Molecular cloning and expression analysis of WRKY transcription factor genes in Salvia miltiorrhiza

Caili Li, Dongqiao Li, Fenjuan Shao and Shanfa Lu*

Abstract

Background: WRKY proteins comprise a large family of transcription factors and play important regulatory roles in plant development and defense response. The WRKY gene family in Salvia miltiorrhiza has not been characterized.

Results: A total of 61 SmWRKYs were cloned from S. miltiorrhiza. Multiple sequence alignment showed that SmWRKYs could be classified into 3 groups and 8 subgroups. Sequence features, the WRKY domain and other motifs of SmWRKYs are largely conserved with Arabidopsis AtWRKYs. Each group of WRKY domains contains characteristic conserved sequences, and group-specific motifs might attribute to functional divergence of WRKYs. A total of 17 pairs of orthologous SmWRKY and AtWRKY genes and 21 pairs of paralogous SmWRKY genes were identified. Maximum likelihood analysis showed that SmWRKYs had undergone strong selective pressure for adaptive evolution. Functional divergence analysis suggested that the SmWRKY subgroup genes and many paralogous SmWRKY gene pairs were divergent in functions. Various critical amino acids contributed to functional divergence among subgroups were detected. Of the 61 SmWRKYs, 22, 13, 4 and 1 were predominantly expressed in roots, stems, leaves, and flowers, respectively. The other 21 were mainly expressed in at least two tissues analyzed. In S. miltiorrhiza roots treated with MeJA, significant changes of gene expression were observed for 49 SmWRKYs, of which 26 were up-regulated, 18 were down-regulated, while the other 5 were either up-regulated or down-regulated at different time-points of treatment. Analysis of published RNA-seq data showed that 42 of the 61 identified SmWRKYs were yeast extract and Ag⁺-responsive. Through a systematic analysis, SmWRKYs potentially involved in tanshinone biosynthesis were predicted.

Conclusion: These results provide insights into functional conservation and diversification of SmWRKYs and are useful information for further elucidating SmWRKY functions.

Background

Salvia miltiorrhiza Bunge (Lamiaceae), known as danshen in Chinese, is one of the most important Traditional Chinese Medicine (TCM) materials. It has been widely used in Chinese medicines treating coronary heart disease, hepatitis, menstrual disorders, menostasis, blood circulation diseases, and other cardiovascular diseases [1]. The main bioactive components of S. miltiorrhiza include the water-soluble (hydrophilic) phenolics, such as rosmarinic acid, salvianolic acid A, salvianolic acid B and lithospermic acid [2], and the lipid-soluble (nonpolar, lipophilic) diterpenoids, known as tanshinones [3]. Enzymes catalyzing the biosynthesis of these bioactive components have been intensively studied recently [4-10]. A large number of genes involved in the biosynthesis of phenolics and terpenoids have been identified through either molecular cloning or transcriptome-wide characterization [3,11-17]. Collectively, S. miltiorrhiza is being developed to be a medicinal model plant [18].

Transcription factors are a class of significant regulators controlling plant growth and development through regulating gene expression at the transcriptional level. They bind to the specific regions, known as cis-elements, in the promoters of genes and then activate or repress the expression of regulated genes in collaboration with other regulatory factors. So far, two large transcription factor gene families, including the plant-specific SQUAMOSA promoter-binding protein-like (SPL) transcription factor gene family and the largest plant transcription factor gene families, have been identified.
factor gene family, termed MYB, have been characterized in *S. miltiorrhiza* [19,20]. A total of 15 SmSPLs and 110 SmMYBs have been identified from *S. miltiorrhiza*. SmSPLs are involved in the regulation of developmental timing in *S. miltiorrhiza* and eight of them are targets of miR156/157 [19]. Similarly, a subset of SmMYBs is regulated by microRNAs, such as miR159, miR319, miR828 and miR858. Many SmMYBs are involved in the biosynthesis of bioactive compounds in *S. miltiorrhiza* [20].

WRKY is a large transcription factor gene family specific to the green lineage, including green algae and land plants. The first WRKY gene, known as SPF1, was cloned from *Ipomoea batatas* about twenty years ago [21]. Since then, great progress has been achieved in WRKY gene identification and functional analysis. Plants with WRKYs identified include green alga (*Chlamydomonas reinhardtii*) [22], moss (*Physcomitrella patens*) [22], fern (*Ceratopteris richardii*) [22], pine (*Pinus monticola*) [23], *Arabidopsis* [24], tobacco (*Nicotiana tabacum*) [25-27], rice (*Oryza sativa*) [28,29], soybean (*Glycine max*) [30], maize (*Zea mays*) [31], barley (*Hordeum vulgare*) [32], grape (*Vitis vinifera*) [33,34], poplar (*Populus trichocarpa*) [35], tomato (*Solanum lycopersicum*) [36], cucumber (*Cucumis sativus*) [37], coffee (*Coffea arabica*) [38], and so forth.

Characterization of the identified WRKY genes showed that they were significant regulators involved in plant developmental processes and responded to biotic and abiotic stresses [39]. The involvement of WRKYs in plant immune response against bacterial, fungal, and viral pathogens has been widely reported [40-51]. Recently, more and more evidence showed the regulatory roles of WRKY in plant response to abiotic stresses. For example, over-expression of three soybean WRKY genes (*GmWRKY13*, *GmWRKY27* and *GmWRKY54*) in *Arabidopsis* showed that *GmWRKY21*-transgenic *Arabidopsis* plants were tolerant to cold stress, *GmWRKY54* conferred salt and drought tolerance, whereas transgenic plants over-expressing *GmWRKY13* had increased sensitivity to salt and mannitol stresses and decreased sensitivity to abscisic acid [52]. It suggests the involvement of WRKY genes in multiple abiotic stress-associated signaling pathways and the association of different WRKY members with different abiotic stresses. Moreover, WRKYs associated with same abiotic stress may show different responses. For instance, *Arabidopsis* *WRKY25* and *WRKY26* are heat-induced, whereas *WRKY33* is heat-repressed [53]. In addition to stress responses, WRKYs also regulate various developmental processes, such as seed dormancy and germination, flowering, fruit maturation, stem elongation, pith secondary cell wall formation, plant senescence, and trichome development [54-58]. It suggests the importance of WRKYs and the complexity of WRKY-associated regulatory networks.

The defining feature of WRKY transcription factors is their DNA-binding domain, known as WRKY domain [39]. It is approximately 60 amino acids in length and includes the conserved amino acid sequence WRYGQK at the N-terminus and an atypical zinc-finger motif either C2H2 (C–X4–5–C–X22–23–H–X1–H) or C2HC (C–X2– C–X23–H–X1–C) at the C-terminus. The structure of the WRKY domain allows it to specifically interact with W-box and SURE (sugar responsive) cis-elements in the promoter of target genes [59-61]. WRKY can be divided into three groups (Groups I, II and III) based on the number of WRKY domains (two domains in Group I and one in the others) and the pattern of zinc finger motif (C2H2 in Groups I and II and C2HC in Group III) [39,40]. Additionally, Group II WRKY proteins can be further divided into subgroups, including IIa, IIb, IIc, IId and IIe, based on the primary amino acid sequence of the WRKY domain.

Although WRKYs have been identified and characterized in various plant species, no information is available for the WRKY gene family in *S. miltiorrhiza*. In this study, we cloned and characterized 61 *S. miltiorrhiza* SmWRKYs.

**Results and discussion**

**Molecular cloning of 61 SmWRKY genes from *S. miltiorrhiza***

It has been shown that 72 *AtWRKY* genes exist in the *Arabidopsis* genome (Additional file 1: Table S1). To identify SmWRKYs, BLAST analysis against the current assembly of the *S. miltiorrhiza* genome was performed using AtWRKY protein sequences as queries. A total of 61 gene models were predicted for SmWRKYs. The 5’-sequence of many SmWRKYs showed low homology with known plant WRKYs. It might cause errors in computational prediction. To verify the predicted gene models and correct errors of computation, full-length coding sequences (CDSs) of all 61 SmWRKYs were PCR-amplified using the primers listed in Additional file 2: Table S2 and then cloned and sequenced. It resulted in the identification of 61 SmWRKYs, which were named SmWRKY1–SmWRKY61, respectively. The deduced SmWRKY proteins have amino acid numbers from 129 to 706, isoelectric points (pI) from 4.76 to 9.9, and molecular weights (Mw) from 19.9 to 76.2 kDa. All of the 61 cloned CDSs have been submitted to GenBank under the accession numbers shown in Table 1. The number of identified SmWRKYs is comparable with that in *Arabidopsis*. Comparable gene numbers were also found for the MYB [20], SPL [19], *Argonaute (AGO)* [62] and RNA-dependent RNA polymerase (RDR) [63] gene families in *S. miltiorrhiza* and *Arabidopsis*. It suggests that *S. miltiorrhiza* and *Arabidopsis* may have similar number of tfmko gene members for many gene families. Thus, the identified SmWRKYs represent an almost complete set of WRKYs in *S. miltiorrhiza*, although it may be not a fully complete set.
| Name   | Gene ID   | AA Len | pI   | Mw (Da) | Group | Conserved motif | Domain pattern | Zinc finger |
|--------|-----------|--------|------|---------|-------|-----------------|----------------|-------------|
| SmWRKY1| KM823124  | 486    | 5.79 | 52075.53| 2b    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY2| KM823125  | 526    | 7.2  | 57916.44| 1     | 2×[WRKYGGK]    | C-X4-C-X22-HXH| C2H2        |
| SmWRKY3| KM823126  | 295    | 9.84 | 31785.84| 2d    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY4| KM823127  | 292    | 7.05 | 32163.19| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY5| KM823128  | 283    | 6.6  | 31674.21| 2e    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY6| KM823129  | 332    | 9.6  | 36424.23| 2d    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY7| KM823130  | 262    | 6.18 | 29689.04| 3     | WRKYGGK        | C-X7-C-X23-HXC| C2HC        |
| SmWRKY8| KM823131  | 300    | 5.18 | 33675.01| 2e    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY9| KM823132  | 269    | 8.91 | 29956.66| 2a    | WRKYGGK        | C-X5-C-X23-HNH| C2H2        |
| SmWRKY10| KM823133 | 343    | 5.37 | 32404.49| 3     | WRKYGGK        | C-X7-C-X23-HXC| C2HC        |
| SmWRKY11| KM823134 | 349    | 5.34 | 39120.3 | 3     | WRKYGGK        | C-X7-C-X23-HXC| C2HC        |
| SmWRKY12| KM823135 | 211    | 8.11 | 24263.86| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY13| KM823136 | 435    | 5.7  | 47512.82| 1     | 2×[WRKYGGK]    | C-X4-C-X22-HXH| C2H2        |
| SmWRKY14| KM823137 | 243    | 8.19 | 27565.82| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY15| KM823138 | 549    | 6.39 | 50911.35| 2b    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY16| KM823139 | 321    | 5.02 | 35954.3 | 3     | WRKYGGK        | C-X7-C-X23-HXC| C2HC        |
| SmWRKY17| KM823140 | 157    | 9.46 | 18126.3 | 2c    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY18| KM823141 | 333    | 4.76 | 36599.9 | 2e    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY19| KM823142 | 221    | 9.8  | 25422.86| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY20| KM823143 | 328    | 9.56 | 35953  | 2d    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY21| KM823144 | 309    | 6.18 | 34870.74| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY22| KM823145 | 407    | 8.72 | 44533.6 | 2b    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY23| KM823146 | 341    | 9.61 | 37956.92| 2d    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY24| KM823147 | 364    | 7.66 | 40559.08| 1     | 2×[WRKYGGK]    | C-X4-C-X22-HXH| C2H2        |
| SmWRKY25| KM823148 | 329    | 5.9  | 36268.73| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY26| KM823149 | 445    | 6.33 | 49392.18| 2b    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY27| KM823150 | 346    | 9.77 | 37310.4 | 2d    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY28| KM823151 | 486    | 8.38 | 53229.94| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY29| KM823152 | 519    | 6.13 | 63253.29| 2b    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY30| KM823153 | 509    | 6.84 | 55355.55| 2b    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY31| KM823154 | 516    | 8.6  | 55885.41| 1     | 2×[WRKYGGK]    | C-X4-C-X23-HXH| C2H2        |
| SmWRKY32| KM823155 | 129    | 9.3  | 15217.26| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY33| KM823156 | 175    | 9.36 | 19944.38| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY34| KM823157 | 309    | 8.11 | 34006.08| 2a    | WRKYGGK        | C-X4-C-X23-HNH| C2H2        |
| SmWRKY35| KM823158 | 508    | 6.12 | 55526.63| 2b    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY36| KM823159 | 246    | 6.61 | 27210.05| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY37| KM823160 | 290    | 6.6  | 32766.16| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY38| KM823161 | 284    | 9.41 | 30581.63| 2d    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY39| KM823162 | 390    | 6.64 | 43510.55| 1     | 2×[WRKYGGK]    | C-X4-C-X23-HXH| C2H2        |
| SmWRKY40| KM823163 | 352    | 6.07 | 38849.51| 2e    | WRKYGGK        | C-X5-C-X23-HXH| C2H2        |
| SmWRKY41| KM823164 | 587    | 9.12 | 64687.26| 1     | 2×[WRKYGGK]    | C-X4-C-X23-HXH| C2H2        |
| SmWRKY42| KM823165 | 706    | 6.02 | 76197.45| 1     | 2×[WRKYGGK]    | C-X4-C-X23-HXH| C2H2        |
| SmWRKY43| KM823166 | 179    | 5.39 | 19963.09| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
| SmWRKY44| KM823167 | 265    | 5.73 | 28206.05| 2c    | WRKYGGK        | C-X4-C-X23-HXH| C2H2        |
Table 1 Sequence features of WRKYs in *S. miltiorrhiza* (Continued)

| SmWRKY   | KMB23168 | 336 | 5.95 | 37211.08 | 3 | WRKYGQK | C-X$_1$-C$_{23}$-HXC | C2HC |
|----------|----------|-----|------|----------|---|---------|---------------------|------|
| SmWRKY46 | KMB23169 | 306 | 6.41 | 33426.96 | 2e | WRKYGQK | C-X$_3$-C$_{23}$-HXC | C2H2 |
| SmWRKY47 | KMB23170 | 268 | 5.86 | 30225.72 | 2c | WRKYGQK | C-X$_4$-C$_{23}$-HXC | C2H2 |
| SmWRKY48 | KMB23171 | 224 | 5.88 | 25174.62 | 2e | WRKYGQK | C-X$_5$-C$_{23}$-HXC | C2H2 |
| SmWRKY49 | KMB23172 | 351 | 9.9  | 39451.59 | 2d | WRKYGQK | C-X$_2$-C$_{23}$-HXC | C2H2 |
| SmWRKY50 | KMB23173 | 291 | 5.45 | 33184.06 | 2e | WRKYGQK | C-X$_7$-C$_{23}$-HXC | C2H2 |
| SmWRKY51 | KMB23174 | 526 | 7.2  | 57916.44 | 1 | 2x[WRKYGQK] | C-X$_4$-C$_{22}$-HXC | C2H2 |
| SmWRKY52 | KMB23175 | 171 | 6.66 | 19066.92 | 2c | WRKYGQK | C-X$_4$-C$_{23}$-HXC | C2H2 |
| SmWRKY53 | KMB23176 | 575 | 6.54 | 62764.44 | 1 | 2x[WRKYGQK] | C-X$_4$-C$_{22}$-HXC | C2H2 |
| SmWRKY54 | KMB23177 | 491 | 7.69 | 53504.9 | 1 | 2x[WRKYGQK] | C-X$_4$-C$_{22}$-HXC | C2H2 |
| SmWRKY55 | KMB23178 | 449 | 9.12 | 49229.93 | 1 | 2x[WRKYGQK] | C-X$_4$-C$_{22}$-HXC | C2H2 |
| SmWRKY56 | KMB23179 | 297 | 5.34 | 33782.19 | 2e | WRKYGQK | C-X$_7$-C$_{23}$-HXC | C2H2 |
| SmWRKY57 | KMB23180 | 281 | 5.18 | 31675.83 | 2c | WRKYGQK | C-X$_4$-C$_{22}$-HXC | C2H2 |
| SmWRKY58 | KMB23181 | 272 | 6.32 | 30091.63 | 2a | WRKYGQK | C-X$_5$-C$_{23}$-HNC | C2H2 |
| SmWRKY59 | KMB23182 | 352 | 8.34 | 38285.97 | 2b | WRKYGQK | C-X$_5$-C$_{23}$-HXC | C2H2 |
| SmWRKY60 | KMB23183 | 379 | 7.63 | 42242.82 | 1 | 2x[WRKYGQK] | C-X$_4$-C$_{22}$-HXC | C2H2 |
| SmWRKY61 | KMB23184 | 168 | 8.96 | 19153.70 | 3 | WRKYGQK | C-X$_2$-C$_{23}$-HXC | C2HC |

**Classification of the WRKY domains and the WRKY proteins**

Transcription factors in a family usually share highly conserved DNA-binding domains. In order to examine the phylogenetic relationships among WRKYs, a neighbor-joining (NJ) phylogenetic tree was constructed for the WRKY domain sequences of AtWRKYs, SmWRKYs, PpWRKYs, GCMa and FLYWCH using MEGA5.0 (Figure 1). According to the classification of AtWRKYs, the WRKY domains were divided into 3 groups (groups 1, 2 and 3). Group 1 WRKY domains come from proteins containing two WRKY domains, one of which is located in the N-terminal (NTWD), while the other one is in the C-terminal (CTWD). The exceptions within this group are the domains from AtWRKY10 and PpWRKY13, each of which possesses a single WRKY domain. Group 1 WRKY domains were further divided into two subgroups, termed group 1 N and group 1C, respectively. Based on their characteristics in the phylogenetic tree, group 2 WRKY domains could be classified into 5 subgroups, including groups 2a, 2b, 2c, 2d and 2e. Multiple sequence alignment of the core WRKY domains, each of which contains approximately 60 residues, showed that 71 of the 74 WRKY domains from 61 SmWRKYs contained the highly conserved sequence WRKYGQK, while the other three (SmWRKY43, SmWRKY52, and SmWRKY61) had WRKYGKKK (Figure 2). Of the 61 SmWRKY, 13 are two-WRKY-domain-containing proteins and all of them have the C2H2-type zinc-finger motif (C–X$_4$–C–X$_{22}$–H–X$_1$–H) (Figure 2, Table 2). The other 48 SmWRKY proteins are one-WRKY-domain-containing proteins, 42 of which contain group 2 WRKY domains and have the same type of finger motif (C–X$_{4.5}$–C$_{23}$–H–X$_1$–H), while the other 6 contain group 3 WRKY domains and have the C2HC zinc finger motif (C–X$_7$–C$_{23}$–H–X$_1$–C) (Figure 2, Table 2). The distribution of residues in the WRKY domains of *S. miltiorrhiza* WRKY proteins is quite similar to that of *Arabidopsis* (Figures 2 and 3), suggesting evolutionary conservation of SmWRKYs and AtWRKYs. Comparing the number of SmWRKYs and AtWRKYs in each group/subgroup showed that the number of SmWRKYs in groups 1 and 2 is similar to that of AtWRKYs in the same group; however the number 6 of SmWRKY members belonging to group 3 is significantly less than the number 14 of AtWRKY members included in the same group. It is consistent with previous results showing group 3 WRKYs to be a newly defined and most dynamic group [22] and suggests the divergence of WRKYs in *S. miltiorrhiza* and *Arabidopsis*.

In order to investigate whether the phylogenies are different between the WRKY domains and the corresponding WRKY proteins, we constructed an NJ tree based on the full-length amino acid sequences of SmWRKYs, AtWRKYs, PpWRKYs, GCMa and FLYWCH (Figure 4). The results showed that the phylogenetic tree of WRKY proteins was quite similar to the tree of WRKY domains with little difference observed (Figures 1 and 4). For instance, AtWRKY1, AtWRKY32 and SmWRKY28 having two WRKY domains and AtWRKY10 belonging to group 1 WRKY domains form separated clades outside group 1. AtWRKY19 and FLYWCH with the WRKY domain belonging to group 1, AtWRKY16 with the WRKY domain belonging to group 2e, and AtWRKY52 and GCMa with the WRKY domain belonging to group 3 form separated clades outside groups 1, 2 and 3. These results indicate the
difference between the WRKY domain and the sequence outside the WRKY domain. Based on the NJ tree constructed with full-length amino acid sequences, we identified 17 pairs of orthologous WRKY genes in *S. miltiorrhiza* and *Arabidopsis* (Table 3). It suggests that many SmWRKYs and AtWRKYs are evolutionarily conserved.

**Analysis of conserved motifs**

In addition to the WRKY domain, other conserved motifs could be important for the diversified functions of WRKY proteins from *S. miltiorrhiza* and *Arabidopsis* [64,65]. Using the MEME program, we identified a total of 20 conserved motifs in WRKYs from *S. miltiorrhiza* and *Arabidopsis* (Figure 5). The length of motifs varies from 8 to 150 amino acids and the number of motifs in each WRKY varies between 2 and 11. The majority of the identified motifs were found in more than one subgroup of WRKYs. Many AtWRKYs in a subgroup contain the same motif(s) as their SmWRKY orthologues in the subgroup. It suggests the conservation of motifs in *S. miltiorrhiza* and *Arabidopsis* WRKYs belonging to a subgroup.

Among the 20 conserved motifs, 9 motifs, including motifs 1, 2, 3, 4, 5, 6, 8, 13, and 19, are located in the WRKY domain, while the other 11 are located outside the domain. Most WRKY proteins in a group have similar
motif compositions (Figure 5). For instance, motif 18 exists in almost all of group 1 WRKY proteins except for SmWRKY15. Motif 14 exists in 16 of the 21 group 1 S. miltiorrhiza and Arabidopsis WRKY proteins. Motifs 9 and 10 commonly exist in groups 2a and 2b, two close subgroups in the phylogenetic tree. Motifs 12, 16 and 17 exist in most WRKY proteins of group 2b but only in a few members of 2a. Motif 7 exists in all of the WRKY proteins belonging to group 2c. Motif 11 is shared by proteins belonging to groups 2d and 2e. Additionally, motif 15, known as the Ca$^{2+}$-dependent CaM-binding domain (CaMBD) [66], commonly exists in most WRKY proteins belonging to groups 2d and 3. Group 2d WRKY proteins usually contain two motif 15s, while the majority of group 3 proteins contain only one. Motif 20, known as the HARF motif [40,67], exists in 5 of 7 AtWRKYs and all 7 SmWRKYs in group 2d. The results indicate functional similarities of WRKY proteins belonging to a group. Group-specific motifs may attribute to functional divergence of WRKYs.

### Table 2 Number of WRKY domains from S. miltiorrhiza and Arabidopsis

| Group | Subgroup | Gene number | Consensus | Exception |
|-------|----------|-------------|-----------|-----------|
|       | AtWRKY   | SmWRKY      |           |           |
| 1     | 1 N      | 27          | 13        | 26        | 13        | C-X$_{4}$-C-X$_{22}$-HXXH | n = 23, AtWRKY26 |
|       | 1C       | 14          | 14        | 13        |           | C-X$_{4}$-C-X$_{22}$-HXXH |
| 2     | 2a       | 44          | 3         | 42        | 3         | C-X$_{5}$-C-X$_{22}$-HXXH |
|       | 2b       | 8           | 8         | 8         |           | C-X$_{5}$-C-X$_{22}$-HXXH |
|       | 2c       | 18          | 18        | 16        |           | C-X$_{5}$-C-X$_{22}$-HXXH |
|       | 2d       | 7           | 7         | 7         |           | C-X$_{5}$-C-X$_{22}$-HXXH |
|       | 2e       | 8           | 8         | 8         |           | C-X$_{5}$-C-X$_{22}$-HXXH |
| 3     | 3        | 14          | 14        | 6         | 6         | C-X$_{7}$-C-X$_{22}$-HXC | n = 22, m = 5, SmWRKY61 |
|       |          |             |           |           |           | AtWRKY52: HNH for HXC |
| Total |          | 85          | 74        |           |           |
Selective constraints on SmWRKY genes

In order to preliminarily examine the evolutionary mechanism of WRKYs, we test the hypothesis of positive selection acting on SmWRKY genes using site models and branch-site models in PAML [68] developed by Nielsen and Yang [69] and Yang et al. [70]. Codon substitution models [71] M0, M1a, M2a, M3, M7 and M8 were applied to the alignments and these models assume variation in $\omega$ among sites. The parameter estimates, log-likelihood and the LRT tests for these models are shown in Table 4. To examine how $dN/dS$ ratios differed among codon positions, we compared models M0 and M3. The log likelihood of M0 for SmWRKY sequences was $l = -5119.032$, with an estimate of $\omega = 0.038$. The low $\omega$ value suggests a strong action of purifying selection in the evolution of SmWRKY analyzed. M3 (discrete) assumes a general discrete distribution with three site classes ($p_0$, $p_1$, $p_2$). The log likelihood of M3 was $l = -5119.251$. However, no site was positively selected at a level of 95%. M7 and M8 fitted the sequences better than M0, M3, M1a and M2a with values of $l = -4857.207$ and $-4857.206$, respectively (Table 4). In both cases, no significant evidence of positive selection was found.

Branch-site models aim to detect positive selection affecting a few sites along particular lineages and allow $\omega$ ratios to vary among sites and lineages simultaneously [68]. It seems that the branch-site models are most suitable for describing evolutionary processes of the WRKY gene family. Therefore, we analyzed positively selected amino acid sites of SmWRKYs using the improved branch-site model [72]. The branches being tested for...
positive selection were used as the foreground, while all other branches on the tree were used as the background. The Bayes empirical Bayes (BEB) method was used to calculate the posterior probabilities. A codon is probably from the site class of positive selection if LRT suggested the presence of codons under positive selection on the foreground branch [73]. The parameter estimates for lineages under positive selection are given in Table 5. A total of 19 residues were found to be under positive selection ($p > 90\%$). It includes 6 in group 2c and 10 in group 2d. The other three residues were found in group 2b, group 2e and group 3. No residues in group 1 and group 2a were found to be under positive selection. The results suggest that different WRKY groups may have different evolutionary rates. Groups 2c and 2d could be confronted with strong positive Darwinian selection, since many highly

Figure 4 Phylogenetic analysis of SmWRKYs (red), AtWRKYs (blue), PpWRKYs (dark green), FLYWCH (light green) and GCMa (pink) proteins. The unrooted phylogenetic tree was constructed using the neighbor-joining method. The names of WRKY proteins not included in a group are shown.
Table 3 WRKY orthologs in *S. miltiorrhiza* and *Arabidopsis*

| SmWRKYs   | Putative Arabidopsis orthologs | Phylogenetic group in the NJ tree |
|-----------|-------------------------------|----------------------------------|
| SmWRKY55  | AtWRKY44                      | Group 1                          |
| SmWRKY54  | AtWRKY7/AtWRKY4               | Group 1                          |
| SmWRKY53  | AtWRKY20                      | Group 1                          |
| SmWRKY51/SmWRKY2 | AtWRKY33                   | Group 1                          |
| SmWRKY28  | AtWRKY32                      | Out of group 1                   |
| SmWRKY34/SmWRKY58 | AtWRKY40                | Group 2a                         |
| SmWRKY59  | AtWRKY72                      | Group 2b                         |
| SmWRKY57  | AtWRKY123                     | Group 2c                         |
| SmWRKY12  | AtWRKY12                      | Group 2c                         |
| SmWRKY19  | AtWRKY13                      | Group 2c                         |
| SmWRKY47  | AtWRKY49                      | Group 2c                         |
| SmWRKY49  | AtWRKY39                      | Group 2d                         |
| SmWRKY56  | AtWRKY29                      | Group 2e                         |
| SmWRKY40  | AtWRKY35/AtWRKY14             | Group 2e                         |
| SmWRKY8   | AtWRKY27                      | Group 2e                         |
| SmWRKY45  | AtWRKY55                      | Group 3                          |
| SmWRKY16  | AtWRKY30                      | Group 3                          |

significant positive sites were detected at the 0.01 significance level (Table 5). The evolution in the other groups seems to be more conservative.

Functional divergence analysis (FDA) of SmWRKY proteins

Using DIVERGE 2.0 that evaluates shifted evolutionary rate and altered amino acid property after gene duplication [74,75], we carried out posterior analysis for estimation of Type-I and Type-II functional divergence between SmWRKY clusters. The estimation was based on the WRKY protein neighbor-joining tree consisting of three major groups (group 1, group 2a–e, and group 3) (Figure 4). Comparison among SmWRKY subgroups showed that all of the coefficients for the type I functional divergence ($\theta_1$) were greater than zero (Additional file 3: Table S3). The $\theta_1$ values of eight group pairs, including 1/2, 1/3, 2a + b/2d, 2a + b/2e, 2a + b/3, 2c/2e, 2c/3 and 2d/3, were ranged from 0.219 to 0.772 and were statistically significant ($P < 0.01$) (Additional file 3: Table S3). It indicates that significant site-specific changes may have happened at certain amino acid sites between these group pairs, leading to a subgroup-specific functional evolution after their diversification.

Type II functional divergence ($\theta_{II}$) values in six group pairs, including 1/2c, 1/2d, 2a + b/2d, 2c/2d and 2d/3, were also greater than zero and ranged from 0.017 to 0.234 (Additional file 4: Table S4). It indicates a radical shift in amino acid properties. In order to extensively reduce positive false, $Q_c > 0.8$ and 1.0 were empirically used as cutoff in the identification of the Type-I and Type-II functional divergence-related residues between gene groups, respectively (Figures 6 and 7). Detailed analysis showed that the number and the distribution of predicted residues for functional divergence in group pairs were different (Additional file 3: Table S3 and Additional file 4: Table S4) and the residues predominantly existed in the WRKY domain. It suggests that these residues probably play important roles in functional divergence of WRKYS during the evolutionary process.

Tissue-specific expression of SmWRKYS

It has been shown that a number of WRKY proteins are involved in plant developmental processes [54,76,77]. In order to preliminarily understand the roles of WRKYS in *S. miltiorrhiza* development, we analyzed the expression of SmWRKYS in roots, stems, leaves and flowers of *S. miltiorrhiza* plants. All of the 61 SmWRKYS identified were expressed in at least a tissue analyzed and exhibited differential expression patterns (Figure 8). Of the 61 SmWRKYS, 22 (36.1%) showed predominant expression in roots, 13 (21.3%) in stems, 4 (6.6%) in leaves and 1 (1.6%) in flowers. The other 21 (34.4%) were mainly expressed in at least two tissues analyzed, indicating these genes are likely to play a more ubiquitous role in *S. miltiorrhiza*. Furthermore, some SmWRKYS in a group shared similar expression patterns, while the others were not. For example, SmWRKY2, SmWRKY24, SmWRKY39, SmWRKY54 and SmWRKY55, belonging to group 1, were predominantly expressed in roots, while the other group 1 members, such as SmWRKY42, SmWRKY13 and SmWRKY60, were mainly expressed in stems, leaves and flowers, respectively (Figure 8). It suggests that SmWRKYS belonging to a group do not necessarily indicate their functions in the same tissues. However, it has been shown that the tissues-specific expression patterns appear to be consistent with their role in the tissues. For example, VvWRKY01, belonging to group 2c, is involved in the regulation of lignin biosynthesis [78]. Over-expression of VvWRKY01 in tobacco resulted in the alteration of expression patterns of genes involved in lignin biosynthesis pathway [78]. Similarly, SmWRKY12, SmWRKY19 and SmWRKY47 in group 2c were predominantly expressed in stems (Figure 8). It indicates the putative roles of these SmWRKYS in the regulation of lignin biosynthesis in *S. miltiorrhiza*. AtWRKY6, a member of group 2b, and AtWRKY53 and AtWRKY70, two AtWRKYS in group 3, are an important regulator in senescent leaves [55,76,79-81]. Of them, AtWRKY6 acts in the upstream of SIRK during leaf senescence [55]. SmWRKY59 belonging to the same WRKY group of AtWRKY6 and SmWRKY7 included in group 3 could be regulators of leaf senescence in *S. miltiorrhiza*, since both of them showed predominant expression in
leaves (Figure 8). It has been shown that AtWRKY44 included in group 1 is involved in trichome differentiation [54]. Several SmWRKYs in group 1, such as SmWRKY2, SmWRKY39, SmWRKY24, SmWRKY54 and SmWRKY55, were highly expressed in roots with abundant root hairs, and SmWRKY13 belonging to the same group was highly expressed in leaves with abundant of trichomes (Figure 8). It implies that these SmWRKY may be associated with trichome development in S. miltiorrhiza.

**Methyl jasmonate (MeJA)-responsive SmWRKYs**

MeJA is a key signaling molecule involved in plant response to stress and in regulating secondary metabolite production in many plant species, including S. miltiorrhiza.
[3,14]. More than 50% (39 of 74) of AtWRKYs were MeJA-responsive in Arabidopsis [82]. More than 1/3 of CrWRKY were regulated by MeJA in hairy roots of Catharanthus roseus [83]. Additionally, CaWRKY27 in Capsicum annuum [84], GbWRKY1 in Gossypium barbadense [85] and GbWRKY3 in Gossypium hirsutum [86] were also regulated by MeJA. In order to test whether WRKYs were responsive to MeJA treatment in S. miltiorrhiza, the expression level of SmWRKYs in roots of plantlets treated with or without MeJA was analyzed using the quantitative RT-PCR method. MeJA treatment showed a wide variety of SmWRKY gene expression profiles (Figure 9). Significant expression level changes were observed for 49 SmWRKYs, of which 26 were up-regulated, 18 were down-regulated, while the other 5, including SmWRKY1, SmWRKY15, SmWRKY17, SmWRKY20 and SmWRKY24, were either up-regulated or down-regulated at different time-points of treatment (Figure 9). It suggests that about 80% of the SmWRKYs analyzed are MeJA-responsive. Examination of the number of SmWRKYs with significant expression level changed at different time-points of treatment showed that the expression of 28, 43, 25 and 23 SmWRKYs was changed after MeJA treatment for 12, 24, 36, and 48 hours, respectively (Figure 9). It suggests that the majority of SmWRKYs have altered expression levels at the time-point of 24 h-treatment. The number of up-regulated SmWRKYs was 13, 26, 15, and 13 at the time-point of 12, 24, 36, and 48 hours, respectively, while that of down-regulated was 15, 17, 10, and 10, respectively. Additionally, 8 SmWRKYs, including SmWRKY9, SmWRKY13, SmWRKY14, SmWRKY25, SmWRKY32, SmWRKY38, SmWRKY44 and SmWRKY52, were significantly up-regulated at all time-points of MeJA treatment, while 7, including SmWRKY7, SmWRKY33, SmWRKY47, SmWRKY49, SmWRKY53, SmWRKY54 and SmWRKY58, were down-regulated (Figure 9). It suggests that the number of up-regulated SmWRKYs is slightly more than down-regulated.

Yeast extract and Ag⁺-responsive SmWRKYs
In order to further investigate the roles of SmWRKYs in S. miltiorrhiza, transcriptome-wide analysis of SmWRKY expression in response to yeast extract and Ag⁺ treatment was performed. RNA-seq data of S. miltiorrhiza hairy roots treated with or without yeast extract (100 μg/ml) and Ag⁺ (30 μM) were downloaded from GenBank under the accession number SRR924662 [12]. RNA-seq reads from non-treated (0 hpi) and treated for 12 h (12 hpi), 24 h (24 hpi) and 36 h (36 hpi) were mapped to SmWRKYs using the SOAP2 software [87]. The log-2-transformed RPKM (RNA-seq reads mapped to a SmWRKY per total million reads from a treatment per kilobases of the SmWRKY length) value of SmWRKYs varied between −3.04 and 8.38 (Additional file 5: Table S5). Using a cutoff of RPKM value ≥2.0, a total of 49 SmWRKYs were found to be expressed in hairy roots. Fisher’s exact test showed that 42 of the 49 SmWRKYs were differentially expressed (Additional file 5: Table S5). It includes 17 significantly up-regulated, 19 significantly down-regulated and 6 significantly up- or down-regulated at different time-points, suggesting the majority of the

| Model                  | lnL  | Estimates of parameter | 2ΔlnL | Positive selection sites |
|------------------------|------|------------------------|-------|--------------------------|
|                        |      | Frequency              | dN/dS |                          |
|                        |      | p = 1.000              | ω = 0.03756 | 508.984 (M3 vs. M0)** | Not allowed |
| M3(discrete)           | −486.454 | p0 = 0.30493            | ω0 = 0.00052 | None |
|                        |      | p1 = 0.32215            | ω1 = 0.02439 | 1 = 0.02439 |
|                        |      | p2 = 0.37292            | ω2 = 0.08755 | 2 = 1.00000 |
| M1a(nearly neutral)   | −5119.251 | p0 = 0.93054            | ω0 = 0.04576 | 0 (M2a vs. M1a) Not allowed |
|                        |      | p1 = 0.06946            | ω1 = 1.00000 | None |
| M2a(positive selection)| −5119.251 | p0 = 0.93052            | ω0 = 0.04576 | None |
|                        |      | p1 = 0.03470            | ω1 = 1.00000 | None |
|                        |      | p2 = 0.03478            | ω2 = 1.00000 | None |
| M7(beta)               | −4857.207 | p = 0.39586            | q = 9.86708 | 0.002 (M8 vs. M7) Not allowed |
|                        |      | q = 9.86708             |           | None |
| M8(beta & ω)          | −4857.206 | p0 = 0.99999            | ω = 1.35981 | None |

Note: *p < 0.05 and **p < 0.01 (χ² test).

1ω was estimated under model M0,M3,M7, and M8. p and q are the parameters of the beta distribution.

2The number of amino acid sites estimated to have undergone positive selection.
identified \textit{SmWRKYs} were responsive to yeast extract and Ag\textsuperscript{+} treatment.

**\textit{SmWRKY} candidates potentially involved in tanshinone biosynthesis**

Terpenoids are plant secondary metabolites with significant physiological and ecological functions and great economic values, and a class of terpenoids, known as tanshiones, is the main bioactive compounds in \textit{S. miltiorrhiza}. Increasing evidence demonstrates the importance of \textit{WRKY} genes in the biosynthesis of secondary metabolites, such as terpenoid indole alkaloid in \textit{Catharanthus roseus} [83]. Additionally, it has been shown that \textit{Gossypium arboresum} \textit{GaWRKY1}, which belongs to group 2a, participates in sesquiterpene biosynthesis in cotton by controlling (+)-\textit{δ}-cadinene synthase (CAD1) activity [88]. \textit{PqWRKY1}, a member of group 2d, responds to MeJA treatment and is a positive regulator of osmotic stress response and triterpene ginsenoside biosynthesis in \textit{Panax quinquefolius} [89]. \textit{AaWRKY1} and \textit{CrWRKY}, belonging to group 3, are the other two terpenoid biosynthesis-related \textit{WRKY}s. \textit{Artemisia annua} \textit{AaWRKY1} was highly expressed in glandular secretory trichomes (GSTs), where the sesquiterpene artemisinin was synthesized [90]. \textit{AaWRKY1} might be strongly induced by MeJA and could bind to the W-box in the promoter of \textit{ADS} gene encoding amorpha-4, 11-diene synthase, a key enzyme in the artemisinin biosynthesis pathway [90]. \textit{CrWRKY1} was preferentially expressed in roots of \textit{C. roseus} and also induced by MeJA [91]. It controlled terpenoid indole alkaloid biosynthesis through positive regulation of \textit{DXS} and \textit{SLS} genes involved in the terpenoid pathway and \textit{AS} and \textit{TDC} genes involved in the indole pathway [91]. Thus, \textit{SmWRKYs} included in group 2a, 2d and 3 probably have an evolutionarily conserved role in regulating terpenoid biosynthesis in \textit{S. miltiorrhiza}.

Terpenoid tanshiones have been mainly produced and accumulated in roots of field-grown \textit{S. miltiorrhiza} during the fast growing period from June to September [92-94]. The process of tanshinone production may be
stimulated by MeJA, yeast extract and Ag⁺ [10,12,95]. Among the 61 identified SmWRKYs, sixteen showed similar responses to the MeJA treatment and the yeast extract and Ag⁺ treatment (Figure 9, Additional file 5: Table S5). SmWRKY2, SmWRKY3, SmWRKY9, SmWRKY11, SmWRKY25, SmWRKY32, SmWRKY37, SmWRKY43 and SmWRKY52 were up-regulated by both the MeJA treatment and the yeast extract and Ag⁺ treatment, while SmWRKY23, SmWRKY26, SmWRKY33, SmWRKY41, SmWRKY47, SmWRKY53 and SmWRKY59 were down-regulated (Figure 9, Additional file 5: Table S5). Among the sixteen SmWRKYs, eight, including six up-regulated (SmWRKY2, SmWRKY3, SmWRKY9, SmWRKY25, SmWRKY37 and SmWRKY52) and two down-regulated (SmWRKY26 and SmWRKY59) were identified as showing similar responses to the MeJA treatment and the yeast extract and Ag⁺ treatment.

**Figure 6 Site-specific prediction for type-I functional divergence between groups of SmWRKYS.** The X-axis represents locations of sites. The Y-axis represents the probability of each group. The red line indicates cutoff = 0.80.
SmWRKY33), were predominantly expressed in roots of field-grown *S. miltiorrhiza* in August (Figure 8), suggesting their specific roles in roots. SmWRKY2, SmWRKY3, SmWRKY9 and SmWRKY26 are members of groups 1, 2d, 2a and 2b, respectively, while SmWRKY25, SmWRKY33, SmWRKY37 and SmWRKY52 are members of group 2c (Figures 1 and 4). Thus, SmWRKY3 and SmWRKY9 are two SmWRKYs (1) with similar responses to the MeJA treatment and the yeast extract and Ag⁺ treatment, (2) having root-specific expression, and (3) belonging to group 2a, 2d or 3 with members probably playing an evolutionarily conserved role in regulating terpenoid biosynthesis.

**Figure 7** Site-specific profile for predicting critical amino acid residues responsible for the type-II functional divergence between groups of SmWRKYs. The X-axis represents locations of sites. The Y-axis represents the probability of each group. The red line indicates cutoff = 1.0.
It implicates that SmWRKY3 and SmWRKY9 are very likely to be activators in tanshinone production. Notably, we may not exclude the possibility that some other SmWRKYs are also involved in tanshinone biosynthesis based on the data currently available. Further analysis of transgenic S. miltiorrhiza plants with over-expressed or silenced SmWRKYs may help to elucidate their function.

**Divergence of paralogous SmWRKY genes**

Gene duplication is an important event for gene family expansion and functional diversity during evolution [65,96,97]. A total of 42 (68.85% of 61) SmWRKY genes appear to be duplicated (Additional file 5: Table S5). In order to preliminarily reveal the mechanism of functional diversity (nonfunctionalization, subfunctionalization and neofunctionalization [98]) of these genes after duplication, the synonymous (Ks) and non-synonymous (Ka) substitution rate were calculated using the CDS of paralogous SmWRKY genes (Additional file 6: Table S6). The Ka/Ks ratios for all of the 21 paralogous SmWRKY gene pairs were less than one with 5 pairs even close to zero. It suggests that the WRKY genes from S. miltiorrhiza have experienced strong purifying selection pressure. Some closely related gene pairs displayed different expression patterns, indicating functional divergences occurred. For example, SmWRKY13 was expressed dominantly in leaves, whereas the other member of the SmWRKY13/SmWRKY31 gene pair, SmWRKY31, was expressed mainly in roots and...
stems (Figure 8). In addition, SmWRKY13, but not SmWRKY31, was significantly up-regulated by MeJA (Figure 9). Expression patterns of other paralogous genes, such as SmWRKY23/49, SmWRKY30/35, SmWRKY41/55, SmWRKY42/53, and SmWRKY43/SmWRKY52, were also different (Figures 8 and 9). It indicates that many SmWRKY gene pairs are divergent under the purifying pressure [99].

Conclusions
In this study, we cloned a total of 61 SmWRKY genes. The cloned genes and the deduced proteins were characterized through a comprehensive approach, including phylogenetic tree construction, WRKY domain characterization, conserved motif identification, selective constraint analysis, functional divergence analysis, and expression profiling. We showed that many SmWRKYs and AtWRKYs were evolutionarily conserved. The WRKY domains could be divided into 3 groups (1, 2 and 3) and 8 subgroups (1 N, 1C, 2a, 2b, 2c, 2d, 2e and 3). Each group of WRKY domains contains characteristic conserved sequences. Additionally, sequence outside the WRKY domain might contribute to the difference between the phylogenetic tree constructed with the WRKY domains and that constructed with the whole WRKY proteins. A total of 20 conserved motifs were identified, of which group-specific motifs might attribute to functional divergence of WRKYS. We identified 17 pairs of orthologous SmWRKY and AtWRKY genes and 21 pairs of

Figure 9 Quantitative RT-PCR analysis of SmWRKY gene expression in S. miltiorrhiza roots treated with MeJA. Fold changes of SmWRKYs in roots of S. miltiorrhiza plantlets treated with MeJA for 12, 24, 36 and 48 h are shown. SmUBQ10 was used as the reference gene. The level of transcripts in roots treated with carrier solution (CK) was arbitrarily set to 1 and the levels in roots treated with MeJA were given relative to this. Mean values and SDs were obtained from three biological and three technical replicates. Y-axis indicates relative expression levels. X-axis indicates different time-points of MeJA treatment. ANOVA (analysis of variance) was calculated using SPSS. P < 0.05 (*) and P < 0.01(**) were considered statistically significant and extremely significant, respectively.
paralogous SmWRKY genes. Selective constraint analysis and functional divergence analysis showed that the SmWRKY subgroup genes have experienced strong positive selection and diverged in function. Gene expression profiles suggest that the majority of 61 SmWRKY genes are tissue-specific and MeJa- and yeast extract and Ag⁺-responsive. These results provide insights into functional conservation and diversification of SmWRKYs and are useful for further investigating SmWRKY functions in S. miltiorrhiza development and defense response.

Methods

Database search and sequence annotation
The nucleotide sequences and amino acids of 72 AtWRKY genes were obtained from the Arabidopsis Information Resource (TAIR; http://www.Arabidopsis.org/). S. miltiorrhiza WRKY (SmWRKY) genes were predicted by tBLASTn [100] search of AtWRKY [101] homologs against the current S. miltiorrhiza genome assembly, which covers about 92% of the entire genome and 96% of the protein-coding genes [18]. An e-value cut-off of 1e-10 was applied to the homologue recognition. The retrieved sequences were used for gene model prediction on the GENSCAN web server (http://genes.mit.edu/GENSCAN.html). Full-length CDSs of SmWRKYs were amplified by reverse transcription-PCR using the primers listed in Additional file 2: Table S2. PCR products were gel-purified, cloned, and then sequenced. The theoretical isoelectric point (pI) and molecular weight (Mw) were predicted using the Compute pI/Mw tool on the ExPASy server (http://web.expasy.org/compute_pi/) [102].

Multiple sequence alignment, phylogenetic analysis and motif detection
PpWRKY genes were obtained from Physcomitrella patens v3.1 (http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Ppatens_er). Human FLYWCH CRAa (EAW85450) and GCMa (BAA13651) were obtained from NCBI (http://www.ncbi.nlm.nih.gov/protein/). Multiple sequence alignment of the WRKY domain from 61 S. miltiorrhiza SmWRKYs and 72 Arabidopsis AtWRKYs was performed using CLUSTALW with BOXSHADE (http://bioweb.pasteur.fr) [103]. Phylogenetic trees were constructed using MEGA 5.0 with the neighbor-joining method [104]. Bootstrap test was replicated 1000 times. Motifs were detected using MEME 5.0 [105].

Plant materials and MeJA treatment
Roots, stems, leaves and flowers from 2-year-old, field-grown S. miltiorrhiza Bunge (line 993) plants were collected in August, 2012 and stored in liquid nitrogen until use. Plantlets cultivated in vitro were grown at 25°C with a photoperiod of 16 h light and 8 h dark for six weeks and treated with 200 μM methyl jasmonate (MeJA) for 12, 24, 36 and 48 h as described previously [3,63]. Plantlets treated with carrier solution were used as controls. Roots of plantlets with or without MeJA treatment were collected and stored in liquid nitrogen until use. Three independent biological replicates were carried out for each experiment.

RNA extraction and quantitative real-time reverse transcription-PCR (qRT-PCR)
Total RNA was extracted from plant tissues using the Quick RNA Isolation Kit (Huayueyung, China). RNA integrity was analyzed on a 1.2% agarose gel. RNA quantity was determined using a NanoDrop 2000C Spectrophotometer (Thermo Scientific, USA). cDNA synthesis was carried out using Superscript III Reverse Transcriptase (TaKaRa, China). qRT-PCRs were performed using the SYBR premix Ex Taq™ kit (TaKaRa, China) and carried out in triplicate for each tissue sample. Gene-specific primers (Additional file 7: Table S7) were designed using Primer Premier 5.0. The length of amplicons is between 80 bp and 250 bp. SmUBQ10 was selected as a reference gene as described previously [3]. Three independent biological replicates were performed. Statistical analysis was carried out as described [20]. Briefly, standardization of gene expression data was performed from three biological replicates as described [106]. 2-ΔΔCq was used to achieve results for relative quantification. For statistical analysis, ANOVA (analysis of variance) was calculated using SPSS (Version 19.0, IBM, USA).

Analysis of SmWRKY expression in response to yeast extract and Ag⁺ treatment
RNA-seq data for S. miltiorrhiza hairy roots treated with yeast extract (100 μg/ml) and Ag⁺ (30 μM) were downloaded from GenBank under the accession number SRR924662 [12]. RNA-seq reads from non-treated (0 hpi) and treated for 12 h (12 hpi), 24 h (24 hpi) and 36 h (36 hpi) were mapped to SmWRKYs using the SOAP2 software [87] and analyzed as described previously [107]. The parameter v cutoff of 3 and parameter r cutoff of 2 were applied. SmWRKYs with the RPKM value greater than 2 were analyzed for differential expression using Fisher’s exact test. P < 0.05 was considered as differentially expressed.

Ka and Ks calculation
Paralogous SmWRKY genes were inferred from phylogenetic analysis. Non-synonymous (Ka) and synonymous (Ks) substitution of each paralogous gene pair were also determined by PAL2NAL program (http://www.bork.embL.de/pal2nal/) [108], which is based on the codon model program in PAML [68].

Tests of positive selection
To determine whether the WRKY gene family exhibited evidence of positive selection under the site model and
branch-site model [71], we applied the codeml program in PAML v4.8 to test the hypothesis of positive selection. An unrooted phylogenetic tree of SmWRKYs was reconstructed using the maximum likelihood method. In the site model, M0 (one ratio), M1a (neutral), M2a (selection), M3 (discrete), M7 (beta) and M8 (beta + ω > 1) were applied to the alignments, and we detected variation in the ω parameter among sites using the LRT for M1a vs. M2a, M0 vs. M3 and M7 vs. M8. Branch-site model [72] was used to compare the non-synonymous/synonymous substitution rate ratio (Ka/Ks) between clades or sequences. The ratio of nonsynonymous-to-synonymous for each branch under model A was calculated. Posterior probabilities (Qks) were calculated using the BEB method [68].

Estimation of functional divergence
The software DIVERGE2 was used to detect the functional divergence among members of SmWRKY subgroups [74]. The method is based on maximum likelihood procedures to estimate significant changes in the site-specific shift. The coefficients of Type-I and Type-II functional divergence (θt and θd) between two clusters were calculated. The coefficients of Type-I and Type-II functional divergence (θt and θd) greater than 0 indicates that site specific altered selective constraints or a radical shift of amino acid physiochemical property occurred after gene duplication and/or speciation [74]. Large posterior probability (Qk) indicates a high possibility that the functional constraint (or the evolutionary rate) and/or the radical change in the amino acid property of a site is different between two clusters [74].

Availability of supporting data
SmWRKY sequences supporting the results of this article are available in GenBank under accession numbers KM823124–KM823184. The other supporting data are included within the article and its additional files.

Additional file 6: Table S6. Ka/Ks and divergence analysis of paralogous SmWRKY genes. Ko, Ks and Ka/Ks are shown.

Additional file 7: Table S7. Primers used for qRT-PCR analysis of SmWRKY genes. Complete set of primers used for qRT-PCR of SmWRKY genes.

Abbreviations
AS: Anthranilate synthase; BEB: Bayes empirical bayes; CAD1: (+)-δ-cadinene synthase; CDS: Coding sequence; Ka: Non-synonymous; Ks: Synonymous; DXS: 1-deoxy-D-xylulose 5-phosphate synthase; GST: Glandular secretory trichome; LRT: Likelihood ratio test; MeJA: Methyl jasmonate; Mw: Molecular weight; π: Isoelectric points; SLS: Secologanin synthase; TCM: Traditional Chinese medicine; TDC: Tryptophan decarboxylase.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
CL contributed to RNA extraction, coding sequence (CDS) cloning, qRT-PCR and bioinformatics analysis, and participated in writing the manuscript. DL analyzed the RNA-seq data. FS prepared the plantlets treated with methyl jasmonate (MeJA). SL designed the experiment, participant in bioinformatics analysis, and wrote the manuscript. All authors have read and approved the version of manuscript.

Acknowledgements
We thank Prof. Shilin Chen and the sequencing group in our institute for kindly providing the S. miltiorrhiza genome sequence. We appreciate Prof. Xian’en Li for providing S. miltiorrhiza plants. This work was supported by grants from the Natural Science Foundation of China (grant no. 31370327), the Beijing Natural Science Foundation (grant nos. 5112026 and 5152021), the Major Scientific and Technological Special Project for Significant New Drugs Creation (grant no. 2012ZX09301002-001-031), the Program for Changjiang Scholars and Innovative Research Team in University (grant no. IRT11150), and the Research Fund for the Doctoral Program of Higher Education of China (grant no. 20111106110033).

Received: 26 October 2014 Accepted: 27 February 2015
Published online: 17 March 2015

References
1. Wu WY, Wang YP. Pharmacological actions and therapeutic applications of Salvia miltiorrhiza depside salt and its active components. Acta Pharmacol Sin. 2011;33:1119–30.
2. Wang KH, Morris-Natschke SL, Lee KH. Developments in the chemistry and biology of the bioactive constituents of Tanshen. Med Res Rev. 2007;27:133–48.
3. Ma Y, Yuan L, Wu B, Li X, Chen S, Lu S. Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in Salvia miltiorrhiza. J Exp Bot. 2012;63:2809–23.
4. Hao X, Shi M, Cui L, Xu C, Zhang Y, Kai G. Effects of methyl jasmonate and salicylic acid on tanshinone production and biosynthetic gene expression in transgenic Salvia miltiorrhiza hairy roots. Biotechnol Appl Biochem. 2014. doi:10.1002/bab.1236.
5. Zhang Y, Yan YP, Wu YC, Hua WP, Chen C, Ge Q, et al. Pathway engineering for phenolic acid accumulations in Salvia miltiorrhiza by combinatorial genetic manipulation. Metab Eng. 2014;21:71–80.
6. Shi M, Luo X, Ju G, Yu X, Hao X, Huang Q, et al. Increased accumulation of the cardio-cerebrovascular disease treatment drug tanshinone in Salvia miltiorrhiza hairy roots by the enzymes 3-hydroxy-3-methylglutaryl CoA reductase and 1-deoxy-D-xylulose 5-phosphate reductoisomerase. Funct Integr Genomics. 2014;14:603–15.
7. Di P, Zhang L, Chen J, Tan H, Xiao Y, Dong X, et al. 13C tracer reveals phenolic acids biosynthesis in hairy root cultures of Salvia miltiorrhiza. ACS Chem Biol. 2013;8:1537–48.
8. Guo J, Zhou YJ, Hillwig ML, Shen Y, Yang L, Wang Y, et al. CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. Proc Natl Acad Sci U S A. 2013;110:12108–13.
9. Hao G, Shi R, Tao R, Fang Q, Jiang X, Ji H, et al. Cloning, molecular characterization and functional analysis of 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase (HDR) gene for diterpenoid tanshinone biosynthesis in Salvia miltiorrhiza Bge. f. alba. Plant Physiol Biochem. 2013;70:31–2.

10. Gao W, Hilwig ML, Huang L, Cui G, Wang X, Kong J, et al. A functional genomics approach to tanshinone biosynthesis provides stelereochemical insights. Org Lett. 2009;11:1510–3.

11. Luo H, Zhu Y, Song G, 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.

12. Gao W, Sun HX, Xiao H, Cui G, Hilwig ML, Jackson A, et al. Combining metabolomics and transcriptomics to characterize tanshinone biosynthesis in Salvia miltiorrhiza. BMC Genomics. 2014;15:73.

13. Yang L, Ding G, Lin H, Cheng H, Kong Y, Wei Y, et al. Transcriptome analysis of medicinal plant Salvia miltiorrhiza and identification of genes related to tanshinone biosynthesis. PLoS ONE. 2013;8:e80466.

14. Hou X, Shao F, Ma Y, Lu S. The phenylalanine ammonia-lyase gene family in Salvia miltiorrhiza: genome-wide characterization, molecular cloning and expression analysis. Mol Biol Rep. 2013;40:4301–10.

15. Wenping H, Yuan Z, Jie S, Lijun Z, Zhezhi W. De novo transcriptome sequencing in Salvia miltiorrhiza to identify genes involved in the biosynthesis of active ingredients. Genomics. 2011;98:272–9.

16. Yan Y, Wang Z, Tian W, Dong Z, Spencer DF. Generation and analysis of expressed sequence tags from the medicinal plant Salvia miltiorrhiza. Sci China Life Sci. 2010;53:838–43.

17. Li Y, Sun C, Luo HM, Li XW, Gao M, Li C, et al. Cloning, functional characterization and identification of WRKY family members in Salvia miltiorrhiza. J Integr Plant Biol. 2014;56:38–50.

18. 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 Genomics. 2014;15:277.

19. Ishiguro SN, Nakamura K. Characterization of a cDNA encoding a novel WRKY gene, WsWRKY1, from the ornamental plant, Wisteria sinensis (Fabaceae). J Plant Res. 2010;123:279–85.

20. Song JY, Luo HM, Li CF, Sun C, Xu J, Chen SL. Genome-wide analysis and molecular dissection of the SPL gene family in Salvia miltiorrhiza. J Integr Plant Biol. 2014;56:38–50.

21. 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 Genomics. 2014;15:277.

22. Zhang Y, Wang L. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. BMC Evol Biol. 2005;5:1.

23. Liu JJ, Ekramoddoullah AKM. Identification and characterization of the WRKY transcription factor family in Pinus monticola Genome. 2005;52:77–88.

24. Wang Q, Wang M, Zhang X, Hao B, Kaushik SK, Pan Y, WRKY gene family evolution in Arabidopsis thaliana. Genetica. 2011;139:973–83.

25. Hara K, Yagi M, Kusano T, Sano H. Rapid systemic accumulation of transcripts encoding a tobacco WRKY transcription factor upon wounding. Mol Gen Genet. 2002;265:30–7.

26. Wang Z, Yang P, Fan B, Chen Z. An oligo selection procedure for identification of sequence-specific DNA-binding activities associated with the plant defence response. Plant J. 1998;16:515–22.

27. Yoda H, Ogawa M, Yamaguchi Y, Koizumi N, Kusano T, Sano H. Identification of early-responsive genes associated with the hypersensitive response to tobacco mosaic virus and characterization of a WRKY-type transcription factor in tobacco plants. Mol Biol Genet. 2002;17:254–61.

28. Ryu HS, Han M, Lee SK, Cho JI, Ryu N, Heu S, et al. A comprehensive expression analysis of the WRKY gene superfamily in rice plants during defense response. Plant Cell Rep. 2006;25:836–47.

29. Wu KL, Guo ZJ, Wang HH, Li J. The WRKY family of transcription factors in rice and Arabidopsis and their origins. DNA Res. 2005;12:99–25.

30. Yin G, Xu H, Xiao S, Qin Y, Li Y, Yan Y, et al. The large soybean (Glycine max) WRKY TF family expanded by segmental duplication events and subsequent divergent selection among subgroups. BMC Plant Biol. 2013;13:148.

31. Wei KF, Chen J, Chen YF, Wu LJ, Xie DX. Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. DNA Res. 2012;19:153–64.

32. Mangelescu E, Killian J, Berendzen KW, Kolukisaoglu UH, Harter K, Jansson C, et al. Phylogenetic and comparative genetic expression analysis of barley (Hordeum vulgare) WRKY transcription factor family reveals putatively retained functions between monocots and dicots. BMC Genomics. 2008;9:194.

33. Wang M, Vannozzi A, Wang G, Liang Y, Tornielli BG, Zenoni S, et al. Genome and transcriptome analysis of the grapevine (Vitis vinifera L.) WRKY gene family. Hortic Res. 2014;1:61.

34. Guo C, Guo R, Xu X, Gao M, Li K, Song J, et al. Evolution and expression analysis of the grape (Vitis vinifera L.) WRKY gene family. J Exp Bot. 2014;65:1513–28.

35. He D, Dong Q, Shao Y, Jiang H, Zhu S, Cheng B, et al. Genome-wide sequence and characterization of the WRKY gene family in Populus trichocarpa. Plant Cell Rep. 2012;31:1199–217.

36. Huang S, Gao Y, Liu J, Peng X, Niu X, Fei Z, et al. Genome-wide analysis of WRKY transcription factors in Solanum lycopersicum. Mol Genet Genomics. 2012;287:495–513.

37. Ling J, Jiang WJ, Zhang Y, Yu HJ, Mao ZC, Gu XF, et al. Genome-wide analysis of WRKY gene family in Cucumis sativus. BMC Genomics. 2011;12:471.

38. Ramiro D, Jalalou A, Pettot AS, Grossi De Sá MF, Maluf MF, Fernandez D. Identification of coffee WRKY transcription factor genes and expression profiling in response to pathogens. Tree Genet Genomes. 2010;6:767–81.

39. Rushton PJ, Somssich IE, Ringler P, Shen QJ. WRKY transcription factors. Trends Plant Sci. 2005;10:247–58.

40. Edulgem T, Rushton PJ, Bobatek S, Somssich IE. The WRKY superfamily of plant transcription factors. Trends Plant Sci. 2005;10:199–206.

41. Dong J, Chen C, Chen Z. Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. Plant Mol Biol. 2003;51:121–37.

42. Xu X, Chen C, Fan B, Chen Z. Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. Plant Cell. 2006;18:1310–26.

43. Li J, Brader G, Karolka T, Palva T. WRKY70 modulates the selection of signaling pathways in plant defense. Plant J. 2006;46:477–91.

44. Oh SK, Yi SY, Yu SH, Moon JS, Park JM, Choi D. CaWRKY2, a chili pepper transcription factor, is rapidly induced by incompatible plant pathogens. Mol Cells. 2006;22:58–64.

45. Zheng Z, Mosher SL, Fan B, Klessig DF, Chen Z. Functional analysis of Arabidopsis WRKY25 transcription factor in plant defense against Pseudomonas syringae. BMC Plant Biol. 2007;7:2.
59. Ciolkoñski J, Wanke D, Birkenbihl RP, Somischi IE. Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. Plant Mol Biol. 2006;68:881–92.
60. Brand LH, Kirchler T, Hummel S, Chaban C, Wanke D. DPELISA: a fast and versatile method to specify the binding of plant transcription factors to DNA in vitro. Plant Methods. 2010;6:25.
61. Sun C, Palnqvist S, Olsson H, Boren M, Ahländberg S, Jansson C. A novel WRKY transcription factor, SUS112A2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. Plant Cell. 2003;15:706–92.
62. Shao F, Lu S. Genome-wide identification, molecular cloning, expression profiling and posttranscriptional regulation analysis of the Argonate gene family in Salvia miltiorrhiza, an emerging medicinal model plant. BMC Genomics. 2013;14:512.
63. Shao F, Lu S. Identification, molecular cloning and expression analysis of five RNA-dependent RNA polymerase genes in Salvia miltiorrhiza. PLoS One. 2014;9:e95117.
64. Lynch M, O’Hely M, Walsh B, Force A. The probability of preservation of a newly arisen gene duplicate. Genetics. 2001;159:789–804.
65. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290:1151–5.
66. Park CY, Lee HJ, Yoo JH, Moon BC, Choi MS, Kang YH, et al. WRKY group IIId transcription factors interact with calmodulin. FEBS Lett. 2005;579:154–50.
67. Xie Z, Zhang ZL, Zou X, Huang J, Rua P, Thompson D, et al. Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant Physiol. 2005;137:176–89.
68. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1386–91.
69. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HV-1 envelope gene. Genetics. 1998;148:929–36.
70. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290:1151–5.
71. Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for detecting positively selected amino acid sites under positive selection. Mol Biol Evol. 2005;22:1107–22.
72. Zhang J, Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites under positive selection. Mol Biol Evol. 2005;22:1107–22.
73. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inferences of amino acid substitution rates and applications to the HIV-1 envelope gene. Genetics. 1998;148:929–36.
74. Gu X. A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. Mol Biol Evol. 2006;23:1937–46.
75. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inferences of amino acid substitution rates and applications to the HIV-1 envelope gene. Genetics. 1998;148:929–36.
76. Park CY, Lee HJ, Yoo JH, Moon BC, Choi MS, Kang YH, et al. WRKY group IIId transcription factors interact with calmodulin. FEBS Lett. 2005;579:154–50.
77. Xie Z, Zhang ZL, Zou X, Huang J, Rua P, Thompson D, et al. Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant Physiol. 2005;137:176–89.
78. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1386–91.
79. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HV-1 envelope gene. Genetics. 1998;148:929–36.
80. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290:1151–5.
81. Park CY, Lee HJ, Yoo JH, Moon BC, Choi MS, Kang YH, et al. WRKY group IIId transcription factors interact with calmodulin. FEBS Lett. 2005;579:154–50.
82. Xie Z, Zhang ZL, Zou X, Huang J, Rua P, Thompson D, et al. Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant Physiol. 2005;137:176–89.
83. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1386–91.
84. Lynch M, O’Hely M, Walsh B, Force A. The probability of preservation of a newly arisen gene duplicate. Genetics. 2001;159:789–804.
85. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290:1151–5.
86. Park CY, Lee HJ, Yoo JH, Moon BC, Choi MS, Kang YH, et al. WRKY group IIId transcription factors interact with calmodulin. FEBS Lett. 2005;579:154–50.
87. Xie Z, Zhang ZL, Zou X, Huang J, Rua P, Thompson D, et al. Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant Physiol. 2005;137:176–89.
88. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1386–91.
89. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HV-1 envelope gene. Genetics. 1998;148:929–36.
90. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290:1151–5.
91. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inferences of amino acid substitution rates and applications to the HIV-1 envelope gene. Genetics. 1998;148:929–36.
106. Willems E, Leyns L, Vandesompele J. Standardization of real-time PCR gene expression data from independent biological replicates. Anal Biochem. 2008;379:127–9.

107. Li D, Shao F, Lu S. Identification and characterization of mRNA-like noncoding RNAs in Salvia miltiorrhiza. Planta. 2015. doi:10.1007/s00425-015-2246-z.

108. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34:W609–12.