RESEARCH ARTICLE

Genome-wide analysis of auxin response factor gene family members in medicinal model plant Salvia miltiorrhiza

Zhichao Xu1,2, Aijia Ji1,2, Jingyuan Song1,2,* and Shilin Chen1,2,3,*

ABSTRACT

Auxin response factors (ARFs) can function as transcriptional activators or repressors to regulate the expression of auxin response genes by specifically binding to auxin response elements (AuxREs) during plant development. Based on a genome-wide strategy using the medicinal model plant Salvia miltiorrhiza, 25 S. miltiorrhiza ARF (SmARF) gene family members in four classes (class Ia, Ila, IIb and III) were comprehensively analyzed to identify characteristics including gene structures, conserved domains, phylogenetic relationships and expression patterns. In a hybrid analysis of the phylogenetic tree, microRNA targets, and expression patterns of SmARFs in different organs, root tissues, and methyl jasmonate or indole-3-acetic acid treatment conditions, we screened for candidate SmARFs involved in various developmental processes of S. miltiorrhiza. Based on this analysis, we predicted that SmARF25, SmARF7, SmARF16 and SmARF20 are involved in flower, leaf, stem and root development, respectively. With the further insight into the targets of miR160 and miR167, specific SmARF genes in S. miltiorrhiza might encode products that participate in biological processes as described for ARF genes in Arabidopsis. Our results provide a foundation for understanding the molecular basis and regulatory mechanisms of SmARFs in S. miltiorrhiza.

KEY WORDS: Developmental processes, Auxin response factors, Auxin response elements, MicroRNA, Salvia miltiorrhiza

INTRODUCTION

The phytohormone auxin, typified by indole-3-acetic acid (IAA), plays a crucial role in controlling the mechanisms by which plants grow and develop, including tropic responses, apical dominance, lateral root formation, vascular differentiation, flower and fruit development, and shoot elongation (Santer and Estelle, 2009). Auxin response factors (ARFs) are important transcription factors that can either activate or repress the transcriptional level of early/primary auxin response genes, such as Aux/IAA, Small Auxin Up RNA (SAUR) and Gretchlen Hagen 3 (GH3) gene family members, by binding to auxin response elements (AuxREs, TGTCTC) or some variation of these elements (TGTCCC or TGTAC) in their promoters (Hagen and Guilfoyle, 2002; Liu et al., 1994; Ulmasov et al., 1997, 1995, 1999b). AtARF1, which binds to the sequence TGTCTC in AuxREs, was the first cloned auxin-related transcription factor and was identified in Arabidopsis using a yeast one-hybrid system (Ulmason et al., 1997). Recently, microarray experiments indicated that AtARF1 and AtARF5 monomers specificity prefer TGTCCG elements to the AuxRE TGTCTC (Boer et al., 2014). The complete genomic sequence of Arabidopsis provides the opportunity to identify the sequence and evolution of all members of a given gene family (Arabidopsis Genome Initiative, 2000). Genome-wide analysis identified 22 full-length ARF genes and one partial-length gene (AtARF23) containing a stop codon in its DNA-binding domain (DBD) in Arabidopsis thaliana (Okushima et al., 2005b; Remington et al., 2004). Furthermore, biochemical and genetic approaches have established crucial functions of ARF genes in the growth and development of Arabidopsis (Guilfoyle and Hagen, 2007).

Taking advantage of the genome-wide identification of A. thaliana ARFs (AtARFs), many studies have found that the ARFs AtARF1 and AtARF2 function as transcriptional repressors related to the regulation of leaf senescence, floral organ abscission and cell growth (Ellis et al., 2005; Li et al., 2004; Okushima et al., 2005a; Schruff et al., 2006); AtARF3 and AtARF4 function in developing reproductive and vegetative tissues (Pekker et al., 2005; Sessions et al., 1997; Finet et al., 2010); AtARF5 functions in Arabidopsis leaf vascular and embryo patterning (Hamann et al., 2002; Krogan et al., 2012); AtARF6 and AtARF8 function in female and male reproduction (Nagpal et al., 2010); AtARF9 acts in suspensor cells to mediate hypophysis specification (Rademacher et al., 2012); and AtARF10, AtARF16 and AtARF17 function in the negative regulation of seed germination and post-germination activities (Liu et al., 2007, 2010). In Arabidopsis, certain AtARF expression patterns are controlled by miRNAs to regulate several developmental events. AtARF6 and AtARF8 are targets of miR167 (Wu et al., 2006), and AtARF10, AtARF16 and AtARF17 are targets of miR160 (Liu et al., 2007, 2010). In some cases, ARF gene expression is altered in response to exogenous auxin signals (Okushima et al., 2005b; Wang et al., 2007a).

A typical ARF contains three conserved domains: an N-terminal B3 DNA binding domain (DBD), a middle regional auxin response factor (MR), and a C-terminal PB1 protein-protein interaction domain (PB1). The DBD can recognize AuxREs or variation elements in the promoter of auxin-responsive genes (Wright and Nemhauser, 2015; Boer et al., 2014), and the PB1 domains are also found in Aux/IAAs (Guilfoyle and Hagen, 2012). Structural and biochemical studies have determined that the PB1 domains of ARFs and Aux/IAAs from AtARF7 (Korosick et al., 2014) and AtARF5
(Nanao et al., 2014) are involved in protein-protein interactions by forming higher order oligomerization or multimerization (Wright and Nemhauser, 2015). The MR, located between the DBD and the PB1 domain, confers functions such as transcriptional activation or repression depending on its amino acid composition (Mun et al., 2012; Yu et al., 2014). Previous studies have shown that glutamine (Q)-rich MRs function as activation regions but that serine (S)-rich, serine and proline (SP)-rich, and serine and glycine (SG)-rich MRs function as repression regions in ARFs from *A. thaliana* (Tiwar et al., 2003; Ulmasov et al., 1999a).

Given the complete genomic sequences of many important species, there has been significant progress in the analysis and identification of the functions of ARFs. Genome-wide analysis has identified many ARFs in many other important plants, such as 25 *Oryza sativa* ARF (OsARF) loci (Wang et al., 2007a), 22 *Solanium lycopersicum* ARFs (SiARFs) (Zouine et al., 2014), 31 *Brassica rapa* (BrARFs) (Mun et al., 2012), 19 *Vitis vinifera* ARFs (VvARFs) (Wan et al., 2014), 47 *Musa acuminata* ARFs (MaARFs) (Hu et al., 2015), 17 *Eucalyptus grandis* (EgARFs) (Yu et al., 2014), 24 *Medicago truncatula* ARFs (MtARFs) (Shen et al., 2015), 39 *Populus trichocarpa* ARFs (PtARFs) (Kalluri et al., 2007), 19 *Citrus sinensis* ARFs (CiARFs) (Li et al., 2015b), 11 *Carica papaya* ARFs (CpARFs) (Liu et al., 2015), and 35 *Gossypium raimondii* ARFs (GrARFs) (Sun et al., 2015). However, the ARF transcription factor family members have not been determined in *Salvia miltiorrhiza*, one of the most commonly used herbs in traditional Chinese medicine (TCM). *S. miltiorrhiza*, also referred to as danshen, belongs to the *Salvia* genus of the Lamiaceae family, and its dried root and rhizome are highly valued (Cheng, 2006). Danshen is well known for its use alone or in combination with other herbs in the treatment of cardiovascular diseases, as well as for its anti-inflammatory, immunomodulatory and anti-oxidative activities; the primary bioactive compounds in danshen are lipophilic diterpenoids and hydrophilic phenolic acids (Wang et al., 2007b; Dong et al., 2011). S. *miltiorrhiza* is also considered a good medicinal plant model in TCM research for studying the biosynthesis and regulation of active compounds (Ma et al., 2012; Xu et al., 2015). Due to the establishment of the *S. miltiorrhiza* genome sequence (Xu et al., 2016a) it has become feasible use in *silico* analysis to isolate its functional gene families such as diterpene; phenolic acid biosynthetic genes; and bHLH, AP2/ERF, WRKY, MYB and SPL transcription factors (Ji et al., 2015; Li et al., 2015a; Li and Lu, 2014; Ma et al., 2012; Wang et al., 2015; Zhang et al., 2014, 2015; Xu et al., 2016b). As ARF gene members are key factors in plant growth and development, identifying these genes in *S. miltiorrhiza* aid in the understanding of developmental processes and cellular responses to auxin in danshen.

Here, we isolated 25 *S. miltiorrhiza* ARF (*SmARF*) genes using a genome-wide approach. Following complete genome sequencing the sequence homology of these *SmARFs* and their gene expression patterns in different organs, root tissues, and methyl jasmonate (MeJA) or IAA treatment conditions, gene structures, and the phylogenetic relationships between *SmARFs* and *AtARFs* were analyzed in detail. This study provided molecular information regarding the *SmARF* gene family and the results will aid in selecting candidate genes related to cell growth and tissue development in *S. miltiorrhiza*, paving the way for further functional characterization of these *SmARF* genes.

RESULTS
Identification and phylogenetic analysis of danshen ARFs
After a BLASTP search and protein domain analysis, 25 non-redundant ARF genes were identified from the genome sequences of *S. miltiorrhiza*. These *SmARFs*, located in the different scaffolds, were named *SmARF*1-*SmARF*25 according to the order of their annotated gene IDs, listed in Table 1. The number of ARF genes in *S. miltiorrhiza* is similar to the number in *A. thaliana* (23), *O. sativa* (25) and *M. acuminata* (24). The predicted proteins encoded by *SmARF* genes varied from 345 amino acids (*SmARF*12) to 1105 amino acids (*SmARF*22), with corresponding molecular weights from 37.78 kDa to 122.17 kDa, and the theoretical isoelectric points ranged from 5.29 (*SmARF*5) to 9.28 (*SmARF*12). Pair-wise analysis of *SmARF* protein homology indicated that the overall homology broadly ranged from 22% (between *SmARF*6 and *SmARF*16) to 89% (between *SmARF*5 and *SmARF*7). The *SmARF*9 and *SmARF*10 genes are located in the same scaffold1069, and the other *SmARFs* are distributed in different scaffolds. Most of the *SmARFs* were predicted to localize to the nucleus, however *SmARF*1 and *SmARF*12 were predicted to localize to chloroplasts.

To characterize the evolutionary relationship between danshen ARF proteins and *Arabidopsis* ARFs, a neighbor-joining tree was constructed using the full-length amino acid sequences (Fig. 1). The results indicated that 25 *SmARFs* were classed together with 23 *AtARFs* into four clusters (classes Ia, Iib, IIIb and IIIc) according to the well-supported bootstrap data. In *S. miltiorrhiza*, *SmARF*3, 5, 7, 18, 21 and 23 belong to class Ia; *SmARF*2, 6, 8, 9, 10, 19, 20, 22, 24 and 25 belong to the largest class Iia; *SmARF*4, 11 and 17 belong to class Ib; and *SmARF*1, 12, 13, 14, 15 and 16 belong to class III. In *A. thaliana*, there is another class Ib that includes *AtARF*12-15 and 20-23. Notably, no *S. miltiorrhiza* ARF proteins were clustered into class Ib from the phylogenetic tree, and this observation implies a diverging trend in the evolution of ARF genes across different plants.

To investigate the biological processes of *SmARFs*, gene ontology (GO) mapping and annotation were performed using Blast2GO. The functional categorization of *SmARFs* as annotated by GO analysis, including their biological processes, molecular functions, and cellular components, is presented in Table S1. Regarding biological processes, eight categories met the criterion of NodeScore >2.0: cellular process (25 genes), metabolic process (25 genes), response to stimulus (25 genes), single-organism process (14 genes), biological regulation (13 genes), signaling (11 genes), developmental process (five genes), and multicellular organismal process (five genes). For developmental process, *SmARF*6 was related to embryo development; five *SmARFs* were classified into post-embryonic developmental process (*SmARF*1, 4, 6, 11, 25); and *SmARF*1, 4, 6 and 25 were predicted to participate in flower development. Notably, the biological process related to secondary metabolism was not identified. Based on the molecular function analysis, all the *SmARFs* were classified into DNA binding; 21 *SmARFs* were grouped into protein binding; and three *SmARFs* were categorized into sequence-specific DNA binding transcription factor activity (*SmARF*1, 11 and 25). According to the cellular component annotation, all *SmARFs* except for *SmARF*1 and *SmARF*12 were localized to the nucleus, in accordance with the subcellular localization predictions.

**Gene structures and conserved domains of danshen ARFs**
To better understand the gene structure of *SmARFs*, the exon-intron features among *SmARFs* were aligned via phylogenetic analysis (Fig. 2). The phylogenetic analysis revealed four clusters in accordance with the group data presented in Fig. 1. Gene structure analysis of all of the *SmARF* genes revealed that the number of exons ranges from 1 to 18, however, *SmARF*12 is intronless. The
<table>
<thead>
<tr>
<th>Gene_namea</th>
<th>Gene_IDb</th>
<th>Scaffoldc</th>
<th>GenBankd</th>
<th>ORF length(bp)e</th>
<th>Polypeptided</th>
<th>Polypeptide length (aa)</th>
<th>MW (kDa)</th>
<th>pl</th>
<th>Domaind</th>
<th>Homologousd</th>
<th>Locationd</th>
</tr>
</thead>
<tbody>
<tr>
<td>SmARF1</td>
<td>SMil_0000332</td>
<td>scaffold7</td>
<td>KX036721</td>
<td>2097</td>
<td>688</td>
<td>78.99</td>
<td>6.8</td>
<td>DBD, MR, PB1</td>
<td>AARF16</td>
<td>Chloroplast</td>
<td></td>
</tr>
<tr>
<td>SmARF2</td>
<td>SMil_0001334</td>
<td>scaffold195</td>
<td>KX036722</td>
<td>3249</td>
<td>1082</td>
<td>119.46</td>
<td>6.12</td>
<td>DBD, MR, PB1</td>
<td>AARF19/ATARF7</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF3</td>
<td>SMil_0001680</td>
<td>scaffold75</td>
<td>KX036723</td>
<td>1950</td>
<td>649</td>
<td>72.02</td>
<td>5.83</td>
<td>DBD, MR, PB1</td>
<td>AARF9</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF4</td>
<td>SMil_0002054</td>
<td>scaffold78</td>
<td>KX036724</td>
<td>2058</td>
<td>685</td>
<td>75.60</td>
<td>7.14</td>
<td>DBD, MR</td>
<td>AARF3</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF5</td>
<td>SMil_0002082</td>
<td>scaffold79</td>
<td>KX036725</td>
<td>2274</td>
<td>757</td>
<td>84.52</td>
<td>5.29</td>
<td>DBD, MR, PB1</td>
<td>AARF2</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF6</td>
<td>SMil_0003923</td>
<td>scaffold191</td>
<td>KX036726</td>
<td>3162</td>
<td>1053</td>
<td>116.63</td>
<td>5.67</td>
<td>DBD, MR, PB1</td>
<td>AARF5</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF7</td>
<td>SMil_0007048</td>
<td>scaffold505</td>
<td>KX036727</td>
<td>2508</td>
<td>835</td>
<td>93.45</td>
<td>6.05</td>
<td>DBD, MR, PB1</td>
<td>AARF2</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF8</td>
<td>SMil_0009504</td>
<td>scaffold820</td>
<td>KX036728</td>
<td>2253</td>
<td>750</td>
<td>82.51</td>
<td>6.1</td>
<td>DBD, MR, PB1</td>
<td>AARF6</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF9</td>
<td>SMil_0011003</td>
<td>scaffold1069</td>
<td>KX036729</td>
<td>2514</td>
<td>837</td>
<td>92.91</td>
<td>5.49</td>
<td>DBD, MR, PB1</td>
<td>AARF8</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF10</td>
<td>SMil_0011004</td>
<td>scaffold1069</td>
<td>KX036730</td>
<td>2499</td>
<td>832</td>
<td>92.79</td>
<td>5.73</td>
<td>DBD, MR, PB1</td>
<td>AARF8</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF11</td>
<td>SMil_0011552</td>
<td>scaffold1157</td>
<td>KX036731</td>
<td>2445</td>
<td>814</td>
<td>89.70</td>
<td>6.08</td>
<td>DBD, MR, PB1</td>
<td>AARF4</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF12</td>
<td>SMil_0011716</td>
<td>scaffold1183</td>
<td>KX036732</td>
<td>1038</td>
<td>345</td>
<td>37.78</td>
<td>9.28</td>
<td>DBD, MR, PB1</td>
<td>AARF17</td>
<td>Chloroplast</td>
<td></td>
</tr>
<tr>
<td>SmARF13</td>
<td>SMil_0011783</td>
<td>scaffold1197</td>
<td>KX036733</td>
<td>2046</td>
<td>681</td>
<td>75.57</td>
<td>7.6</td>
<td>DBD, MR, PB1</td>
<td>AARF16</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF14</td>
<td>SMil_0011888</td>
<td>scaffold1217</td>
<td>KX036734</td>
<td>1992</td>
<td>663</td>
<td>73.45</td>
<td>6.36</td>
<td>DBD, MR, PB1</td>
<td>AARF6</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF15</td>
<td>SMil_0012294</td>
<td>scaffold3445</td>
<td>KX036735</td>
<td>2046</td>
<td>681</td>
<td>74.39</td>
<td>7.88</td>
<td>DBD, MR, PB1</td>
<td>AARF16</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF16</td>
<td>SMil_0014125</td>
<td>scaffold1650</td>
<td>KX036736</td>
<td>1965</td>
<td>654</td>
<td>72.00</td>
<td>6.11</td>
<td>DBD, MR, PB1</td>
<td>AARF6</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF17</td>
<td>SMil_0015390</td>
<td>scaffold1939</td>
<td>KX036737</td>
<td>1422</td>
<td>473</td>
<td>51.98</td>
<td>6.7</td>
<td>DBD, MR, PB1</td>
<td>AARF4</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF18</td>
<td>SMil_0016249</td>
<td>scaffold2256</td>
<td>KX036738</td>
<td>2079</td>
<td>692</td>
<td>77.19</td>
<td>6.65</td>
<td>DBD, MR, PB1</td>
<td>AARF18</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF19</td>
<td>SMil_0017439</td>
<td>scaffold2499</td>
<td>KX036739</td>
<td>2013</td>
<td>670</td>
<td>74.35</td>
<td>5.82</td>
<td>DBD, MR, PB1</td>
<td>AARF8</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF20</td>
<td>SMil_0018692</td>
<td>scaffold2890</td>
<td>KX036740</td>
<td>2361</td>
<td>786</td>
<td>87.33</td>
<td>5.77</td>
<td>DBD, MR, PB1</td>
<td>AARF19/ATARF7</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF21</td>
<td>SMil_0020421</td>
<td>scaffold3557</td>
<td>KX036741</td>
<td>1752</td>
<td>583</td>
<td>64.88</td>
<td>5.66</td>
<td>DBD, MR, PB1</td>
<td>AARF11</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF22</td>
<td>SMil_0020906</td>
<td>scaffold3760</td>
<td>KX036742</td>
<td>3318</td>
<td>1105</td>
<td>122.17</td>
<td>6.14</td>
<td>DBD, MR, PB1</td>
<td>AARF19/ATARF7</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF23</td>
<td>SMil_0022033</td>
<td>scaffold4292</td>
<td>KX036743</td>
<td>2322</td>
<td>773</td>
<td>86.88</td>
<td>5.87</td>
<td>DBD, MR, PB1</td>
<td>AARF1</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF24</td>
<td>SMil_0022205</td>
<td>scaffold4381</td>
<td>KX036744</td>
<td>2493</td>
<td>830</td>
<td>91.72</td>
<td>6.19</td>
<td>DBD, MR, PB1</td>
<td>AARF6</td>
<td>Nuclear</td>
<td></td>
</tr>
<tr>
<td>SmARF25</td>
<td>SMil_0026186</td>
<td>scaffold7111</td>
<td>KX036745</td>
<td>2625</td>
<td>874</td>
<td>96.29</td>
<td>5.86</td>
<td>DBD, MR, PB1</td>
<td>AARF6</td>
<td>Nuclear</td>
<td></td>
</tr>
</tbody>
</table>

*Names of SmARF genes in *S. miltiorrhiza*.  
*Gene ID annotated in *S. miltiorrhiza* genome.  
*Scaffold numbers assembled in *S. miltiorrhiza* genome.  
The GenBank accession numbers.  
*Length of open reading frame in base pairs.  
The number of amino acids, theoretical pl (isoelectric point) and MW (molecular weight) of polypeptide.  
*DBD: B3 DNA-binding domain; MR: middle regional auxin response factor domain; PB1: C-terminal dimerization domain.  
*Homology: the homology with Arabidopsis ARFs.  
*Location: predicted subcellular location of SmARFs.

**Prediction of miRNA targets among SmARFs and analysis of the AuxREs in SmARF gene promoters**

Using the BLASTN algorithm to identify targets of miRNA160 and miRNA167 within SmARF gene sequences, target sites of *At-miRNA160* (UGCCUGGCUCCCGUGUA) and *miRNA167* were predicted within the 1300-1319 bp region of *SmARF1*, the 1359-1379 bp region of *SmARF13*, the 1438-1367 bp region of *SmARF14*, the 1332-1352 bp region of *SmARF15* and the 1363-1376 bp region of *SmARF16*. Additionally, target sites of *miRNA167* (UCAAGCUUGCGCUAGUACUA) were predicted within the 1975-1993 bp region of *SmARF8*, the 2302-2320 bp region of *SmARF9*, the 2287-2305 bp region of *SmARF10*, the 1777-1795 bp region of *SmARF19*, the 2419-2437 bp region of *SmARF24* and the 2350-2368 bp region of *SmARF25* (Fig. 2). The results of this analysis suggested that miR160-167-mediated post-transcriptional regulation of ARFs is conserved between *S. miltiorrhiza* and *Arabidopsis*.
We surveyed 20 AUX/IAA and 10 GH3 primary/early auxin response gene members in *S. miltiorrhiza* based on a genome-wide strategy. The promoters (−1000 to −1 bp) of these two auxin response gene families were selected to screen for AuxREs. As expected, 19 of 20 AUX/IAA and 9 of 10 GH3 gene promoters contain one or more AuxREs (Table S3). These results indicated that these auxin response genes could be regulated by SmARFs in *S. miltiorrhiza*.

**Expression patterns of SmARF genes in different plant organs or tissues**

To better probe the physiological function of SmARFs, the tissue-specific expression of 25 SmARF genes in different danshen organs (leaf, root, stem and flower) was determined by analyzing the RNA-seq data (Fig. 3; Table S4). Most SmARF genes, but not SmARF9, 12 or 17 presented ubiquitous expression and high variability in all studied organs, and this result implies that these SmARFs might function in danshen growth and development. There were significant differences in SmARF expression between organs. SmARF3, 7 and 21, all of which belong to class Ia based on phylogenetic analysis, showed higher expression in danshen leaf than in other organs. SmARF4, 5, 11, 18, 20 and 23 were expressed more strongly in danshen root than in other organs, however, only SmARF16 showed stem-specific expression in *S. miltiorrhiza*. When comparing phylogenetic tree analysis with the expression cluster analysis, SmARF8, 10, 19, 22, 24 and 25, which belong to class IIa, showed significantly lower expression in danshen leaf than in other organs. Most of the SmARF genes from class III (SmARF13-17) also clustered in one expression branch. These results indicated that ARF genes from the same class might perform a similar physiological function in plants.

Previous evidence revealed that the periderm of danshen root is the primary site of biosynthesis and accumulation of tanshinones. The expression pattern of SmARF genes in different root tissues (periderm, phloem and xylem) was also examined using RNA-seq data (Fig. S4, Table S4). SmARF9, 12 and 17 displayed no expression in danshen root tissues. SmARF13 showed the greatest expression in periderm, more than four and 14 times greater than that in phloem and xylem, respectively.
expression in phloem and xylem than in periderm. Furthermore, SmARF20 exhibited stronger expression in phloem and xylem than in periderm.

**Expression patterns of SmARF genes upon auxin or MeJA treatment**

Auxin is a central regulator of plant growth and development. To investigate the response of SmARF genes to exogenous IAA stimulation, we analyzed the variation in SmARF gene expression at 0, 0.5, 1 and 3 h after 20 μM IAA treatment using qRT-PCR (Fig. 4). As expected, most SmARF genes were significantly auxin-sensitive. The overall expression patterns of SmARFs varied, with 11 SmARF mRNAs (SmARF2, 3, 4, 5, 11, 13, 15, 18, 21, 22 and 23) showing up-regulation and eight SmARF mRNAs (SmARF1, 7, 8, 10, 14, 19, 24 and 25) showing down-regulation at 3 h of IAA treatment (P<0.01 for all). One SmARF gene (SmARF20) did not display significant changes in expression (P>0.05) regardless of the treatment duration. The unmentioned SmARF6 and SmARF16 displayed significantly down-regulated expression at 0.5 and 1 h, and at 3 h, the expression of these genes returned to the same level as that for mock IAA treatment. The most strongly up-regulated SmARF genes, SmARF13, 15 and 23, were markedly induced after IAA treatment [greater than twofold increase, log (expression level) >1]. Similarly, the expression of five SmARF genes (SmARF1, 10, 16, 19, 25) showed marked down-regulation [greater than twofold decrease, log (expression level) >1]. For IAA treatment, 13 SmARF genes (SmARF1, 3, 5, 6, 7, 8, 11, 16, 18, 21, 22, 24 and 25) displayed significant up- or downregulation over the three examined time points. For example, the expression level of SmARF16 was decreased by greater than twofold at 1 h but was significantly increased at 3 h compared with the control levels.

In *Nicotiana benthamiana*, transient silence of *NbARF1* and MeJA treatment resulted in significant enrichment of leaf nicotine (Todd et al., 2010). In addition, MeJA treatment significantly alters the biosynthesis of active compounds (tanshinones or phenolic acids) in *S. miltiorrhiza*, hence the expression variation of SmARF genes after MeJA treatment was studied using RNA-seq data (Table S4). The results showed that SmARF24 and 25 displayed significant up-regulation and that SmARF1 exhibited downregulation after MeJA treatment. The expression of other SmARF genes showed no evident changes following MeJA treatment. This finding suggests that SmARFs might perform a small role in post-developmental processes.

**DISCUSSION**

The cultivation of medicinal plants has faced intense pressure due to social and environmental concerns. Studying the molecular mechanisms of medicinal plant growth processes would help resolve potential questions related to cultivation of these plants. Genome-wide characterization and analysis of SmARFs could improve the understanding of their regulatory roles in danshen growth and development. In this study, 25 ARF gene members in *S. miltiorrhiza* were identified, and this number was similar to that for other model plants, such as *A. thaliana* (23) and *O. sativa* (25). Protein domain analysis provided useful information for predicting the biological functions of SmARFs, which primarily depend on their characteristic DBD, MR and PB1. ARFs rely on the DBD to specifically bind to AuxREs in the promoters of auxin-responsive genes. Their PB1 is involved in homomeric and heteromeric interactions with ARFs and Aux/IAA proteins. The percentage of PB1-truncated SmARFs (8%) was much lower than that of the ARF members identified in other plants, such as *Arabidopsis* (17%), rice (24%) and *M. truncatula* (54%). ARFs can function as transcriptional activators or repressors according to the amino acid composition of the MR. The activator/repressor ratio of SmARFs was 0.39 (7/18) and this value was also much lower than the ratios for *Arabidopsis* (0.59) and rice (0.56). Phylogenetic analysis and divergence time estimation based on 1824 single-copy true orthologous genes indicated that *S. miltiorrhiza* was distantly related to *Arabidopsis*, with an estimated divergence time of approximately 139 million years ago (Xu et al., 2016a). We also constructed a
phylogenetic tree to analyze the relationship of ARF family members between *S. miltiorrhiza* and *Arabidopsis* (Fig. 1). Phylogenetic tree analysis revealed five sister gene pairs with high bootstrap values (≥98%) between *S. miltiorrhiza* and *Arabidopsis*; this evidence supports the high homology of ARFs between species. Although there are similar numbers of ARFs between the *S. miltiorrhiza* and *A. thaliana* genomes, the absence of class Ib ARFs from *S. miltiorrhiza* reflects genomic expansion and rearrangements resulting from extensive duplication and deletion over a long period of evolutionary history according to the phylogenetic tree analysis. The detection of close relationships based on comparative analysis may help in the selection of candidate ARFs with specific biological functions in *S. miltiorrhiza*. According to the motif analysis, the motifs from different classes in SmARFs present high conservation (Fig. S2). Motif 8 and motif 7, located in the PB1 domain of the C-terminal in most SmARFs, include conserved residues (lysine motif and OPCA-like motif) of the positive and negative face found only in the ARF family, thus indicating the evolutionary conservation of ARF function (Korasick et al., 2014; Nanao et al., 2014). Much evidence demonstrates that *miRNAs* play dominant roles in post-transcriptional gene regulation by binding to their complementary mRNA targets, especially to target transcription factors, in plants (Jones-Rhoades et al., 2006; Li and Zhang, 2016). In *Arabidopsis*, *miR167* controls the expression patterns of *AtARF6* and *AtARF8* to regulate female and male reproduction or to promote jasmonic acid production and flower maturation (Nagpal et al., 2005; Wu et al., 2006). Phylogenetic tree analysis showed that SmARF8, 9, 10, 19, 24 and 25 were closely related to *AtARF6* and *AtARF8*, both of which are in class IIa; all of these six SmARFs contain a target site of *miR167* (Fig. 2). GO analysis also categorized SmARF25 into flower development; thus in *S. miltiorrhiza* we predicted that the expression of SmARF8, 10, 19, 24 and 25 might be inhibited by *miR167* to regulate certain developmental processes as described for *AtARF6* and *AtARF8*. Among them, SmARF25 was identified as the best candidate regulator of flower development. In addition, SmARF10, 19, 24 and 25 might function as transcriptional activators due to their Q-rich MRs. Additionally, *miR160* was found to bind to *AtARFs* (*AtARF10, AtARF16* and *AtARF17*) to negatively regulate seed germination and post-germination activities (Liu et al., 2007, 2010). These *AtARFs* were closely related to SmARF1, 12, 13, 14, 15 and 16 in class III, and these ARFs might function as transcriptional repressors due to the amino acid compositions of their MRs. Aside from SmARF12, other class III *SmARF* genes were identified to

![Color Key](image)

**Fig. 3. A heat map showing SmARF gene expression patterns in different organs.** The red color represents upregulation of expression, the white color represents an unchanged expression level, and the blue color represents downregulation of expression. Red dot (SmARF6) is not belong to the Class III.
contain miR160 target sites; this finding implies that these SmARFs perform functions that are similar to the functions of AtARF10, 16 and 17.

Comprehensive analysis of SmARF gene expression patterns and the evolution of their sequences helped us screen for candidate SmARF genes with potentially distinct functions. Most SmARF genes displayed ubiquitous but highly variable expression in all studied organs, and this expression pattern suggests their functional divergence. In Arabidopsis, AtARF2 regulates leaf senescence and floral organ abscission independently of the ethylene and cytokinin response pathways (Ellis et al., 2005). In S. miltiorrhiza the expression levels of SmARF3, 7 and 21 were significantly higher in leaves than in other studied organs, and SmARF5 and SmARF7 were closely related to AtARF2 in class Ia; these findings indicate that SmARF7 might play a crucial role in leaf development. In Arabidopsis AtARF7 and AtARF19 promote leaf expansion and auxin-induced lateral root formation (Wilmoth et al., 2005). In S. miltiorrhiza, SmARF20 was grouped with AtARF7 and 19 in class IIa, and the expression of SmARF20 was much higher in danshen root than in other organs. These observations suggest that SmARF20 likely regulates auxin-induced lateral root formation. The differential expression of SmARF20 between periderm, phloem, and xylem further support its role in root development. Notably, SmARF8, 10, 19, 22, 24 and 25 also belong to class IIA with AtARF7 and 19. The expression patterns of these SmARFs were much lower in danshen leaf, reflecting that they might be involved in leaf expansion. SmARF16, a stem-specifically expressed transcription factor, likely participates in stem development.

Recently, synthetic biology, particularly the biosynthesis of natural products, has advanced by leaps and bounds. The tanshinone and phenolic acid biosynthetic pathways, which have gradually been elucidated, have attracted increasing attention (Cui et al., 2015; Guo et al., 2013; Ma et al., 2012; Xu et al., 2015); however the molecular mechanism of danshen development has been an unpopular subject despite the importance of this medicinal plant. In this study the basic functional characteristics of SmARFs, such as the presence of the conserved DBD, MR and PB1 in 88% (22/25) of the SmARFs; the significant variation in the expression of 95% (21/22) of the examined SmARFs after 0.5 h, 1 h and 3 h of IAA treatment; and the presence one or more AuxREs in 85% (17/20) of the AUX/IAA gene promoters and 90% (9/10) of the GH3 gene promoters in SmARFs indicated their regulatory roles in danshen growth and development. Further biochemical and genetic studies of candidate ARFs in S. miltiorrhiza will lead to the production of a working model for the cultivation and selective breeding of fine varieties of medicinal plants.

In summary, 25 ARF gene members (seven transcriptional activators and 18 repressors) in S. miltiorrhiza were identified, and a comprehensive account of this gene family has been performed. SmARFs were grouped into four classes with AtARFs in Arabidopsis, and the gene structures, functional domains, and miRNA targets of SmARFs were analyzed in detail. Expression patterns were used to predict candidate SmARFs involved in the regulation of various developmental processes. The results of this study will provide a basic foundation for the verification of the functions and evolution of SmARF gene family members in this model medicinal plant.

MATERIALS AND METHODS

Genome-wide survey of ARF genes in S. miltiorrhiza

The Arabidopsis ARF protein sequences (AtARF1 to AtARF23) were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov/protein/). BLASTP searches were used to identify the corresponding ARF gene members...
in *S. miltiorrhiza* using a cut-off e-value of 1.0E\(^{-10}\). The hidden Markov model (HMM) profiles of ARF gene family members including B3-DBD (Pfam02362), AUX_RESP (MR, Pfam06507), and AUX/IAA family (PBI, Pfam02309) members were applied to identify ARF genes based on the *S. miltiorrhiza* genome. The domains of all obtained ARFs were analyzed using BLAST from the Conserved Domain Database (http://www.ncbi.nlm.nih.gov/ncdd). The auxin response genes, AUX/IAA gene family (Pfam02309) members and GH3 gene family (Pfam03211) members were also selected using the same approach. The Compute pl/Mw tool on the ExPASy server (http://web.expasy.org/compute_pi/) was employed to predict the theoretical isoelectric point (pl) and the molecular weight (Mw) of each SmARF protein.

**Gene structure, conserved motif and subcellular localization analyses**

The Gene Structure Display Server (GSDS 2.0; http://gsds.cbi.pku.edu.cn/index.php) was used to analyze the gene structure of SmARFs with the input of coding sequences (CDSs) and corresponding genomic sequences. Conserved motifs in SmARF transcription factors were identified using MEME (Suite version 4.9.1; http://meme-suite.org/tools/meme) according to the following criteria: maximum number of 15 motifs and an optimum width of 8-50 amino acids. subCELlular LOcalization predictor (CELLO v.2.5; http://cello.life.nctu.edu.tw/) was used to predict the subcellular localization of SmARF proteins.

**Phylogenetic tree construction and miRNA target site analysis**

All SmARF and AtARF protein sequences were pooled into MEGA6 (http:// www.megasoftware.net/) to perform multiple sequence alignments. Then neighbor-joining trees were constructed using the bootstrap method with 1000 replications and pairwise deletion of gaps/missing data. The miRNA target sites of *AtmiR160* and *AtmiR167* in the SmARFs were searched using the PMRD database (http://bioinformatics.cau.edu.cn/PMRD/).

**Plant resources**

*S. miltiorrhiza* (line 99-3) was cultivated at the Institute of Medicinal Plant Development (IMPLAD), Chinese Academy of Medical Sciences (CAMS), in an open experimental field. Three-year-old roots, stems, and flowers were collected. The roots were peeled into three parts (periderm, phloem and xylem) (Xu et al., 2015). Leaves with or without MeJA treatment (12 h, 200 μM; Sigma-Aldrich, MO, USA) were collected from tissue culture plantlets of *S. miltiorrhiza* at 25°C under a long day of 16-h light/8-h dark (Zhang et al., 2015). For auxin treatment, seedlings from tissue culture plantlets were incubated for 0.5 h, 1 h, or 3 h at 20°C, 25°C and 30°C in an open experimental field. Three-year-old roots, stems, and flowers were generated using Illumina-derived short reads to the *S. miltiorrhiza* genome. The domains of all obtained ARFs were analyzed using BLAST from the Conserved Domain Database (http://www.ncbi.nlm.nih.gov/ncdd). The auxin response genes, AUX/IAA gene family (Pfam02309) members and GH3 gene family (Pfam03211) members were also selected using the same approach. The Compute pl/Mw tool on the ExPASy server (http://web.expasy.org/compute_pi/) was employed to predict the theoretical isoelectric point (pl) and the molecular weight (Mw) of each SmARF protein.

**Gene expression analysis by qRT-PCR**

Four RNA samples of seedlings from tissue culture plantlets that were treated with IAA (mock, 0.5 h, 1 h, or 3 h) were isolated. Total RNA was isolated from three biological replicates for each sample using the RNeasy Plus Mini Kit (Qiagen, Germany). Reverse transcription was performed using PrimeScript™ Reverse Transcriptase (TaKaRa, Japan). The qRT-PCR primers were designed using Primer Premier 6 (Table S5), and their specificity was verified by PCR. qRT-PCR analysis was conducted in triplicate using SYBR® Premix Ex Taq™ II (TaKaRa, Japan), with *SmActin* as a reference gene, with a LightCycler 480 real-time PCR system (Roche, Switzerland). Ct values were calculated to analyze the relative expression levels using the 2^−ΔΔCt method (Livak and Schmittgen, 2001). To detect differences in the expression of candidate genes between IAA treatment durations, one-way ANOVA was performed using IBM SPSS 20 software (IBM Corporation, USA). P<0.05 (*) and P<0.01 (**) were considered to indicate significant differences in expression.

**Competing interests**

The authors declare no competing or financial interests.

**Author contributions**

Z.X., J.S. and S.C. designed and coordinated the study. Z.X. and A.J. performed experiments. Z.X. analyzed the data. Z.X. and J.S. wrote the manuscript.

**Funding**

This work was supported by the National Natural Science Foundation of China [Grant no. 81573398] and the National Science-Technology Support Plan of China [Grant no. 2012BAI29B01].

**Supplementary information**

Supplementary information available online at http://bio.biologists.org/lookup/doi/10.1242/bio.017178.supplemental

**References**


