

# SCIENTIFIC REPORTS



OPEN

## Systematic analysis of *DEMETER*-like DNA glycosylase genes shows lineage-specific Smi-miR7972 involved in *SmDML1* regulation in *Salvia miltiorrhiza*

Jiang Li, Caili Li & Shanfa Lu 

DEMETER-like DNA glycosylases (DMLs) initiate the base excision repair-dependent DNA demethylation to regulate a wide range of biological processes in plants. Six putative *SmDML* genes, termed *SmDML1–SmDML6*, were identified from the genome of *S. miltiorrhiza*, an emerging model plant for Traditional Chinese Medicine (TCM) studies. Integrated analysis of gene structures, sequence features, conserved domains and motifs, phylogenetic analysis and differential expression showed the conservation and divergence of *SmDMLs*. *SmDML1*, *SmDML2* and *SmDML4* were significantly down-regulated by the treatment of 5Aza-dC, a general DNA methylation inhibitor, suggesting involvement of *SmDMLs* in genome DNA methylation change. *SmDML1* was predicted and experimentally validated to be target of Smi-miR7972. Computational analysis of forty whole genome sequences and almost all of RNA-seq data from Lamiids revealed that *MIR7972s* were only distributed in some plants of the three orders, including Lamiales, Solanales and Boraginales, and the number of *MIR7972* genes varied among species. It suggests that *MIR7972* genes underwent expansion and loss during the evolution of some Lamiids species. Phylogenetic analysis of *MIR7972s* showed closer evolutionary relationships between *MIR7972s* in Boraginales and Solanales in comparison with Lamiales. These results provide a valuable resource for elucidating DNA demethylation mechanism in *S. miltiorrhiza*.

*Salvia*, widely distributed in the world, is the largest genus in the Labiatae family. It includes about 900 species, of which many have significant economical and medicinal value. *S. miltiorrhiza* Bunge is a well-known *Salvia* species widely used in traditional Chinese medicine (TCM) for the treatment of dysmenorrhoea, amenorrhoea and cardiovascular diseases<sup>1–4</sup>. It is also one of the best selling TCM materials with long usage history and the first Chinese herb entering international market. Recently, *S. miltiorrhiza* is emerging as a model system for medicinal plant biology<sup>5</sup>. The whole genomes of two *S. miltiorrhiza* lines have been decoded<sup>6,7</sup>. A huge amount of RNA-seq-based transcriptome data have been obtained and are available for comparative analysis (<https://www.ncbi.nlm.nih.gov/sra>). It provides useful information for further elucidating the genetic and epigenetic regulatory mechanisms of *S. miltiorrhiza* development and bioactive compound production.

Epigenetic regulation is an important regulatory mechanism affecting many plant cellular processes. It is known that epigenetic phenotypes are caused by changes in a chromosome without alterations in the DNA sequence, including individual or combined changes in DNA methylation and histone modification, as well as the action of chromatin-remodeling factors and noncoding RNAs<sup>8</sup>. DNA methylation is an important epigenetic modification. It is involved in multiple biological processes, such as transposon silencing, genomic imprinting, and X-chromosome inactivation<sup>9–13</sup>. In mammals, DNA methylation predominantly occurs in CG sequence contexts<sup>14</sup>; however, in plants, three sequence contexts, including CG, CHG and CHH (H represents either A, T or G), are major DNA methylation targets<sup>15,16</sup>. DNA methylation levels and patterns in plants are dynamically regulated by DNA methyltransferases and demethylases. DNA METHYLTRANSFERASE 1 (MET1) and

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, No.151 Malianwa North Road, Haidian District, Beijing, 100193, China. Correspondence and requests for materials should be addressed to S.L. (email: [sflu@implad.ac.cn](mailto:sflu@implad.ac.cn))

CHROMOMETHYLASE 3 (CMT3) maintain symmetric CG<sup>17</sup> and CHG methylation<sup>18</sup>, respectively. Methylation of asymmetric CHH contexts must be *de novo* created by DOMAINS REARRANGED METHYLTRANSFERASE (DRM) during cell replication<sup>19</sup>. DNA demethylation can be regulated through passive or active processes. Passive demethylation occurs in newly synthesized DNA strand and is caused by dysfunction of DNA methyltransferase, whereas active demethylation is an outcome of replacement of methylated cytosine with non-methylated cytosine under the catalysis of DEMETER-like DNA glycosylases (DMLs)<sup>20,21</sup>.

The DMLs act as bifunctional glycosylase/AP-lyase, which removes 5-methylcytosine (5 mC) followed by cleaving the abasic site<sup>8</sup>. They contain three essential domains, including a DNA glycosylase domain and two additional conserved domains, termed domain A and domain B<sup>22</sup>. The DNA glycosylase domain has a helix-hairpin-helix (HhH) motif, a glycine/proline-rich region followed by a conserved aspartic acid (GPD), and a [4Fe-4S] cluster motif<sup>22</sup>. The [4Fe-4S] cluster motif consists of four cysteine residues functioning to hold a [4Fe-4S] cluster and is required for 5 mC excision<sup>22</sup>. Amino acids in domain A are required for nonspecific DNA binding and all three domains are necessary and sufficient for 5 mC excision<sup>22</sup>.

DMLs play significant roles in many developmental and biological processes, such as reproduction, seed development, and plant response to biotic and abiotic stresses<sup>20,23–27</sup>. *Arabidopsis* contains four DML gene family members, known as *AtDME*, *AtROS1*, *AtDML2* and *AtDML3*, respectively. *AtDME* is necessary for seed viability and is a core regulator of imprinted genes<sup>28</sup>. It is preferentially expressed in the central cell before fertilization, ensuring maternal expression of the imprinted genes *MEA* and *FWA*<sup>29,30</sup>. *AtROS1*, *AtDML2* and *AtDML3* are broadly expressed in vegetative tissues to maintain a proper genome methylation pattern and to regulate relevant genes and transposons<sup>31</sup>. *AtROS1* counteracts RNA-directed DNA Methylation (RdDM) pathway for dynamic transcriptional regulation<sup>32</sup>, whereas *AtDML2* and *AtDML3* are involved in removing improper methylated cytosines and maintaining methylation levels of certain targeted sites to make sure an appropriate distribution of genome methylation<sup>31</sup>. A total of six DML genes exist in the rice genome. It includes *OsROS1a*, *OsROS1b*, *OsROS1c*, *OsROS1d*, *OsDML3a* and *OsDML3b*, of which four are *AtROS1* orthologs, whereas the other two are *AtDML3* orthologs<sup>24</sup>. The function of *OsROS1a* is analogous, to certain extent, to that of *AtDME* in both male and female gametophytes<sup>33</sup>. *OsROS1c* is responsible for the demethylation of retrotransposon *Tos17* and plays a critical role in promoting the transposition of *Tos17*<sup>21</sup>. Little is known about the function of other rice DMLs and DMLs in other plants.

Based on current knowledge of epigenetics, we presume that epigenetic regulation is involved in secondary metabolism and Dao-di herb formation, two important issues in medicinal plant biology. With the long-term goal to test this hypothesis and to elucidate the underlying mechanisms, we analyze and report here the DML gene family in *S. miltiorrhiza*. Comparative analysis showed that *SmDMLs* were conserved with *Arabidopsis* and rice DMLs. They differentially expressed in various *S. miltiorrhiza* tissues and were responsive to DNA methylation inhibitor (5-aza-2'-deoxycytidine, 5Aza-dC) treatments. The expression of *SmDML1* is posttranscriptionally regulated by *Smi-miR7972*. The existence of *MIR7972s* in plants was systematically investigated. *MIR7972* genes were identified in some species belonging to three orders, including Lamiales, Solanales, and Boraginales. The distribution patterns were scattered and the number of gene members was varied among species. The results provide the first hand of information for elucidating the role of epigenetic regulation in medicinal plants.

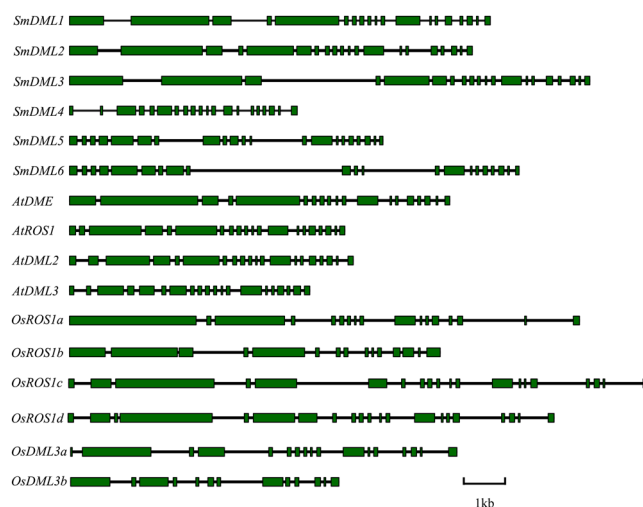
## Results

**Identification and comparative analysis of *SmDMLs*.** Since *S. miltiorrhiza* has significant medicinal value and is being developed rapidly as a model system for medicinal plant biology, two research groups sequenced the whole genomes of two different *S. miltiorrhiza* lines, one of which is known as 99-3<sup>6</sup>, whereas the name of the other one is unknown<sup>6</sup>. The draft genome assemblies of line 99-3 and the name-unknown line are about 559 and 641 Mb, respectively, although the estimated genome sizes are 615 and 645.78 Mb, respectively<sup>6,7</sup>. It suggests that both of the current *S. miltiorrhiza* genome assemblies are incomplete. In order to identify *S. miltiorrhiza* DEMETER-like DNA glycosylase gene family (*SmDMLs*), we searched the two genome assemblies using four *Arabidopsis* AtDML proteins as queries. The retrieved genomic DNA sequences were first computationally predicted for gene models on the GENSCAN web server<sup>34</sup> and then manually examined and corrected by comparison with DML genes identified from other plants ([www.ncbi.nlm.nih.gov/blast/](http://www.ncbi.nlm.nih.gov/blast/)) and by alignment with RNA-seq data of *S. miltiorrhiza* transcriptome (<http://www.ncbi.nlm.nih.gov/sra>). After correcting various genomic DNA sequence errors, we obtained six full-length *SmDML* genes from the genome assembly of the name-unknown *S. miltiorrhiza* line and four full-length *SmDML* genes and several *SmDML* gene fragments from the genome assembly of line 99-3. Amino acid sequence alignment showed that each of the four *SmDML* proteins from line 99-3 was identical to a *SmDML* protein from the name-unknown *S. miltiorrhiza* line. Sequence comparison of line 99-3 *SmDML* gene fragments, transcriptome of line 99-3 and the other two *SmDML* genes from the name-unknown *S. miltiorrhiza* line showed that line 99-3 *SmDML* gene fragments encoded partial *SmDML* proteins identical to the other two *SmDML* proteins from the name-unknown *S. miltiorrhiza* line. Taken together, we conclude that the *SmDML* gene family includes six members, which are designated as *SmDML1–SmDML6*, respectively (Table 1). The number of *SmDMLs* is more than *AtDMLs*, whereas it is same as *OsDMLs*<sup>24</sup>.

*SmDML* genes contain 18–20 introns (Table 1, Fig. 1). Although similarities of gene structures exist among the six *SmDMLs*, it is particularly high among *SmDML1*, *SmDML2* and *SmDML3* and between *SmDML5* and *SmDML6*. In addition, the structures of *SmDML1*, *SmDML2* and *SmDML3* are highly similar to four *AtDMLs*, *OsROS1a–OsROS1d* and *OsDML3a* (Fig. 1). The deduced *SmDML* proteins have amino acid number varying from 720 (*SmDML4*) to 2030 (*SmDML3*), isoelectric point (pI) from 6.76 (*SmDML5*) to 9.27 (*SmDML4*), and molecular weight (Mw) from 80.5 kDa (*SmDML4*) to 226.5 (*SmDML3*) (Table 1). Wide ranges of gene length, pI and Mw also exist among *AtDMLs* and *OsDMLs* (Table 1). It suggests the conservation and divergence of sequence features of plant DMLs.

Gene name	Gene model	Gene length	No. of intron	ORF (bp)	Protein (aa)	MW (kDa)	pI
<i>SmDML1</i>	MG602215	9194	18	5865	1954	217.9	6.90
<i>SmDML2</i>	MG602216	8800	19	5598	1865	207.2	6.97
<i>SmDML3</i>	MG602217	11369	19	6093	2030	226.5	8.20
<i>SmDML4</i>	MG602218	4976	19	2163	720	80.5	9.27
<i>SmDML5</i>	MG602219	6859	20	3165	1054	118.4	6.76
<i>SmDML6</i>	MG602220	9820	20	3237	1078	121.8	7.37
<i>AtDME</i>	AT5G04560.2	8306	18	5964	1987	221.1	7.61
<i>AtROS1</i>	AT2G36490.1	6014	18	4182	1393	156.5	7.38
<i>AtDML2</i>	AT3G10010.1	6197	19	3999	1332	151.5	8.62
<i>AtDML3</i>	AT4G34060.1	5251	19	3135	1044	120.3	9.63
<i>OsROS1a</i>	LOC_Os01g11900.1	11143	16	5859	1952	215.9	6.77
<i>OsROS1b</i>	LOC_Os02g29230.1	8100	14	4911	1636	182.6	6.53
<i>OsROS1c</i>	LOC_Os05g37350.1	12675	19	5544	1847	205.9	6.61
<i>OsROS1d</i>	LOC_Os05g37410.1	10616	20	5490	1829	204.3	6.53
<i>OsDML3a</i>	LOC_Os02g29380.1	8438	16	3624	1207	134.4	6.14
<i>OsDML3b</i>	LOC_Os04g28860.1	5860	12	2883	960	108.4	8.94

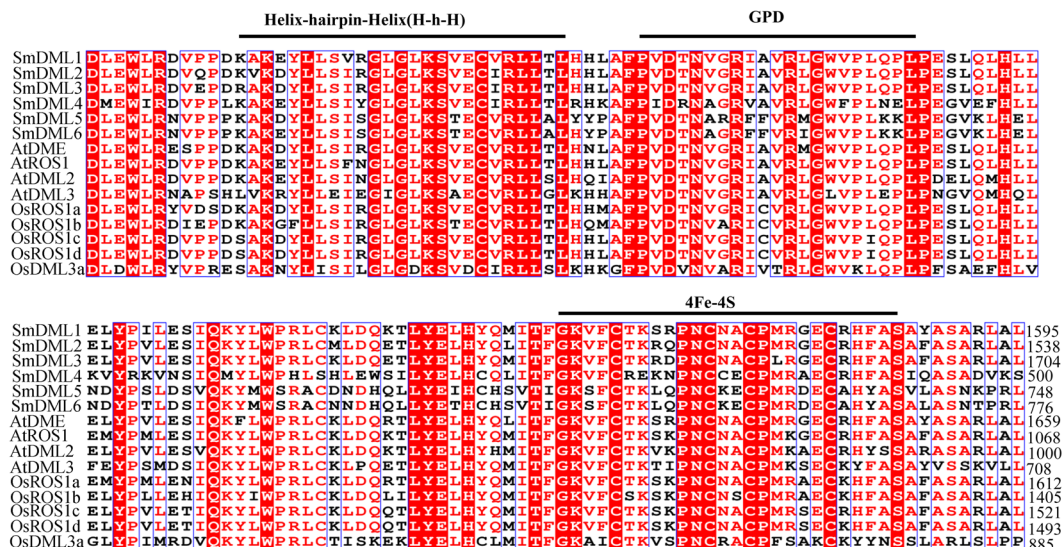
**Table 1.** Sequence features and intron numbers of *SmDMLs*, *AtDMLs* and *OsDMLs*.



**Figure 1.** Gene structures of *DMLs* in *S. miltiorrhiza*, *Arabidopsis* and rice. Exons are presented by filled green boxes. Introns are presented by lines.

**Conserved domains and motifs of DMLs.** It has been shown that DML proteins contain a DNA glycosylase domain and two additional conserved domains A and B<sup>22</sup>. Multiple sequence alignment of deduced amino acid sequences showed that all of the six *SmDMLs* also contain the three domains (Fig. 2, see Supplementary Figs S1–3). The DNA glycosylase domain is highly conserved among 15 analyzed DMLs, including six *SmDMLs*, four *AtDMLs* and five *OsDMLs*, and contains HhH, GPD and [4Fe-4S] cluster motifs (Fig. 2). Compared with other DMLs analyzed, *OsDML3b* has less conserved DNA glycosylase domain. Differing from the DNA glycosylase domain, the sequences of domains A and B are highly conserved among all 16 *S. miltiorrhiza*, *Arabidopsis* and rice DMLs (see Supplementary Figs S2,3). In addition to the amino acid sequence, the location of domains is conserved among all DMLs from *S. miltiorrhiza*, *Arabidopsis* and rice (see Supplementary Fig. S1). These results suggest that the majority of domains in DMLs from different plant species are deeply conserved. It is consistent with the critical function of these domains in 5 mC excision.

In order to further determine the conservation and divergence of plant DMLs, we searched conserved motifs in *SmDMLs*, *AtDMLs* and *OsDMLs* using the MEME suite (<http://meme.sdsc.edu/meme/meme.html>). A total of 15 conserved motifs were identified (Fig. 3). The length of motifs ranges from 18 to 150 amino acids (Table 2). The number of motifs in each DML varies between 7 and 15. Motif 1 is actually the DNA glycosylase domain. It exists in 15 of the 16 DMLs analyzed (Fig. 3). No motif 1 was detected in *OsDML3b*. It is consistent with the fact that the DNA glycosylase domain is less conserved in *OsDML3b* compared with that in other DMLs analyzed. Motifs 2 and 9 are located in domain A. Motif 2 exists in all 16 DMLs analyzed, whereas motif 9 was found in 13 of the 16 DMLs. No motif 9 was detected in *SmDML4*, *OsDML3a* and *OsDML3b*. It suggests that motif 2 is more conserved than motif 9, although both of them are located in the conserved domain A. Motifs 3, 4, 5, 7, 8 and



**Figure 2.** Amino acid sequence alignment of the conserved DNA glycosylase domain of DML proteins in *S. miltiorrhiza*, *Arabidopsis* and rice. Numbers show the position of amino acids. Identical amino acids are highlighted in red. Similar amino acid residues are showed in red.



**Figure 3.** Distribution of conserved motifs of DML proteins from *S. miltiorrhiza*, *Arabidopsis* and rice.

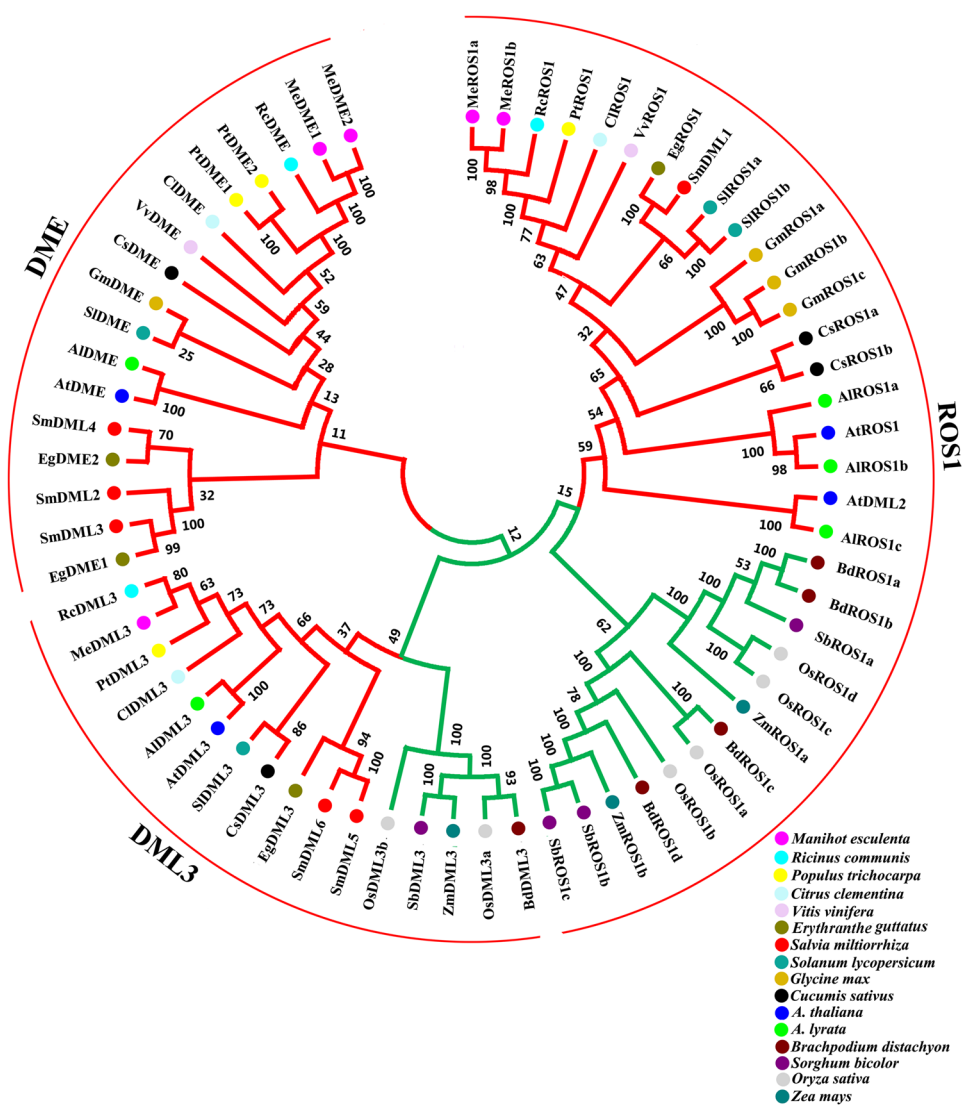
11 are located in domain B. Among them, motifs 3, 4, 5, 7 and 8 existing in 14 or 15 DMLs show relatively high conservation, whereas motif 11, which was detected in 11 of 16 DMLs, is the least conserved. In addition to the nine motifs found in conserved domains, other six, including motifs 6, 10, 12, 13, 14 and 15, are located in less conserved regions. Among them, motifs 6, 10 and 13 are highly conserved, whereas motifs 12, 14 and 15 are specific to OsROS1c and OsROS1d. Motifs commonly existing in DMLs are probably associated with the conserved functions of DMLs, but those specific to a few DMLs seem to be related to gene-specific functions.

**Phylogenetic analysis of DML proteins.** In order to determine the evolutionary relationship among DMLs, an unrooted neighbor-joining tree was constructed using 66 full length protein sequences from 16 plant species. DML proteins were clustered into three orthology groups, including the DME group, the ROS1 group, and the DML3 group (Fig. 4). The ROS1 group is the largest group. It contains 33 DML proteins. Based on the phylogenetic tree constructed, the ROS1 group may be further divided into the monocot subgroup and the dicot subgroup. Similar to the ROS1 group, the DML3 group may also be divided into the monocot subgroup and the dicot subgroup. DME group is only restricted to dicots. It indicates that DME may be phylogenetically monophyletic in dicots. These results are consistent with previous phylogenetic tree constructed using conserved DNA glycosylase domains of DMLs from flowering plants<sup>24</sup>. It suggests the conservation and divergence of DML proteins in monocots and dicots.

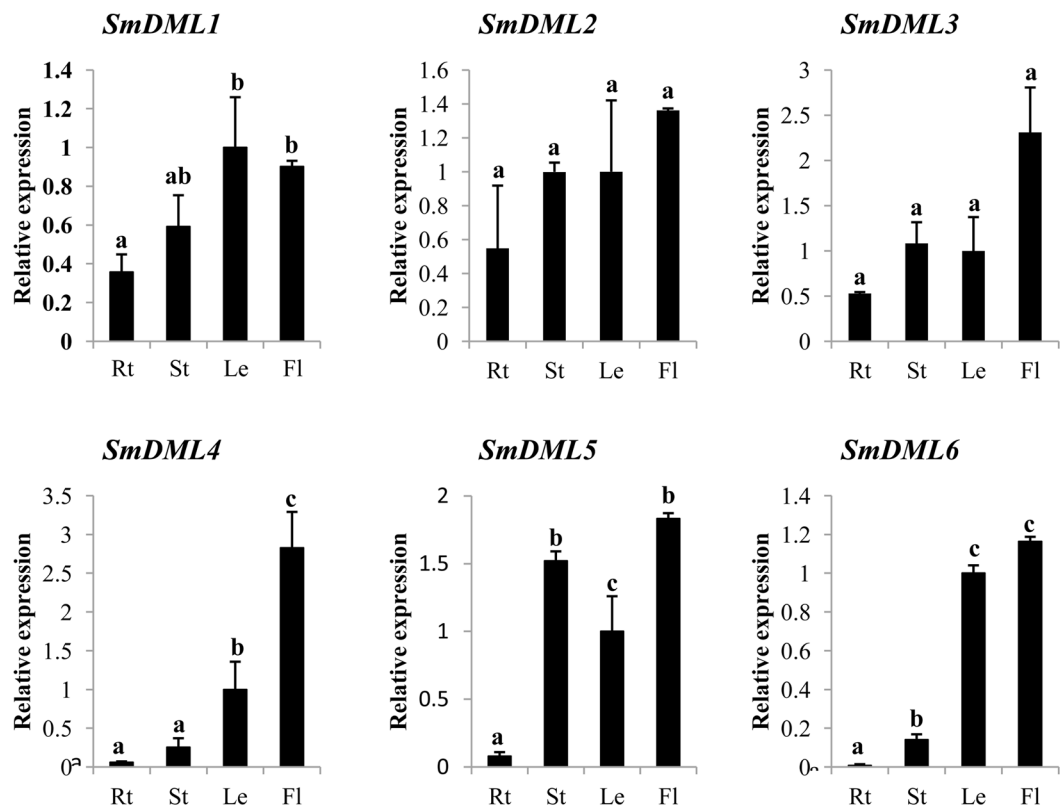
All of SmDMLs showed close phylogenetic relationships with DMLs from monkey flower (*Erythranthe guttatus*) (Fig. 4). It is consistent with the fact that both *S. miltiorrhiza* and *E. guttatus* are members of Lamiales. Based on gene structures, conserved motifs and phylogenetic relationships, SmDML2 and SmDML5 are closely related

Motif	Length (aa)	Consensus sequence
1	150	FLNRLVKDHGSIDLEWLRDVPDPKAKDYLLSIRGLGLKSVCEVRLTLHHLAFPVDTNVGRIAVRLGWVPLQPLPESLQLHLELYPVLESIQKYLWPRL
2	63	ERRVFRGRADSFARMHLVQGDRRFSPWKGSVVDSVVGVLFTQNVSDHLSSAFMSLAAKFPV
3	43	RGTLIPCRTAMRGSFPLNGTYFQVNEVFADHESLNPIDVPR
4	29	QYCFWRGFVFCVRFDRKTRAPRPLVARLH
5	33	ASIPTPKLKNVSRRLTEHQVYELPDSHPLLEGF
6	41	ERTEDTMSVDWEAVRCADVKEIADTIKERGMNMLAERIK
7	21	DKREPDDPCPYLLAIWTPGET
8	21	WNLPRRTVYFGTSVPSIFKGL
9	28	VKKKKPRPKVDLDPETTRVWNLMLGKDA
10	29	EAFYEDPDEIPTIKLNMEEFTQNLKSYIQ
11	21	LCSNETCFSCNSIREAQSQTV
12	150	YIKFMTKRSQKARLNSPNSIQPNIDQKNRFSSETIFSGGFNGLKRSEETFQKTLQPDPDKRINLDIHCEVPVENSNTSTPPYMDYLQGVTSKFRYFD
13	18	NCEPIIEEPASPEPEIEE
14	150	MACEKIHMEPKGNATVNELTKGENYSLHCKEPAGSLCDHETEIDHKAKSISDFSAVELTACMKNLHATQFQKEISLSQSVVTSSELQPLSSGMDHA
15	150	HRPKVIREDRPAKKQMATTSEEKPLNQPKRKYVWKNRNPSSLEKCAEPFSDHSISRESRTTVRSSIASVRRRLQFEFGEHGVQRDQSSRTNSWYRNQEK

**Table 2.** Consensus sequences of 15 motifs identified in SmDMLs, AtDMLs and OsDMLs.



**Figure 4.** Phylogenetic analysis of DML proteins from 16 plant species. Monocot and dicot are colored green and red, respectively. The bootstrap values are shown.

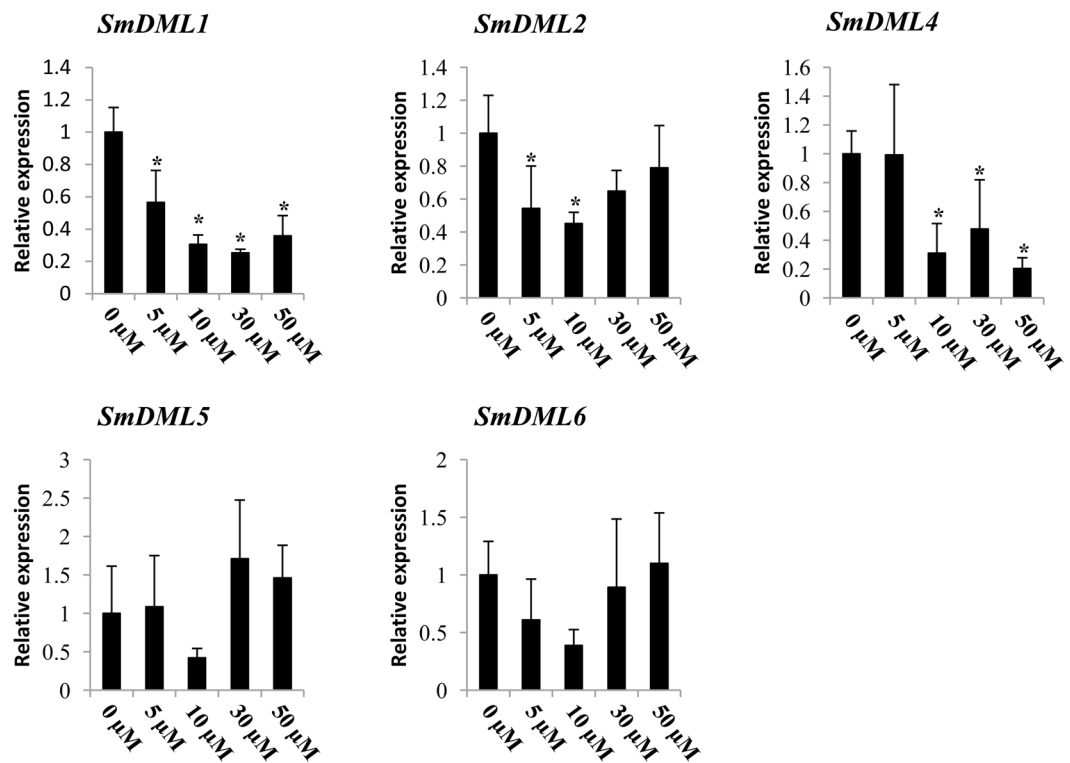


**Figure 5.** Expression patterns of *SmDMLs* in roots (Rt), stems (St), leaves (Le) and flowers (Fl) of *S. miltiorrhiza*. The expression levels were analyzed using the quantitative RT-PCR method. Fold changes of *SmDML* expression are shown. Expression level in leaves was arbitrarily set to 1 and the levels in other organs were given relative to this. One-way ANOVA was calculated using IBM SPSS 20 software.  $P < 0.05$  was considered statistically significant and was represented by different letters. Error bars were indicated by the standard deviations of three biological replicates.

with *SmDML3* and *SmDML6*, respectively. *SmDML2* and *SmDML3* share 62.88% protein sequence identity and 66.16% cDNA similarity. Protein sequence identity and cDNA similarity between *SmDML5* and *SmDML6* are 89.58% and 93.17%, respectively. It implies duplication events occurred during *SmDMLs* evolution. To test the phylogenetic selection pressure on these genes after duplication, we analyzed the substitution rate ratios of non-synonymous (Ka) versus synonymous (Ks) mutations (Ka/Ks)<sup>35</sup>. Generally, Ka/Ks = 1 suggests that the genes are pseudogenes with neutral selection, whereas less than 1 implies purifying or stabilizing selection<sup>35</sup>. The calculated Ka/Ks values between *SmDML2* and *SmDML3* and between *SmDML5* and *SmDML6* are 0.3833 and 0.9799, respectively. It indicated that *SmDML2* and *SmDML3* paralogous pairs experienced strong purifying selection, whereas *SmDML5* and *SmDML6* paralogous pairs could be pseudogenes with neutral selection. These results suggested that specific *SmDMLs* experienced distinct phylogenetic selection pressure.

**Differential expression of *SmDMLs* in *S. miltiorrhiza*.** As the main participants of DNA demethylation, DMLs play important roles in plant growth and development<sup>36,37</sup>. To preliminarily explore the biological function of *SmDMLs*, the expression of six *SmDMLs* in flowers, leaves, roots and stems of 2-year-old, field nursery-grown *S. miltiorrhiza* plants was analyzed using quantitative RT-PCR technology. Although the transcripts of all six *SmDMLs* were detected in the tissues analyzed, significant differential expression patterns were observed (Fig. 5). *SmDML1* showed the highest expression in leaves and flowers. Its expression in roots was the lowest among the four tissues analyzed. The expression patterns of *SmDML1* are different from its *Arabidopsis* counterpart, *AtROS1*, which showed high expression in stems and roots and low in flowers<sup>31</sup>. *SmDML2* and *SmDML3* are paralogs closely related (Fig. 4). Consistently, they showed similar expression patterns with the highest in flowers, followed by stems, leaves and roots. *SmDML4* was predominantly expressed in flowers. Although *SmDML5* and *SmDML6* are probably pseudogenes based on Ka/Ks, they were expressed. *SmDML5* showed the highest expression in flowers and stems, followed by leaves, and the lowest in roots, whereas *SmDML6* was expressed mainly in flowers and leaves (Fig. 5). Expressed pseudogenes have been reported in other plants, such as rice<sup>35</sup> and barley<sup>38</sup>.

To further analyze the expression patterns of *SmDMLs*, we analyzed RNA-seq data generated for periderm, phloem and xylem of *S. miltiorrhiza* roots. *SmDML1* exhibited the highest expression in xylem, followed by phloem and periderm. The expression of *SmDML2* and *SmDML3* paralogs showed similar patterns with the highest in phloem, followed by xylem and periderm. However, the differential expression for each of them in the three tissues of *S. miltiorrhiza* roots was not significant as analyzed using TopHat2.0.12 and Cufflinks2.2.1<sup>39</sup> (Table S3).

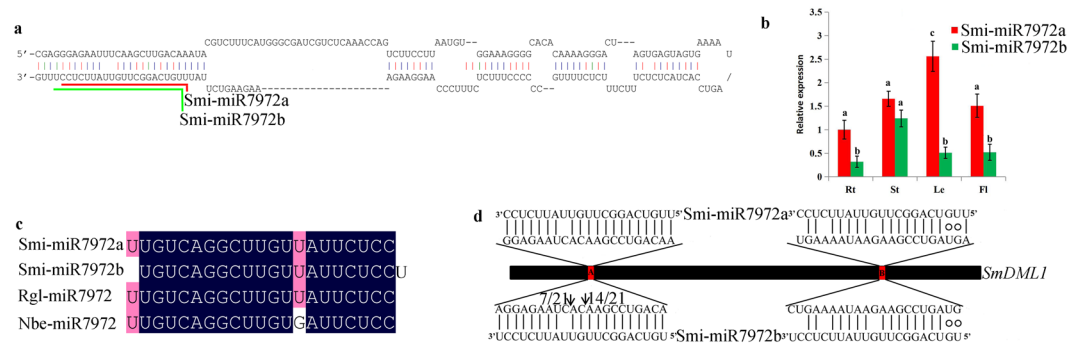


**Figure 6.** Responses of *SmDML* genes to 5Aza-C treatment. Fold changes of *SmDMLs* in leaves of *S. miltiorrhiza* plantlets treated with 1, 5, 10, 30 or 50  $\mu\text{M}$  of 5Aza-C for 15 days are shown. The expression levels were analyzed using the quantitative RT-PCR method. Expression level in leaves without treatment (0  $\mu\text{M}$ ) was arbitrarily set to 1 and the levels in leaves of 5Aza-C-treated plantlets were given relative to this. One-way ANOVA was calculated using IBM SPSS 20 software.  $P < 0.05$  was considered statistically significant and was represented by asterisks.

The FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values of *SmDML4–SmDML6* were very low, which are consistent with the low expression levels in roots analyzed using the quantitative RT-PCR technology (Fig. 5).

**Expression patterns of *SmDML* genes in response to 5Aza-dC.** It has been shown that DNA methylation-related genes are responsive to DNA methylation inhibitor treatment and methylation-sensitive expression of *ROS1* in plants is conserved and adaptive<sup>40</sup>. To elucidate the expression patterns of *SmDML* genes in response to DNA methylation inhibitor treatment, we treated *S. miltiorrhiza* plantlets with different concentrations of the DNA methyltransferase inhibitor, 5Aza-dC. Gene expression analysis showed that *SmDML1*, *SmDML2* and *SmDML4* were significantly methylation-responsive (Fig. 6). *SmDML1* transcripts were significantly reduced regardless of the concentration of 5Aza-dC treated (Fig. 6). The expression of *SmDML2* was significantly decreased at 5  $\mu\text{M}$  and 10  $\mu\text{M}$  5Aza-dC treatments; however, reduced expression of *SmDML2* was not significant at higher 5Aza-dC concentrations (Fig. 6). The expression of *SmDML4* was significantly decreased at 10  $\mu\text{M}$ , 30  $\mu\text{M}$  and 50  $\mu\text{M}$  5Aza-dC treatments. *SmDML5* and *SmDML6* transcript levels were not influenced by 5Aza-dC treatments, indicating they could be not involved in DNA demethylation. Interestingly, the expression of *SmDML3* was not detected in *S. miltiorrhiza* plantlets (Fig. 6); however, it was expressed in all mature tissues analyzed (Fig. 6), implying its important roles in *S. miltiorrhiza* plant maturation. Taken together, these data suggest that DNA methylation changes can influence *SmDML* expression in *S. miltiorrhiza*.

**miRNA-mediated posttranscriptional regulation of *SmDML1*.** In *Arabidopsis*, *AtDML3* transcripts are cleaved by miR402<sup>41,42</sup>. In *Nicotiana benthamiana*, *NbROS1* is targeted by Nb\_miRC1\_3p<sup>43</sup>. In order to know whether *SmDMLs* are regulated by miRNAs, we searched high-throughput *S. miltiorrhiza* small RNA database for miRNAs potentially targeting *SmDMLs* using psRNATarget<sup>44,45</sup>. A total of 44 small RNAs with sequence reads greater than four were identified under the maximum expectation of 3.0. The retrieved small RNA sequences were aligned with *S. miltiorrhiza* 99-3 genome and then secondary structures of genomic DNA sequences surrounding the small RNAs were predicted and manually checked. It resulted in the identification of *Smi-MIR7972*. This miRNA gene generates two 21nt-miRNAs, termed Smi-miR7972a and Smi-miR7972b, respectively. Smi-miR7972a and Smi-miR7972b are overlapped, and Smi-miR7972a starts 1nt upstream relative to Smi-miR7972b (Fig. 7a). Quantitative qRT-PCR analysis using Smi-miR7972a- and Smi-miR7972b-specific primers showed that Smi-miR7972a was highly expressed in leaves of 2-year-old, field nursery-grown *S. miltiorrhiza* plants (Fig. 7b). Its expression in roots, stems and flowers is similar and relatively low compared with its



**Figure 7.** Smi-miR7972 in *S. miltiorrhiza*. **(a)** The hairpin structure of Smi-miR7972. Smi-miR7972a and Smi-miR7972b are indicated by red and green lines. **(b)** Expression patterns of Smi-miR7972a and Smi-miR7972b in roots (Rt), stems (St), leaves (Le) and flowers (Fl) of *S. miltiorrhiza*. Fold changes of Smi-miR7972a and Smi-miR7972b are shown. Expression level of Smi-miR7972a in roots was arbitrarily set to 1 and the levels of Smi-miR7972a and Smi-miR7972b were given relative to this. Error bars were indicated by the standard deviations of three biological replicates. **(c)** Sequence alignment of miR7972s from *S. miltiorrhiza*, *R. glutinosa* and *N. benthamiana*. **(d)** Validation of Smi-miR7972a- and Smi-miR7972b-mediated cleavage using the modified 5' RLM-RACE method. Heavy black line represents open reading frame of *SmDML1*. The complementary sites of Smi-miR7972 in *SmDML1* are represented by A and B and shown in red. The nucleotide sequences of Smi-miR7972a and Smi-miR7972b from 3' to 5' and the complementary sites of *SmDML1* from 5' to 3' are shown in the expanded regions. Vertical dashes indicate Watson-Crick pairing. Circles indicate G:U wobble pairing. Vertical arrows indicate the 5' termini of Smi-miR7972-mediated cleavage products, as obtained by 5' RLM-RACE, with the frequency of clones shown.

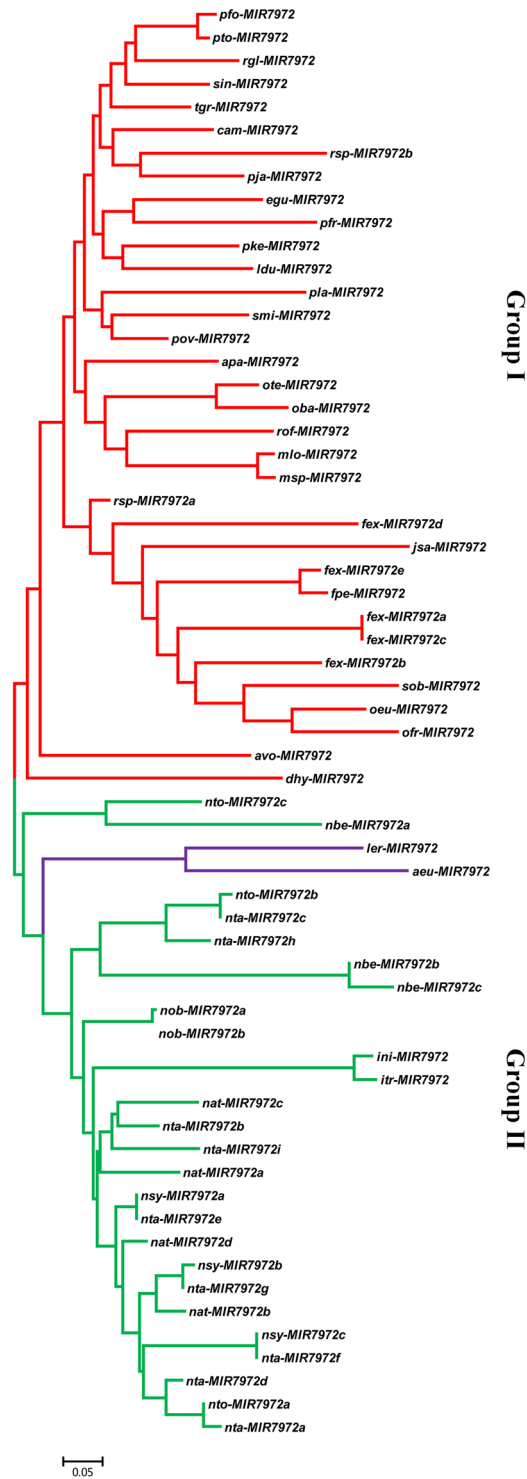
expression in leaves. Smi-miR7972b showed the highest expression in stems. Its expression in roots, leaves and flowers is similar. The expression of Smi-miR7972a was higher than Smi-miR7972b in all tissues analyzed (Fig. 7b). Blast analysis of Smi-miR7972a and Smi-miR7972b against miRBase (<http://www.mirbase.org/>) showed that Smi-miR7972a was identical to the functionally unknown Rgl-miR7972 from *Rehmannia glutinosa*<sup>46</sup>. Examination of Smi-miR7972a and Nb\_miRC1\_3p sequences revealed that they were highly similar with only a mismatched nucleotide. It suggests that Nb\_miRC1\_3p is actually a member of the *MIR7972* family. Thus, Nb\_miRC1\_3p was renamed to Nbe-miR7972 in this report (Fig. 7c).

Computational target prediction showed that *SmDML1* contained two regions near-perfectly complementary to Smi-miR7972a and Smi-miR7972b (Fig. 7d). The regions were named region A and region B, respectively. The expectations are 1.0 and 3.0 for region A and region B, respectively, as calculated using psRNATarget<sup>44</sup>. To analyze the cleavage site of Smi-miR7972a and Smi-miR7972b in *SmDML1*, the modified 5'RLM-RACE experiments were carried out as previously described<sup>47</sup>. The results confirmed that *SmDML1* was indeed cleaved at region A *in vivo* (Fig. 7d). However, the cleavage at region B was not validated. It indicates that region B is not cleaved due to the greater expectation. It is also possible that region B is cleaved only in certain tissues or plant developmental stages. Plant miRNAs usually cleave target mRNAs at the tenth complementary nucleotide from the 5' end of the miRNA<sup>48,49</sup>. Investigation of cleavage site at region A showed that there were two cleavage sites. One corresponds to the tenth complementary nucleotide from the 5' end of Smi-miR7972b, confirming the cleavage by Smi-miR7972b. The other one corresponds to the twelfth and thirteenth complementary nucleotides from the 5' end of Smi-miR7972a and Smi-miR7972a, respectively. It implied that this site was cleaved neither Smi-miR7972b nor Smi-miR7972a, indicating the complexity of mRNA cleavage mechanism.

**Phylogeny of *MIR7972* genes.** So far, miR7972 has been identified only in three plant species, including *S. miltiorrhiza*, *N. benthamiana* and *R. glutinosa*, all of which are placed in Lamiids. In order to know the phylogeny of *MIR7972* genes in plants, systematic and comprehensive investigation was carried out on the whole genome sequences of 40 Lamiids species (Table S4) and almost all RNA-seq data available for Lamiids species in SRA database. *MIR7972* genes exist only in Solanales, Boraginales and Lamiales. It was not found in other orders of the Lamiids clade. In total, 62 *MIR7972* genes were identified (Table S5), of which 34 were from 29 Lamiales species, 26 from 8 Solanales species, whereas the other two were from two Boraginales species. Interestingly, in Solanales, *MIR7972s* were identified only in *Nicotianeae* of Solanaceae and *Ipomoea* of Convolvulaceae. In Boraginales, *MIR7972s* were identified only in two plant species of *Lithospermum*. However, Lamiales *MIR7972s* widely exist in various families, including Oleaceae, Gesneriaceae, Plantaginaceae, Pedaliaceae, Verbenaceae, Lamiaceae, Phrymaceae, Paulowniaceae, and Orobanchaceae (Fig. 8). The majority of these families are not evolutionarily close to each other (<http://www.mobot.org/MOBOT/Research/APweb/welcome.html>). It indicates the complexity of *MIR7972* origin and evolution. The underlying mechanism remains to be elucidated. Most of the Lamiales species contain one *MIR7972* gene, whereas multiple copies were found in all *Nicotianeae* species, suggesting the occurrence of *MIR7972* duplication in *Nicotianeae*.

In order to reveal the phylogenetic relationship of *MIR7972s*, a neighbor-joining (NJ) phylogenetic tree was constructed for precursor sequences of the identified *MIR7972s* using MEGA7.0 (Fig. 8). *MIR7972s* could be divided into two groups. Group I contains Lamiales *MIR7972s*, whereas group II consists of *MIR7972s* from





**Figure 8.** Phylogenetic relationships of MIR7972 precursors in various Lamiids species. It includes *Ruellia speciosa* (rsp), *Mentha longifolia* (mlo), *Ocimum tenuiflorum* (ote), *Fraxinus excelsior* (fex), *Doroceras hygrometricum* (dhy), *Sesamum indicum* (sin), *Erythranthe guttata* (egu), *Nicotiana obtusifolia* (nob), *Ipomoea nil* (ini), *Ipomoea trifida* (itr), *Nicotiana attenuata* (nat), *Nicotiana sylvestris* (nsy), *Nicotiana tomentosiformis* (nto), *Nicotiana tabacum* (nta), *Nicotiana benthamiana* (nbe), *Andrographis paniculata* (apa), *Jasminum sambac* (jsa), *Syringa oblata* (sob), *Fraxinus pennsylvanica* (fpe), *Olea europaea* (oeu), *Osmanthus fragrans* (ofr), *Alectra vogelii* (avo), *Rehmannia glutinosa* (rgl), *Phtheirospermum japonicum* (pja), *Pedicularis keiskei* (pke), *Conopholis americana* (cam), *Paulownia fortunei* (pfo), *Paulownia tomentosa* (pto), *Plantago ovata* (pov), *Plantago lagopus* (pla), *Lippia dulcis* (ldu), *Tectona grandis* (tgr), *Ocimum basilicum* (oba), *Perilla frutescens* (pfr), *Rosmarinus officinalis* (rof), *Mentha spicata* (msp), *Lithospermum erythrorhizon* (ler), *Arnebia euchroma* (aeu). Species from Lamiales, Solanales and Boraginales are shown red, green and purple, respectively. MIR7972s could be divided into two groups, including group I and group II. The branch length is shown.

Solanales and Boraginales. It suggests closer evolutionary relationships between *MIR7972s* in Boraginales and Solanales in comparison with Lamiales. The results are consistent with previous studies showing that Lithospermeae species exhibit closer evolutionary relationships with Solanales species<sup>50</sup>.

## Discussion

*S. miltiorrhiza* is a widely cultivated medicinal crop in East Asia and a model medicinal plant for TCM studies. It has been used as medicinal materials to treat cerebrovascular diseases and cardiovascular diseases for hundreds of years<sup>51,52</sup>. Many genes associated with the production of bioactive compounds, including lipid-soluble tanninones and water-soluble phenolic acids, have been characterized<sup>2,53,54</sup>. However, little is known about epigenetic factors regulating gene expression in *S. miltiorrhiza*. Elucidation of regulatory process of DNA methylation is important for understanding gene expression regulatory mechanism associated with secondary metabolism and Dao-di herb formation in *S. miltiorrhiza*. In this study, six *S. miltiorrhiza* *SmDML* genes involved in DNA demethylation were identified and characterized. The number of *SmDML* genes is similar to those from castor bean, tomato, *Arabidopsis* and rice, ranging from three to six<sup>24,55,56</sup>.

Phylogenetic analysis of 66 DMLs from 16 plant species showed that *SmDML1* was a member of the ROS1 group, *SmDML2–SmDML4* belonged to the DME group, and *SmDML5* and *SmDML6* were included in the DML3 group (Fig. 4). The results from *Arabidopsis* and rice showed that DMLs in a group may have different functions. For instance, both rice *OsROS1a* and *OsROS1c* are members of the ROS1 group. *OsROS1a* is indispensable in both male and female gametophytes and critical to gametophytes<sup>33</sup>, whereas *OsROS1c* promotes expression and transposition of *Tos1721*. Among the three *SmDMLs* of the DME group, *SmDML4* is the smallest with only 720 amino acids. It lacks conserved motifs 9, 10, 11, and 13 compared with *SmDML2* and *SmDML3* (Fig. 3). *SmDML2–SmDML4* showed differential expression in mature plants (Fig. 5) and *SmDML3* was not expressed in plantlets. Similarly, differential expression was observed for *SmDML5* and *SmDML6* (Fig. 5), two *SmDMLs* of the DML3 group. It indicates that *SmDMLs* from a group may also be involved in distinct biological processes.

Generally, DNA methylation suppresses gene expression, whereas DNA demethylation promotes gene expression. In this study, we found that the expression of *SmDMLs* was down-regulated after 5Aza-dC treatment. It is consistent with significant down-regulation of *DMLs* in DNA methylation mutants in *Arabidopsis*<sup>57,58</sup>. Low expression of *SmDMLs* may contribute to maintain the certain DNA methylation level under the presence of 5Aza-dC. Balance between DNA methylation and demethylation is important for plant growth and development<sup>59,60</sup>. In *Arabidopsis*, the balance is monitored by a DNA methylation monitoring sequence (MEMS) in the *ROS1* promoter region<sup>40,60</sup>. It is unknown whether this mechanism also exists in *S. miltiorrhiza*. Alternatively, *SmDMLs* were not directly regulated by DNA methylation. Down-regulation of *SmDMLs* under 5Aza-dC treatment was mediated by a complex network with various mediators.

Plant miRNAs are a class of small non-coding RNAs with about 21–22 nt in length. They play vital roles in many biological processes through RNA cleavage<sup>61</sup>. *Arabidopsis AtDML3* is regulated by miR402. The regulation is important for seed germination under stress conditions<sup>42</sup>. In *N. benthamiana*, miRNA-mediated repression of *ROS1* may strength transcriptional gene silencing<sup>43</sup>. Although *S. miltiorrhiza* miRNAs have been reported<sup>3,45,62–65</sup>, their functions are largely unknown. Analysis of high-throughput sRNA data allowed us to identify *Smi-miR7972a* and *Smi-miR7972b* targeting *SmDML1* for cleavage. It indicates that miRNAs play significant roles in the regulation network of DNA methylation in *S. miltiorrhiza*. *Smi-miR7972b* accumulated at lower levels than *Smi-miR7972a* in the tissues analyzed (Fig. 7b). The variance between the levels of *Smi-miR7972a* and *Smi-miR7972b* could be caused by differential sequence preference of DICER-LIKE 1 (DCL1) responsible for miRNA biogenesis<sup>66,67</sup>. Experimental evidence showed that the cleavage of *SmDML1* was mediated by *Smi-miR7972b* rather than *Smi-miR7972a* in the tissues analyzed (Fig. 7d). The possibility that *Smi-miR7972a* cuts *SmDML1* in other tissues cannot be ruled out. Alternatively, *Smi-miR7972b* is easier to be recruited by AGO proteins for the formation of RNA induced silencing complex (RISC)<sup>68</sup>.

*MIR7972* genes were only identified in some species of the three Lamiids orders, including Solanales, Lamiales and Boraginales, and the number of *MIR7972* genes varied among species. The actual origin and evolution mechanism of *MIR7972* is unknown. One of the possibilities is that the *MIR7972* genes in different plant species were originated from a common ancestor of Solanales, Lamiales and Boraginales. Loss and duplication of *MIR7972* occurred in some plant species during evolution. This possibility is consistent with frequent birth and death of some *MIRNA* genes<sup>69</sup>. Independent origin and evolution of *MIR7972s* in different lineage is the other possibility. Evidence to support this possibility is that some miRNAs evolved are rarely lost and highly conserved across taxa<sup>70</sup>. The gain and loss of *MIR7972* could be important for plants of some lineages to survive in the stressful environments. Further investigating the biological function of *MIR7972* will help to elucidate the evolution mechanism of *MIR7972*.

## Materials and Methods

**Plant materials.** *Salvia miltiorrhiza* Bunge (line 99-3) plants were cultivated in a field nursery at the Institute of Medicinal Plant Development, Beijing, China. Roots, stem, leaves and flowers of two-year-old plants were collected and stored immediately in liquid nitrogen until use. For 5-aza-2'-deoxycytidine (5Aza-dC) treatment, plantlets were grown on Murashige and Skoog (MS) agar medium<sup>71</sup> supplemented with 0, 5, 10, 30 or 50  $\mu$ M 5Aza-dC (Sigma) for 15 days at 25 °C under a photoperiod of 16 h light and 8 h dark. Newly generated leaves were collected and immediately stored in liquid nitrogen until use. Three independent biological replicates were carried out for each treatment.

***SmDML* gene identification.** The deduced amino acid sequences of four *Arabidopsis* DML proteins were downloaded from the TAIR database (<http://www.arabidopsis.org>). To predict *SmDML* genes, *AtDMLs* were used as queries to search the two databases of *S. miltiorrhiza* whole genome sequence<sup>6,7</sup>. The searches were carried

out using the tBLASTN program<sup>72</sup>. An e-value cut off of  $1e^{-10}$  was applied. Gene models were predicted on the GENSCAN web server (<http://genes.mit.edu/GENSCAN.html>) for retrieved genomic DNA sequences<sup>34</sup>. The predicted gene models were then manually examined and corrected by comparison with DML genes identified from other plants using the BLASTx algorithm ([www.ncbi.nlm.nih.gov/blast/](http://www.ncbi.nlm.nih.gov/blast/)) and by alignment with RNA-seq data of *S. miltiorrhiza* transcriptome (<http://www.ncbi.nlm.nih.gov/sra>). The INTERPRO database (<http://www.ebi.ac.uk/interpro/>) was finally used to confirm each predicted protein sequence to be a DML.

**Gene structure and protein sequence analysis.** Gene structures of *SmDMLs*, *AtDMLs* and *OsDMLs* were determined on the Gene Structure Display Server (GSDS 2.0; <http://gsds.cbi.pku.edu.cn/index.php>). Coding sequences and corresponding genomic sequences were used as inputs. The deduced protein sequences of *SmDMLs*, *AtDMLs* and *OsDMLs* were analyzed for amino acid number, molecular weight (Mw), theoretical isoelectric point (pI) using the EXPASY PROTOPARAM tool (<http://www.expasy.org/tools/protparam.html>). Multiple sequence alignment was performed for *SmDML*, *AtDML* and *OsDML* amino acid sequences using ClustalW. Conserved motifs in *SmDML*, *AtDML* and *OsDML* proteins were detected using the MEME suite (<http://meme.sdsc.edu/meme/meme.html>).

**Phylogenetic analysis.** Unrooted neighbor-joining (NJ) trees were constructed using MEGA (version 7.0) with 1000 bootstrap replicates<sup>73</sup>. Protein sequences of DMLs from 16 plant species were downloaded from Phytozome (<http://phytozome.jgi.doe.gov/pz/portal.html>) (Table S1). Ka and Ks values were calculated for two gene pairs, *SmDML2/SmDML3* and *SmDML5/SmDML6*, using DNASP5 software<sup>74</sup>.

**RNA extraction and qRT-PCR analysis.** Total RNA was isolated from *S. miltiorrhiza* tissues using the plant total RNA extraction kit (Aidlab, China). The isolation was carried out following the manufacturer's instructions. RNA integrity was analyzed on an agarose gel. RNA quantity was determined using a NanoDrop 2000C spectrophotometer (Thermo Scientific, USA). Reverse transcription was conducted using PrimeScript™ RT reagent kit (TaKaRa, Japan). Gene specific primers were designed using Primer Premier 6 (PREMIER Biosoft Int, USA) based on *SmDML* coding sequences. *SmUBQ10* was used as an internal control as described previously<sup>2</sup>. The expression of *Smi-miR7972a* and *Smi-miR7972b* was analyzed using Mir-X miRNA qRT-PCR SYBR Kit (TaKaRa, Japan). The primers were listed in Table S2. qRT-PCR was performed in triplicate for each tissue sample using the SYBR premix Ex Taq™ kit (TaKaRa, China) on a CFX96 Touch™ real-time PCR system (Bio-Rad, USA). Three independent biological replicates were performed. Gene relative expression levels were calculated for Ct values using the  $2^{-\Delta\Delta C_t}$  method<sup>75</sup>. Differential expression among tissues and treatments was determined by one-way ANOVA using IBM SPSS 20 software (IBM Corporation, USA).

**RAN-Seq data and bioinformatic analysis.** Transcriptome sequencing data generated for periderm, phloem and xylem of *S. miltiorrhiza* roots was downloaded from SRA database of NCBI (SRX751296)<sup>54</sup>. Differential expression of *SmDML* genes was analyzed using TopHat2.0.12 and Cufflinks2.2.1<sup>39</sup>.

**Identification of *S. miltiorrhiza* miRNAs potentially targeting *SmDMLs*.** *S. miltiorrhiza* small RNAs potentially targeting *SmDMLs* for cleavage were predicted using psRNATarget<sup>44</sup>. Small RNAs from roots, stems, leaves and flowers of *S. miltiorrhiza* were downloaded from SRA database (SRX686651, SRX686652, SRX686653 and SRX686654)<sup>45</sup>. The maximum expectations of 3.0 and the target accessibility-allowed maximum energy to unpair the target site of 25 were applied. The predicted small RNAs were mapped to *S. miltiorrhiza* 99-3 genome using Bowtie<sup>76</sup>. No mismatch was allowed. Secondary structures of genomic sequences surrounding small RNA-aligned regions were predicted on the mfold web server<sup>77</sup>. The structures were manually checked and miRNAs were annotated under the criteria described<sup>70</sup>.

**5' RLM-RACE validation of miR7972-directed cleavage.** Roots, stems, leaves and flowers of two-year-old *S. miltiorrhiza* were used for validation of miR7972-directed cleavage. The modified RNA ligase-mediated rapid amplification of 5' cDNAs (5' RLM-RACE) was carried out using the FirstChoice® RLM-RACE Kit (Invitrogen, Carlsbad, CA). The nesting and nested primers were 5'-GGGGCAACCTGGTGGATTCATCT-3' and 5'-ACCGTTAACACCATTTTCCGA-3', respectively. Nesting PCR was carried out under the touchdown conditions: 94°C for 3 min, 5 cycles of 94°C for 30 s and 72°C for 90 s, 5 cycles of 94°C for 30 s, 70°C for 30 s and 72°C for 50 s, 25 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 1 min, followed by a final extension at 72°C for 10 min. Nested PCR amplification was performed under following conditions: 94°C for 3 min, 35 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 30 s, followed by a final extension at 72°C for 10 min.

**Identification of *MIR7972* genes in Lamiids.** *MIR7972* precursors from *S. miltiorrhiza*, *Rehmannia glutinosa* and *Nicotiana benthamian* were used to blast genomes of 40 Lamiids plant species listed in Table S3 using BLASTn<sup>72</sup>. Transcriptome-wide identification of *MIR7972* was performed through BLAST analysis of *Smi-MIR7972* or *Nbe-MIR7972* against RNA-seq reads (<https://www.ncbi.nlm.nih.gov/sra>) from Lamiids plants using BLASTn<sup>72</sup>.

## References

- Cheng, T. O. Danshen a popular chinese cardiac herbal drug. *J. Am. Coll. Cardiol.* **47**, 1487–1501 (2006).
- Ma, Y. *et al.* Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *J. Exp. Bot.* **63**, 2809–2823 (2012).
- Shao, F., Qiu, D. & Lu, S. Comparative analysis of the *Dicer*-like gene family reveals loss of miR162 target site in *SmDCL1* from *Salvia miltiorrhiza*. *Scientific Reports* **5**, 9891 (2015).
- Ji, A. J. *et al.* Genome-wide identification of the *AP2/ERF* gene family involved in active constituent biosynthesis in *Salvia miltiorrhiza*. *Plant. Genome* **9**, 1–11 (2016).

5. Song, J. Y. *et al.* *Salvia miltiorrhiza* as medicinal model plant. *Yao Xue Xue Bao* **48**, 1099–1106 (2013).
6. Zhang, G. *et al.* Hybrid de novo genome assembly of the Chinese herbal plant danshen (*Salvia miltiorrhiza* Bunge). *GigaScience* **4**, 62 (2015).
7. Xu, H. *et al.* Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Molecular Plant* **6**, 949–952 (2016).
8. Zhang, H. & Zhu, J. K. Active DNA demethylation in plants and animals. *Cold Spring Harb Symp. Quant. Biol.* **77**, 161–173 (2012).
9. Bala Tannan, N. *et al.* DNA methylation profiling in X;autosome translocations supports a role for L1 repeats in the spread of X chromosome inactivation. *Hum. Mol. Genet.* **23**, 1224–1236 (2014).
10. Fu, Y. *et al.* Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor. *EMBO J.* **32**, 2407–2417 (2013).
11. Chan, S. W., Henderson, I. R. & Jacobsen, S. E. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat. Rev. Genet.* **6**, 351–360 (2005).
12. Bird, A. P. & Wolffe, A. P. Methylation-induced repression—belts, braces, and chromatin. *Cell* **99**, 451–454 (1999).
13. Macdonald, W. A. Epigenetic mechanisms of genomic imprinting: common themes in the regulation of imprinted regions in mammals, plants, and insects. *Genet Res Int.* **2012**, 585024 (2012).
14. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* **10**, 295–304 (2009).
15. Dunoyer, P. *et al.* An endogenous, systemic RNAi pathway in plants. *EMBO J.* **29**, 1699–1712 (2010).
16. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet.* **11**, 204–220 (2010).
17. Finnegan, E. J. & Kovac, K. A. Plant DNA methyltransferases. *Plant Mol. Biol.* **43**, 189–201 (2000).
18. Lindroth, A. M. *et al.* Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* **292**, 2077–2080 (2001).
19. Wassenegger, M., Matzke, M. A. & Matzke, A. J. M. RNA-directed DNA methylation. *Plant Mol. Biol.* **43**, 203–220 (2000).
20. Hsieh, T. F. *et al.* Genome-wide demethylation of *Arabidopsis* endosperm. *Science* **324**, 1451–1454 (2009).
21. La, H. *et al.* A 5-methylcytosine DNA glycosylase/lyase demethylates the retrotransposon Tos17 and promotes its transposition in rice. *Proc. Natl. Acad. Sci. USA* **108**, 15498–15503 (2011).
22. Mok, Y. G. *et al.* Domain structure of the DEMETER 5-methylcytosine DNA glycosylase. *Proc. Natl. Acad. Sci. USA* **107**, 19225–19230 (2010).
23. Penterman, J. *et al.* DNA demethylation in the *Arabidopsis* genome. *Proc. Natl. Acad. Sci. USA* **104**, 6752–6757 (2007).
24. Zemach, A. *et al.* Local DNA hypomethylation activates genes in rice endosperm. *Proc. Natl. Acad. Sci. USA* **107**, 18729–18734 (2010).
25. Ibarra, C. A. *et al.* Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337**, 1360–1364 (2012).
26. Yu, A. *et al.* Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. *Proc. Natl. Acad. Sci. USA* **110**, 2389–2394 (2013).
27. Bharti, P., Mahajan, M., Vishwakarma, A. K., Bhardwaj, J. & Yadav, S. K. *AtROS1* overexpression provides evidence for epigenetic regulation of genes encoding enzymes of flavonoid biosynthesis and antioxidant pathways during salt stress in transgenic tobacco. *J. Exp. Bot.* **66**, 5959–5969 (2015).
28. Choi, Y. *et al.* DEMETER, a DNA Glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell* **110**, 33–42 (2002).
29. Kinoshita, T. *et al.* One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303**, 521–523 (2004).
30. Gehring, M. *et al.* DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**, 495–506 (2006).
31. Ortega-Galisteo, A. P., Morales-Ruiz, T., Ariza, R. R. & Roldán-Arjona, T. *Arabidopsis* DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Mol. Biol.* **67**, 671–681 (2008).
32. Gong, Z. *et al.* *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell* **111**, 803–814 (2002).
33. Ono, A. *et al.* A null mutation of *ROS1a* for DNA demethylation in rice is not transmittable to progeny. *Plant J.* **71**, 564–574 (2012).
34. Burge, C. B. & Karlin, S. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* **8**, 346–354 (1998).
35. Thibaudnissen, F., Ouyang, S. & Buell, C. R. Identification and characterization of pseudogenes in the rice gene complement. *BMC Genomics* **10**, 317 (2009).
36. Agius, F., Kapoor, A. & Zhu, J. K. Role of the *Arabidopsis* DNA glycosylase/lyase *ROS1* in active DNA demethylation. *Proc. Natl. Acad. Sci. USA* **103**, 11796–11801 (2006).
37. Liu, R. *et al.* A DEMETER-like DNA demethylase governs tomato fruit ripening. *Proc. Natl. Acad. Sci. USA* **112**, 10804–10809 (2015).
38. Prade, V. M. *et al.* The Pseudogenes of Barley. *Plant J.* **93**, 502–514 (2018).
39. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
40. Mittelsten Scheid, O., Williams, B. P., Pignatta, D., Henikoff, S. & Gehring, M. Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS genetics* **11**, e1005142 (2015).
41. Sunkar, R. & Zhu, J. K. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell* **16**, 2001–2019 (2004).
42. Kim, J. Y., Kwak, K. J., Jung, H. J., Lee, H. J. & Kang, H. MicroRNA402 affects seed germination of *Arabidopsis thaliana* under stress conditions via targeting DEMETER-LIKE protein3 mRNA. *Plant Cell Physiol.* **51**, 1079–1083 (2010).
43. Baksa, I. *et al.* Identification of *Nicotiana benthamiana* microRNAs and their targets using high throughput sequencing and degradome analysis. *BMC Genomics* **16**, 1–21 (2015).
44. Dai, X. & Zhao, P. X. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* **39**, W155–W159 (2011).
45. Xu, X. *et al.* Deep sequencing identifies tissue-specific microRNAs and their target genes involving in the biosynthesis of tanshinones in *Salvia miltiorrhiza*. *PLoS one* **9**, e111679 (2014).
46. Li, M. J. *et al.* Transcriptome/degradome-wide identification of *R. glutinosa* miRNAs and their targets: the role of miRNA activity in the replanting disease. *PLoS one* **8**, e68531 (2013).
47. Lu, S. *et al.* Novel and mechanical stress responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell* **17**, 2186–2203 (2005).
48. Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513–520 (2002).
49. Schwab, R. *et al.* Specific effects of microRNAs on the plant transcriptome. *Dev. Cell* **8**, 517–527 (2005).
50. Wu, F. Y. *et al.* Transcriptome analysis explores genes related to shikonin biosynthesis in Lithospermeae plants and provides insights into Boraginales' evolutionary history. *Scientific Reports* **7**, 4477 (2017).
51. Zhou, L., Zuo, Z. & Chow, M. S. Danshen: an overview of its chemistry, pharmacology, pharmacokinetics, and clinical use. *J. Clin. Pharmacol.* **45**, 1345–1359 (2005).
52. Yang, H., Han, L., Sheng, T., He, Q. & Liang, J. Effects of Replenishing Qi, promoting blood circulation and resolving phlegm on vascular endothelial function and blood coagulation system in senile patients with hyperlipemia. *J. Tradit. Chin. Med.* **26**, 120–124 (2006).

53. Cui, G. *et al.* Functional divergence of diterpene syntheses in the medicinal plant *Salvia miltiorrhiza*. *Plant Physiol.* **169**, 1607–1618 (2015).
54. Xu, Z. *et al.* Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J.* **82**, 951–961 (2015).
55. Cao, D. *et al.* Genome-wide identification of cytosine-5 DNA methyltransferases and demethylases in *Solanum lycopersicum*. *Gene* **550**, 230–237 (2014).
56. Xu, W., Yang, T., Dong, X., Li, D. Z. & Liu, A. Genomic DNA methylation analyses reveal the distinct profiles in castor bean seeds with persistent endosperms. *Plant Physiol.* **171**, 1242–1258 (2016).
57. Jia, Y. *et al.* Loss of RNA-Dependent RNA Polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. *PLoS genetics* **5**, e1000737 (2009).
58. Madzima, T. F., Huang, J. & McGinnis, K. M. Chromatin structure and gene expression changes associated with loss of MOP1 activity in *Zea mays*. *Epigenetics* **9**, 1047–1059 (2014).
59. Cao, X. & Jacobsen, S. E. Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr. Biol.* **12**, 1138–1144 (2002).
60. Lei, M. *et al.* Regulatory link between DNA methylation and active demethylation in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **112**, 3553–3557 (2015).
61. Kidner, C. A. & Martienssen, R. A. The developmental role of microRNA in plants. *Curr. Opin. Plant Biol.* **8**, 38–44 (2005).
62. Zhang, L. *et al.* Genome-wide analysis and molecular dissection of the SPL gene family in *Salvia miltiorrhiza*. *J. Integr. Plant Biol.* **56**, 38–50 (2014).
63. Shao, F. & Lu, S. Genome-wide identification, molecular cloning, expression profiling and posttranscriptional regulation analysis of the *Argonaute* gene family in *Salvia miltiorrhiza*, an emerging model medicinal plant. *BMC Genomics* **14**, 512 (2013).
64. 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* **15**, 1–12 (2014).
65. Zhang, H. *et al.* Identification and characterization of *Salvia miltiorrhizain* miRNAs in response to replanting disease. *PLoS One* **11**, e0159905 (2016).
66. Voinnet, O. Origin biogenesis and activity of plant microRNAs. *Cell* **136**, 669–687 (2009).
67. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).
68. Zhang, H., Xia, R., Meyers, B. C. & Walbot, V. Evolution, functions, and mysteries of plant ARGONAUTE proteins. *Curr. Opin. Plant Biol.* **27**, 84–90 (2015).
69. Fahlgren, N. *et al.* High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS One* **2**, e219 (2007).
70. Meyers, B. C. *et al.* Criteria for annotation of plant microRNAs. *Plant Cell* **20**, 3186–3190 (2008).
71. Skoog, F. & Murashige, T. A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiologia plantarum* **15**, 473–497 (1962).
72. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
73. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**, 1870–1874 (2016).
74. Rozas, J. & Rozas, R. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Comput. Appl. Biosci.* **11**, 621–625 (1995).
75. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).
76. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
77. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

## Acknowledgements

This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS) (2016-I2M-3-016), the Natural Science Foundation of China (31370327, 31570667 and 81773836), and the Beijing Natural Science Foundation (5152021).

## Author Contributions

J.L. analyzed the data, performed qRT-PCR and RACE experiments and participated in writing the manuscript. C.L. contributed to data analysis and RNA extraction. S.L. designed the experiment, participant in bioinformatics analysis, and wrote the manuscript. All authors have read and approved the version of manuscript

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-25315-w>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018