T, Katayama-Ikegami A, Kobayashi S, Sato A, Kono A, Yonemori K. Seasonal abscisic acid signal and a basic leucine zipper transcription factor, DkbZIP5, regulate proanthocyanidin biosynthesis in persimmon fruit. *Plant Physiol* 2012;**158**:1089–102.
29 Zhang Y, Zheng S, Liu Z, Wang L, Bi Y. Both HY5 and HYH are necessary regulators for low temperature-induced anthocyanin accumulation in *Arabidopsis* seedlings. *J Plant Physiol* 2011;**168**:367–74.
30 Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, et al. bZIP transcription factors in *Arabidopsis*. *Trends Plant Sci* 2002;**7**:106–11.
31 Liao Y, Zou HF, Wei W, Hao YJ, Tian AG, Huang J, et al. Soybean *GmbZIP44*, *GmbZIP62* and *GmbZIP78* genes function as negative regulator of ABA signaling and confer salt and freezing tolerance in transgenic *Arabidopsis*. *Planta* 2008;**228**:225–40.
32 Nijhawan A, Jain M, Tyagi AK, Khurana JP. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiol* 2008;**146**:333–50.
33 Wang J, Zhou J, Zhang B, Vanitha J, Ramachandran S, Jiang SY. Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher Plants with an emphasis on sorghum. *J Integr Plant Biol* 2011;**53**:212–31.
34 Jin Z, Xu W, Liu A. Genomic surveys and expression analysis of bZIP gene family in castor bean (*Ricinus communis* L.). *Planta* 2014;**239**:299–312.
35 Liu J, Chen N, Chen F, Cai B, Dal Santo S, Tornielli GB, et al. Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (*Vitis vinifera*). *BMC Genom* 2014;**15**:281.
36 Zhao J, Guo R, Guo C, Ho H, Wang X, Gao H. Evolutionary and expression analyses of the apple basic leucine zipper transcription factor family. *Front Plant Sci* 2016;**7**:376.
37 Xu H, Song J, Luo H, Zhang Y, Li Q, Zhu Y, et al. Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol Plant* 2016;**9**:949–52.
38 Xing B, Yang D, Guo W, Liang Z, Yan X, Zhu Y, et al. Ag^+ as a more effective elicitor for production of tanshinones than phenolic acids in *salvia miltiorrhiza* hairy roots. *Molecules* 2014;**20**:309–24.
39 Xu Z, Luo H, Ji A, Zhang X, Song J, Chen S, et al. Global identification of the full-length transcripts and alternative splicing related to phenolic acid biosynthetic genes in *Salvia miltiorrhiza*. *Front Plant Sci* 2016;**7**:100.
40 Lv H, Zhang X, Liao B, Liu W, He L, Song J, et al. Cloning and analysis of 1-hydroxy-2-methyl-2-(*E*)-butenyl-4-diphosphate reductase genes *HsHDR1* and *HsHDR2* in *Huperzia serrate*. *Acta Pharm Sin B* 2015;**5**:583–9.
41 Wang Z, Cheng K, Wan L, Yan L, Jiang H, Liu S, et al. Genome-wide analysis of the basic leucine zipper (bZIP) transcription factor gene family in six legume genomes. *BMC Genom* 2015;**16**:1053.
42 Wei K, Chen J, Wang Y, Chen Y, Chen S, Lin Y, et al. Genome-wide analysis of bZIP-encoding genes in maize. *DNA Res* 2012;**19**:463–76.

43 Tu D, Ma X, Zhao H, Mo C, Tang Q, Wang L, et al. Cloning and expression of *SgCYP450-4* from *Siraitia grosvenorii*. *Acta Pharm Sin B* 2016;**6**:614–22.
44 Baloglu MC, Eldem V, Hajyzadeh M, Unver T. Genome-wide analysis of the bZIP transcription factors in cucumber. *PLoS One* 2014;**9**: e96014.
45 Chen S, Song J, Sun C, Xu J, Zhu Y, Verpoorte R, et al. Herbal genomics: examining the biology of traditional medicines. *Science* 2015;**347**:S27–9.
46 Li D, Fu F, Zhang H, Song F. Genome-wide systematic characterization of the bZIP transcriptional factor family in tomato (*Solanum lycopersicum* L.). *BMC Genom* 2015;**16**:771–90.
47 Pourabed E, Golmohamadi FG, Monfared PS, Razavi SM, Shobbar ZS. Basic leucine zipper family in barley: genome-wide characterization of members and expression analysis. *Mol Biotechnol* 2015;**57**:12–26.
48 Liu X, Chu Z. Genome-wide evolutionary characterization and analysis of bZIP transcription factors and their expression profiles in response to multiple abiotic stresses in *Brachypodium distachyon*. *BMC Genom* 2015;**16**:227.
49 Zhang F, Fu X, Lv Z, Lu X, Shen Q, Zhang L, et al. A basic leucine zipper transcription factor, AabZIP1, connects abscisic acid signaling with artemisinin biosynthesis in *Artemisia annua*. *Mol Plant* 2015;**8**:163–75.
50 Fricke J, Hillebrand A, Twyman RM, Prüfer D, Gronover CS. Abscisic acid-dependent regulation of small rubber particle protein gene expression in *Taraxacum brevicorniculatum* is mediated by TbbZIP1. *Plant Cell Physiol* 2013;**54**:448–64.
51 Okada A, Okada K, Miyamoto K, Koga J, Shibuya N, Nojiri H, et al. OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. *J Biol Chem* 2009;**284**:26510–8.
52 Miyamoto K, Nishizawa Y, Minami E, Nojiri H, Yamane H, Okada K. Overexpression of the bZIP transcription factor OsbZIP79 suppresses the production of diterpenoid phytoalexin in rice cells. *J Plant Physiol* 2015;**173**:19–27.