



Global Identification of the Full-Length Transcripts and Alternative Splicing Related to Phenolic Acid Biosynthetic Genes in *Salvia miltiorrhiza*

Zhichao Xu¹, Hongmei Luo¹, Aijia Ji¹, Xin Zhang¹, Jingyuan Song^{1*} and Shilin Chen^{1,2*}

¹ Institute of Medicinal Plant Development – Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China, ² Institute of Chinese Materia Medica – Chinese Academy of Chinese Medical Science, Beijing, China

OPEN ACCESS

Edited by:

Eugenio Benvenuto,
Italian National Agency for New
Technologies, Energy and Sustainable
Economic Development, Italy

Reviewed by:

Jianbo Xiao,
University of Macau, China
Laura Bassolino,
CREA, Italy
Shujuan Zhao,
Shanghai University of Traditional
Chinese Medicine, China

*Correspondence:

Jingyuan Song
jysong@implad.ac.cn;
Shilin Chen
slchen@icmm.ac.cn

Specialty section:

This article was submitted to
Plant Biotechnology,
a section of the journal
Frontiers in Plant Science

Received: 21 September 2015

Accepted: 19 January 2016

Published: 05 February 2016

Citation:

Xu Z, Luo H, Ji A, Zhang X, Song J
and Chen S (2016) Global
Identification of the Full-Length
Transcripts and Alternative Splicing
Related to Phenolic Acid Biosynthetic
Genes in *Salvia miltiorrhiza*.
Front. Plant Sci. 7:100.
doi: 10.3389/fpls.2016.00100

Salvianolic acids are among the main bioactive components in *Salvia miltiorrhiza*, and their biosynthesis has attracted widespread interest. However, previous studies on the biosynthesis of phenolic acids using next-generation sequencing platforms are limited with regard to the assembly of full-length transcripts. Based on hybrid-seq (next-generation and single molecular real-time sequencing) of the *S. miltiorrhiza* root transcriptome, we experimentally identified 15 full-length transcripts and four alternative splicing events of enzyme-coding genes involved in the biosynthesis of rosmarinic acid. Moreover, we herein demonstrate that lithospermic acid B accumulates in the phloem and xylem of roots, in agreement with the expression patterns of the identified key genes related to rosmarinic acid biosynthesis. According to co-expression patterns, we predicted that six candidate cytochrome P450s and five candidate laccases participate in the salvianolic acid pathway. Our results provide a valuable resource for further investigation into the synthetic biology of phenolic acids in *S. miltiorrhiza*.

Keywords: *Salvia miltiorrhiza*, hybrid-seq, full-length transcripts, phenolic acid biosynthesis, alternative splicing, cytochrome P450s, laccases

INTRODUCTION

The alternative splicing events of multiexon genes in multicellular eukaryotes can enhance the functional diversity of the encoded proteins and regulate gene expression through complex post-transcriptional mechanisms (Reddy et al., 2013). Recent alternative splicing analysis originating from next-generation sequencing (NGS, Illumina) has revealed that over 60% of multiexon genes undergo alternative splicing events in plants, such as *Arabidopsis thaliana* (Marquez et al., 2012), *Glycine max* (Shen et al., 2014), *Brachypodium distachyon* (Walters et al., 2013), and *Oryza sativa* (Zhang et al., 2010). However, the short-read assembly strategy of NGS limits its capacity to precisely quantify and predict alternative splicing events. In contrast, the long reads from SMRT sequencing (single molecule, real-time DNA sequencing using Pacific Biosciences RS II, PacBio) have demonstrated their advantage in sequencing full-length transcripts to identify and predict alternative splicing isoforms in human embryonic stem cells (Au et al., 2013; Roberts et al., 2013). Research has also addressed the disadvantage of high sequencing errors by correction with high-quality NGS reads (Au et al., 2012; Koren et al., 2012). Our recent study successfully

demonstrated the localization of tanshinones to the root periderm and revealed the molecular mechanism of tanshinone biosynthesis using hybrid-seq (next-generation and single molecular real-time sequencing, NGS and TGS) of the root transcriptome of *Salvia miltiorrhiza* (Xu et al., 2015).

Salvia miltiorrhiza Bunge is one of the most commonly used medicinal plants in Traditional Chinese Medicine (TCM), as its dried root or rhizome is of great phytochemical value in the treatment of cardiovascular diseases and inflammation, and as an anti-oxidant, among other uses (Cheng, 2006; Wang et al., 2007; Dong et al., 2011). The main active components of *S. miltiorrhiza* are hydrophilic salvianolic acids (SAs), such as rosmarinic acid (RA) and lithospermic acid B (LAB; Wang et al., 2007), and lipophilic diterpenoid components, such as tanshinones I/IIA, dihydrotanshinone, and cryptotanshinone (Lei et al., 2014). The availability of the nuclear and chloroplast genomes (Qian et al., 2013) and transcriptome (Hua et al., 2011; Luo et al., 2014), along with research related to the molecular regulation (Zhang et al., 2013, 2015; Tan et al., 2014; Li et al., 2015) and biosynthesis of its bioactive components (Guo et al., 2013, 2015; Bloch and Schmidt-Dannert, 2014), strongly favors *S. miltiorrhiza* as a potential model medicinal plant for TCM research.

There are two pathways for RA synthesis, namely, the phenylpropanoid pathway and the tyrosine-derived pathway, and many of the key genes encoding enzymes in *S. miltiorrhiza* have been identified (Di et al., 2013; Hou et al., 2013; Bloch and Schmidt-Dannert, 2014). In the phenylpropanoid pathway, phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumarate-CoA ligase (4CL) sequentially catalyze the conversion of L-phenylalanine into 4-coumaroyl-CoA. In the tyrosine-derived pathway, tyrosine aminotransferase (TAT) and 4-hydroxyphenylpyruvate reductase (HPPR) sequentially catalyze the conversion of L-tyrosine into 4-hydroxyphenyllactic acid, which is then catalyzed into 3,4-dihydroxyphenyllactic acid by an unknown CYP450 in *S. miltiorrhiza* (Di et al., 2013). Rosmarinic acid synthase (RAS) catalyzes conversion of the products from the two pathways to form 4-coumaroyl-3', 4'-dihydroxyphenyllactic acid (4C-DHPL), and SmCYP98A78 (allelic variant of SmCYP98A14, Chen et al., 2014) has been indicated as the specific hydroxylase that catalyzes the conversion of 4C-DHPL to RA (Di et al., 2013). Finally, oxidative dimerization of hydroxystilbene occurs, and laccase has been proposed to catalyze the oxidative reaction from RA to LAB (Giardina et al., 2010; Di et al., 2013). Although the phenolic acid biosynthetic pathway has in essence been proposed and identified, many homologous genes encoding key enzymes were uncovered by genome-wide strategy. Indeed, a total of 28 homologous genes of *SmPAL*, *SmC4H*, *Sm4CL*, *SmTAT*, *SmHPPR*, *SmRAS*, and *SmCYP98A78* have been identified by genome annotation (Wang et al., 2015).

In this study, using the hybrid-seq transcriptome of *S. miltiorrhiza* roots, we systematically analyzed the full-length transcripts and alternative splicing events of these 28 gene loci predicted as being related to RA biosynthesis. We then analyzed co-expression patterns and predicted candidate CYP450s and laccases related to the SA pathway. Our experiments not only

reveal full-length transcript and alternative splicing data but also provide a reference tool for future studies on the genes involved in the biosynthesis of phenolic acids.

MATERIALS AND METHODS

Plant Resources

Salvia miltiorrhiza (line 99-3) plants were cultivated at the Institute of Medicinal Plant Development (IMPLAD), Chinese Academy of Medical Sciences (CAMS) in an open experimental field. Roots, stems, and flowers were collected from 3-years-old plants growing in the field on May 27th, 2014. The roots were separated into three parts (periderm, phloem, and xylem) according to morphology and microstructure. Leaves with and without MeJA treatment (12 h, 200 μ M; Sigma-Aldrich, St. Louis, MO, USA) were collected from tissue culture *S. miltiorrhiza* (line 99-3) plantlets at 25°C under long-day condition of 16-h light/8-h dark. All of the collected tissues originated from the same clone of *S. miltiorrhiza* (line 99-3).

Transcriptomic Data

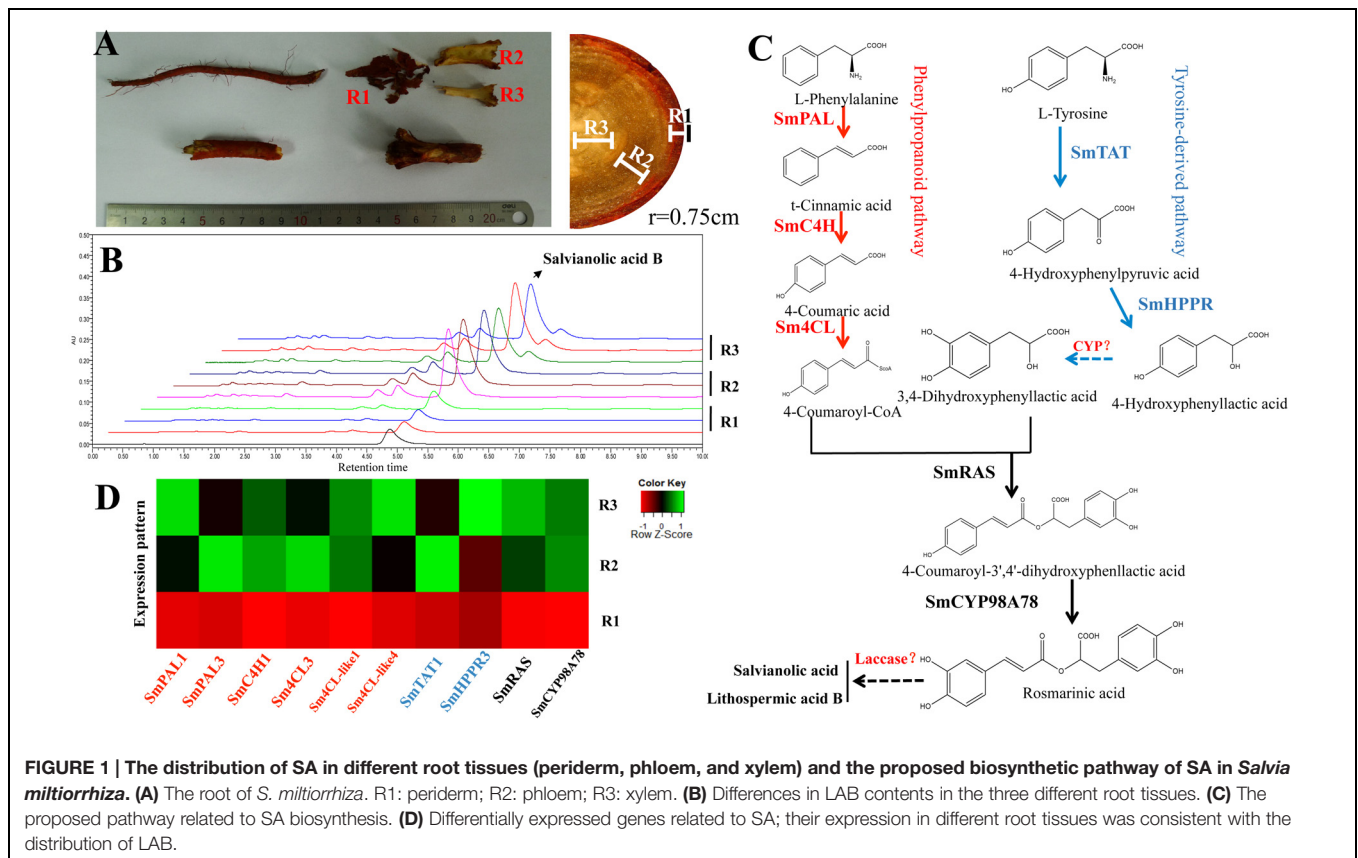
Single molecule real-time DNA sequencing data from pooled root tissues (periderm, phloem, and xylem) using the PacBio RS II platform (Pacific Biosciences of California, USA; Accession, SRX753381) and RNA-seq reads from different root tissues (periderm, phloem, and xylem) using the Illumina HiSeq 2500 platform (Illumina, USA) are reported in our recent study (Xu et al., 2015; Accession, SRR1640458). RNA-seq reads for different organs (root, stem, and flower) were generated using the Illumina HiSeq 2000 platform (Illumina, USA; Accession, SRP028388), and Illumina reads from leaves with and without 12 h MeJA treatment were obtained in a previous study (Luo et al., 2014; Accession, SRP051564).

Bioinformatic Analysis

Single molecule real-time DNA sequencing data were corrected with Illumina short reads using LSC 1.alpha software (Au et al., 2012). Alternative splicing isoforms were analyzed using IDP 0.1.7 software, employing SMRT sequencing reads, Illumina short reads, and genome scaffolds (Au et al., 2013). Differential gene expression in various root tissues, organs and under MeJA treatment was analyzed using Tophat 2.0.12 and Cufflinks 2.2.1 (Trapnell et al., 2012) by mapping the Illumina short reads to *S. miltiorrhiza* genome sequences. Heat maps were constructed using R statistical project (Gentleman et al., 2004).

Gene Structures and Phylogenetic Analysis

The alternative splicing isoforms found by IDP were viewed using the IGV 2.3.34 software (<http://www.broadinstitute.org/software/igv/>). The annotated gene sequences were corrected with the SMRT sequencing reads using Apollo software (Lee et al., 2013). Gene structures (e.g., intron, exon, intron phase) were also analyzed with Apollo. The full-length amino acid sequences of candidate CYP450s and laccases from *S. miltiorrhiza* and



other species were aligned with MEGA 6 (Tamura et al., 2013). Neighbor-joining trees were then constructed using the bootstrap method with 1,000 replications.

UPLC Analysis of LAB Content

The detection methods followed the Pharmacopeia of the People’s Republic of China. Periderm, phloem, and xylem samples were ground into powder (with three biological replicates for each sample), and each weighed sample of ground powder (0.2 g) was extracted with 50 mL of 75% methanol. After 1 h of heating reflux extraction, 75% methanol was added to complement and maintain a constant weight, and the sample was filtered through a 0.45- μ m syringe filter. In addition, an LAB standard was dissolved with 75% methanol at a concentration of 140 mg/L. Chromatographic separation was performed using an ACQUITY UPLC BEH C18 column (2.1 mm \times 100 mm, 1.7 μ m) with a mobile phase of 30% methanol, 10% acetonitrile, 1% methanoic acid, and 59% H₂O in a Waters UPLC system (Waters, USA). The detection wavelength was set to 286 nm.

Gene Expression Analysis by qRT-PCR

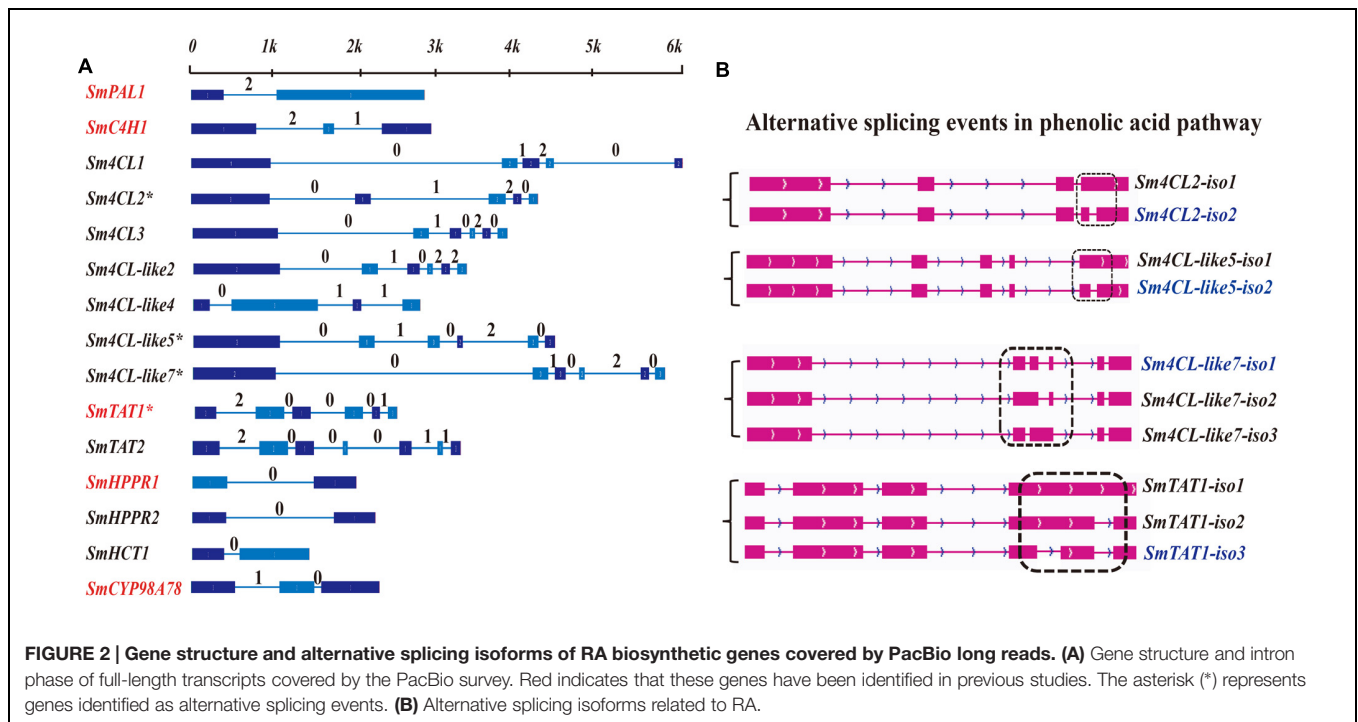
Nine RNA samples were isolated from different *S. miltiorrhiza* tissues (periderm, phloem, xylem, root, stem, leaf, and flower), which were collected from experimental field, and leaves from tissue culture plantlets were treated with MeJA (control or 12-h MeJA treatment). Total RNA (three biological replicates for each sample) was isolated using the RNeasy Plus Mini

kit (Qiagen, Germany). Reverse transcription was performed with PrimeScript™ Reverse Transcriptase (TaKaRa, Japan). The qRT-PCR primers were designed with Primer Premier 6 (Supplementary Table S1), and their specificity was verified by PCR. The qRT-PCR analysis was performed in triplicate using SYBR® Premix Ex Taq™ II (TaKaRa, Japan), with *SmActin* as a reference gene, and a 7500 real-time PCR system (ABI, USA). The Ct value was calculated for analyzing relative expression levels using the 2^{- $\Delta\Delta$ CT} method (Livak and Schmittgen, 2001). To detect differences in the expression of candidate genes among various tissues, one-way ANOVA was performed using IBM SPSS 20 software (IBM Corporation, USA). *P* < 0.01 was considered highly significant. Gene co-expression analysis of candidate genes was performed using Pearson’s correlation test.

RESULTS

Localization of SA Accumulation in *S. miltiorrhiza* Root

The rhizome or root of *S. miltiorrhiza* is the primary medicinal part of this well-known herb. The hydrophilic phenolic acids in the *S. miltiorrhiza* root are mainly distributed in the phloem and xylem. UPLC identification demonstrated a similar LAB content in the phloem and xylem, which were five times higher than that in the periderm (Figures 1A,B). These results provided a



potential basis for co-expression analysis of SA biosynthetic genes in the *S. miltiorrhiza* root.

Isoform Detection and Prediction of RA Biosynthetic Genes

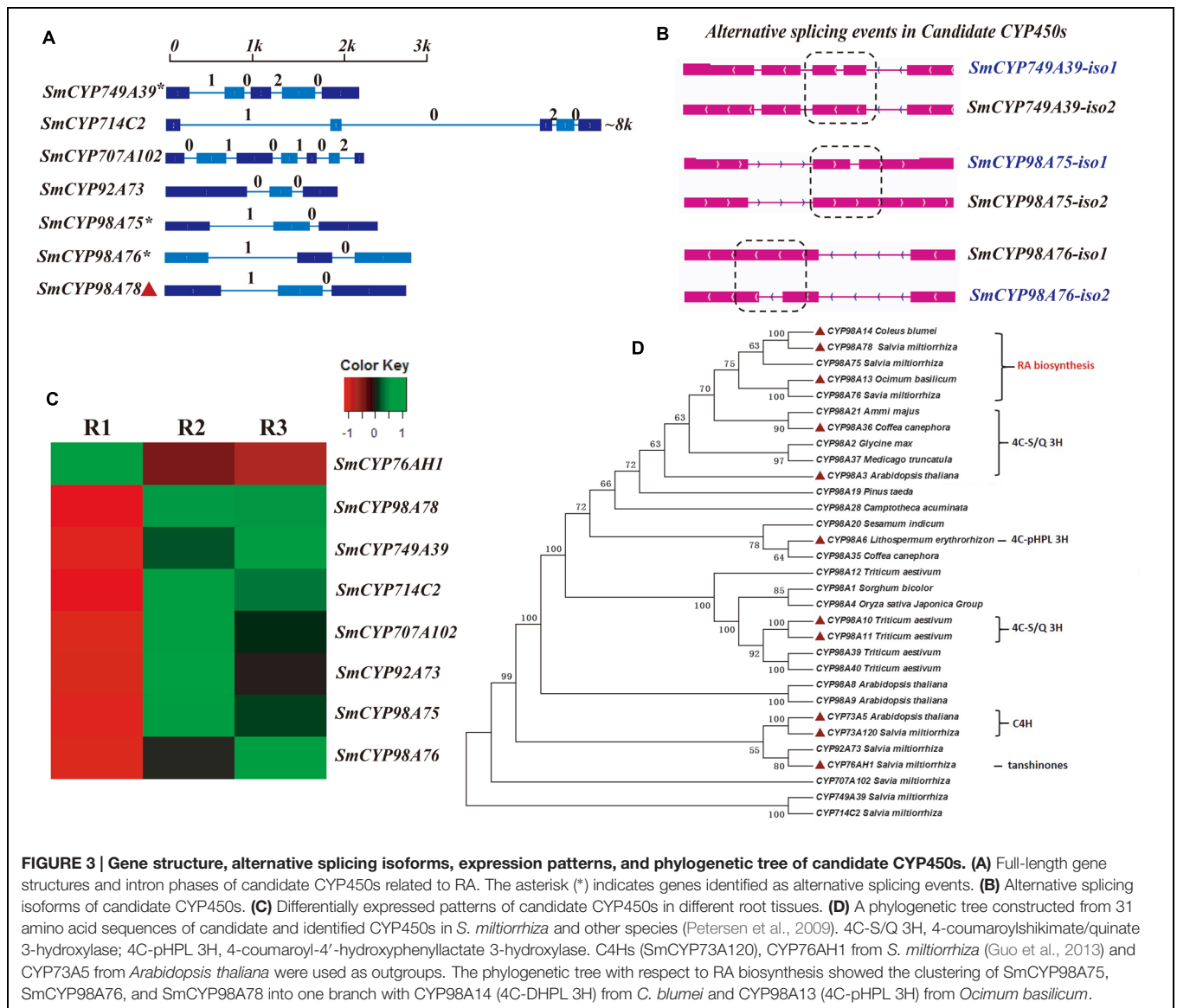
Using the next-generation sequencing platform (Illumina), RNA-seq data (a total of 867,864,885 reads) from *S. miltiorrhiza* periderm, phloem, xylem, root, stem, flower, leaf, and leaf after 12 h of MeJA treatment were collected. Using SMRT sequencing (PacBio RS II platform), full-length cDNA libraries from pooled periderm, phloem, and xylem samples were performed for a long-read survey, and 796,011 subreads were employed to identify full-length transcripts and alternative splicing events by hybrid-seq using the IDP (isoforms detection and prediction) pipeline.

A total of 28 candidate genes from the phenylpropanoid pathway and tyrosine-derived pathway, related to RA biosynthesis, were selected based on a genome-wide approach; these gene included *SmPALs* (3), *SmC4Hs* (2), *Sm4CLs* (10), *SmTATs* (3), *SmHPPRs* (3), *SmRASs* (6), and *SmCYP98A78* (Supplementary Table S2 and Figure 1C). The same approach was previously used to detect tanshinone biosynthetic genes (Xu et al., 2015). Fifteen gene loci were detected as full-length transcripts (Supplementary Figure S1), and their gene structures and intron phases are described in Figure 2A. *SmC4H2* might be a duplicated pseudogene of *SmC4H1* with an N-terminal deletion, as *SmC4H2* exhibits 74% homology with *SmC4H1* and is located at a distance of 7.5 kb from *SmC4H1* in the genome. *Sm4CL2*, *Sm4CL-like5*, *Sm4CL-like7*, and *SmTAT1* were identified as expressing alternatively spliced isoforms (Figure 2B), and all of the alternatively spliced junctions were characterized as intron retention. *Sm4CL2* and *Sm4CL-like5* each

expressed two isoforms, whereas *Sm4CL-like7* and *SmTAT1* each expressed three isoforms (Supplementary Table S5). Among their respective alternative splicing events, *Sm4CL2-iso2*, *Sm4CL-like5-iso2*, and *SmTAT1-iso3* were the dominantly expressed isoforms (Supplementary Table S5), though three isoforms of *Sm4CL-like7* all exhibited similar expression. We found that all of the intron retentions introduced premature termination codons (PTCs), and the PTC locations in *Sm4CL2-iso1*, *Sm4CL-like5-iso1*, *Sm4CL-like7-iso2*, *Sm4CL-like7-iso3*, *SmTAT1-iso1*, and *SmTAT1-iso2* were in intron 4, intron 5, intron 2, intron 3, intron 4, and intron 4, respectively, (Supplementary Figure S2).

Expression Profiles of Candidate RA Biosynthetic Genes

In this study, analysis of differentially expressed genes in the three root tissues showed that *SmPAL1*, *SmPAL3*, *SmC4H1*, *Sm4CL3*, *Sm4CL-like1*, *Sm4CL-like4*, *SmTAT1*, *SmHPPR3*, *SmRAS*, and *SmCYP98A78* exhibited low expression in the periderm and high expression in the phloem and xylem, in accord with the distribution of LAB (Figure 1D). In addition, the transcript levels of *SmPAL1*, *SmC4H1*, *Sm4CL3*, *Sm4CL-like1*, *SmTAT1*, *SmHPPR3*, *SmRAS*, and *SmCYP98A78* were significantly up-regulated after 12 h of MeJA treatment (Supplementary Table S2); however, the expression of *Sm4CL2*, *Sm4CL-like4*, *Sm4CL-like6*, and *SmHPPR2* was down-regulated after MeJA treatment. *SmTAT3*, *SmHCT2*, *SmHCT3*, and *SmHCT4* were identified as silenced genes. *SmHCT1* exhibited remarkably specific expression in the root xylem, yet *SmHCT5* showed only slight expression in the stem. Phylogenetic trees for 18 hydroxycinnamoyltransferase (HCT family) amino acid sequences including hydroxycinnamoyl-CoA:shikimate/quininate



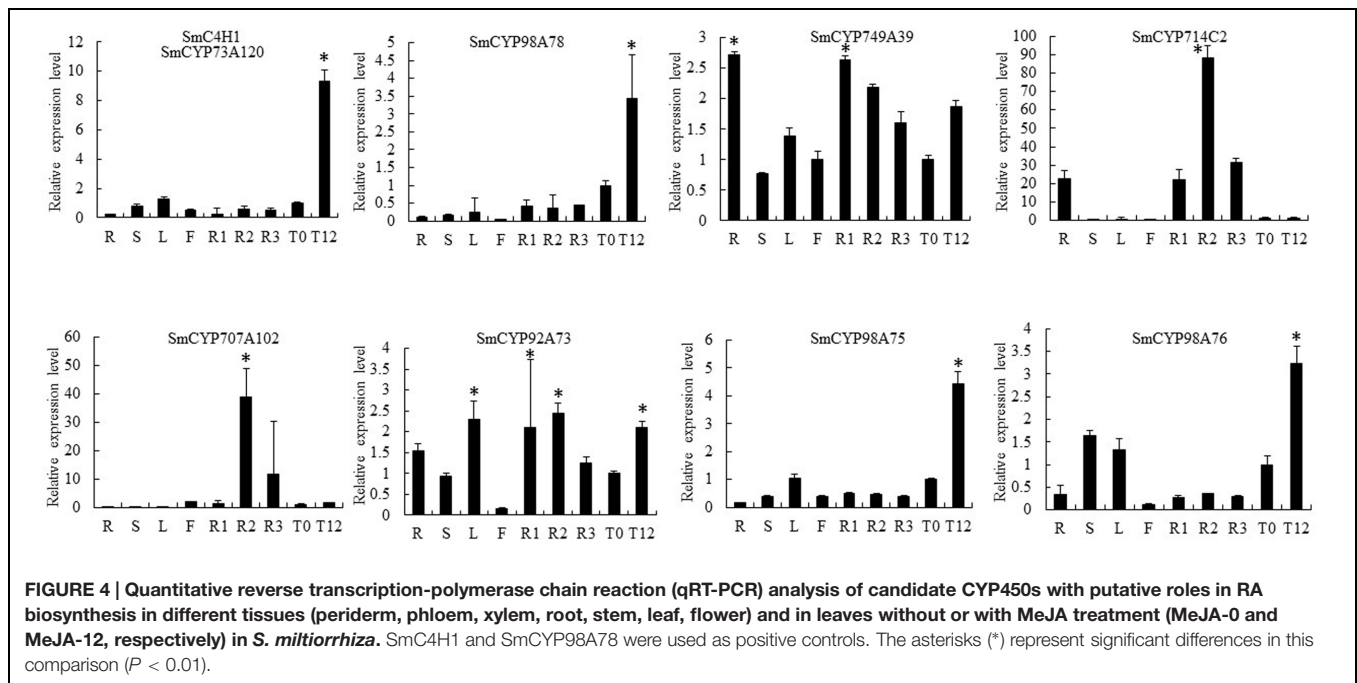
hydroxycinnamoyltransferases (HCS/QTs), RASs, and hydroxycinnamoyl/benzoyltransferases (HCBTs), in different species revealed clustering of five unidentified HCTs from *S. miltiorrhiza* with RASs, rather than with HCS/QTs (Supplementary Figure S3).

Co-expression Analysis and Isoform Identification of Candidate CYP450s

Our RNA-seq results showed opposite expression patterns for *SmCYP76AH1* and *SmCYP98A78*, in accord with the different distribution of tanshinones and phenolic acids in the periderm, phloem, and xylem. Moreover, six CYP450s were selected as candidate CYP450s related to RA biosynthesis based on the criteria of a phloem/periderm FPKM greater than 1.5 and a xylem/periderm FPKM greater than 1.5. The selected CYP450s included *SmCYP749A39*, *SmCYP714C2*, *SmCYP92A73*, *SmCYP98A75*, and *SmCYP98A76* (Supplementary Table S3).

A comprehensive evaluation of eight RNA-seq and qRT-PCR analyses of these CYP450s, including *SmCYP98A78* and *SmC4H1* (*SmCYP73A120*), indicated that expression level of the candidate CYP450s was significantly up-regulated by MeJA, with the exception of *SmCYP714C2*, which was not expressed in leaves (Figures 3C and 4). Furthermore, Pearson's correlation analysis of the qRT-PCR results showed highly significant co-expression of *SmCYP98A75*, *SmCYP98A76*, *SmCYP98A78*, and *SmC4H1* ($P < 0.01$).

All six CYP450s were identified as full-length transcripts in the PacBio transcriptome, and their gene structures and intron phases are described in Figures 3A,B. In addition, gene loci *SmCYP749A39*, *SmCYP98A75*, and *SmCYP98A76* were found to undergo alternative splicing events, with each expressing two gene isoforms. All alternative splicing events of these candidate CYP450s were classified as intron retention. For their respective alternative splicing events, *SmCYP749A39-iso1*,



SmCYP98A75-iso1, and *SmCYP98A76-iso1* were found to be the dominantly expressed isoforms (Supplementary Table S5), and intron retention for *SmCYP749A39-iso2*, *SmCYP98A75-iso2*, and *SmCYP98A76-iso2* introduced PTCs in exon 3, intron 2, and intron 2, respectively (Supplementary Figure S4).

To better understand the putative functions of these candidate CYP450s, we constructed a phylogenetic tree with 31 full-length amino acid sequences from various species, including some functionally identified CYP98As. CYP92A73 clustered into one branch with CYP76AH1, which was identified as catalyzing the miltiradiene to ferruginol step in tanshinone biosynthesis (Figure 3D). These two CYP450s were found to be neighbors of C4H from *S. miltiorrhiza* and *A. thaliana* (Figure 3D). The other predicted CYP450s, SmCYP707A102, SmCYP749A39, and SmCYP714C2, were distant from the CYP98A subfamily (Figure 3D).

Co-expression Analysis and Isoform Identification of Candidate Laccases

Eighty laccases were identified through genome-wide analysis in *S. miltiorrhiza*, with five identified based on the criterion of higher expression in the phloem and xylem than in the periderm: *SmLAC1*, *SmLAC2*, *SmLAC3*, *SmLAC4*, and *SmLAC5* (Figure 5C, Supplementary Table S4). Furthermore, RNA-seq and qRT-PCR analyses indicated that MeJA up-regulated the expression of *SmLAC1*, *SmLAC2* and *SmLAC5* (Figures 5C and 6). According to the qRT-PCR data, *SmLAC5* was significantly co-expressed ($P < 0.05$) with *SmCYP98A78* and *SmC4H1*.

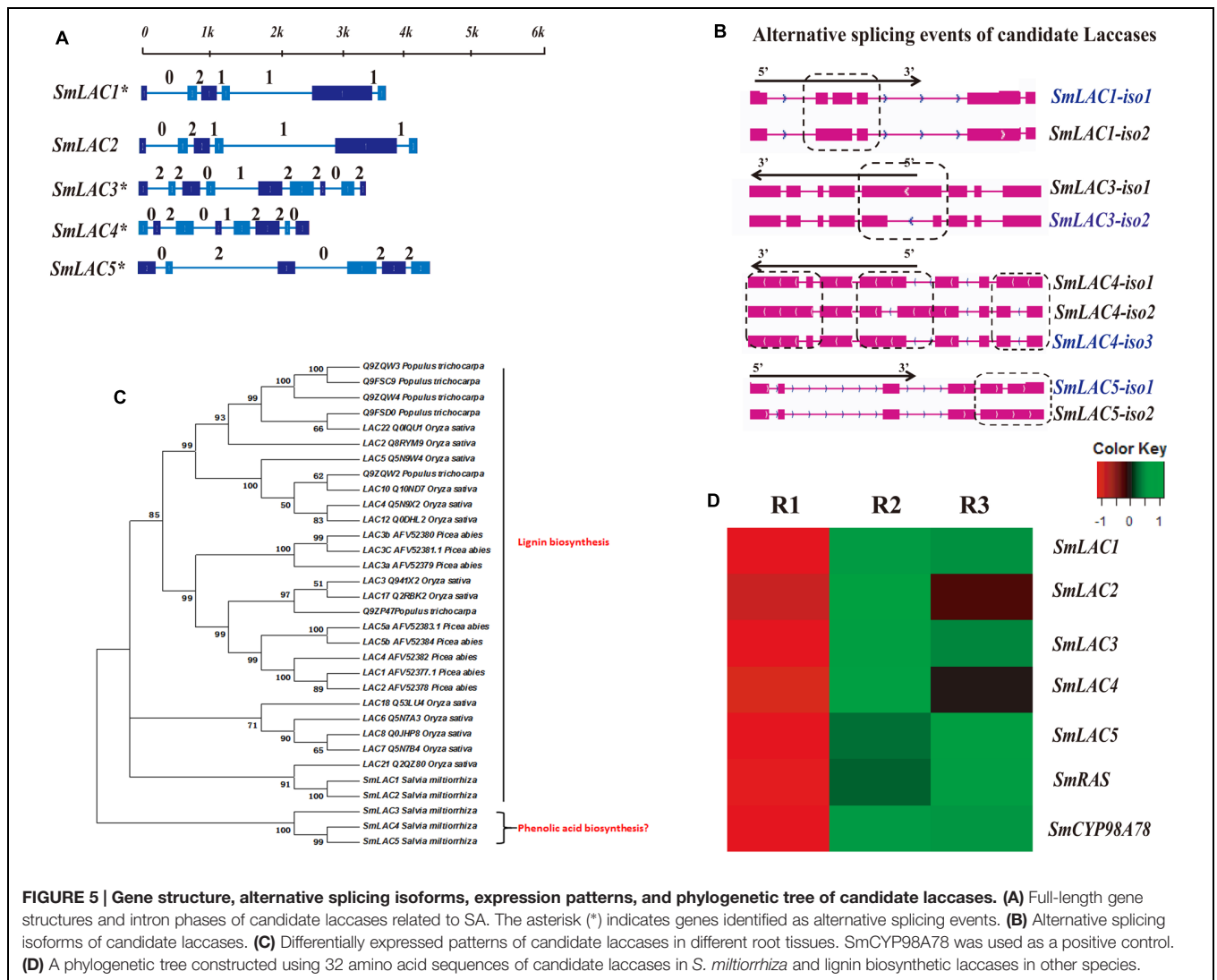
All five laccases were identified as full-length transcripts in the PacBio transcriptome, and their gene structures and intron phases are described in Figures 5A,B. In addition, the *SmLAC1*, *SmLAC3*, *SmLAC4*, and *SmLAC5* gene loci

were found to undergo alternative splicing events, with *SmLAC1*, *SmLAC3*, and *SmLAC5* each expressing two isoforms and *SmLAC4* three isoforms. All of the alternative splicing events in these selected laccases were classified as intron retention. *SmLAC1-iso1*, *SmLAC3-iso2*, *SmLAC4-iso3*, and *SmLAC5-iso1* were the dominantly expressed isoforms among their respective alternative splicing events (Supplementary Table S5), and the intron retention in *SmLAC1-iso2*, *SmLAC3-iso1*, *SmLAC4-iso1*, *SmLAC4-iso2*, and *SmLAC5-iso2* introduced PTCs in intron 2, intron 4, intron 1, intron 4, and intron 5, respectively (Supplementary Figure S5).

To predict the functions of the candidate laccases, a phylogenetic tree was constructed using 32 amino acid sequences from *Populus trichocarpa*, *Picea abies*, *Oryza sativa*, and *S. miltiorrhiza*. *SmLAC3*, *SmLAC4*, and *SmLAC5* were classified into different branches with other laccases that have been described as closely correlated with lignin biosynthesis in other species (Figure 5D).

DISCUSSION

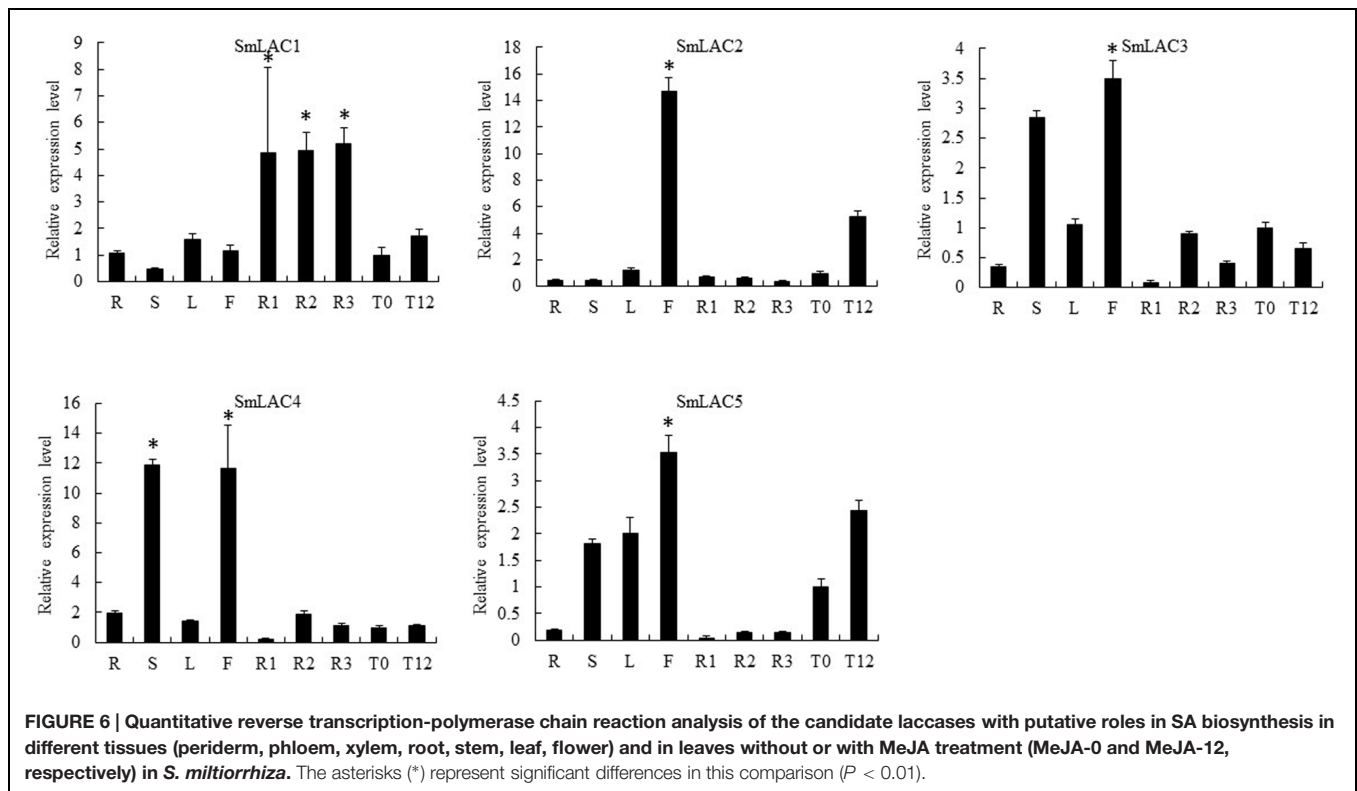
In this study, we analyzed full-length transcripts and alternative splicing events related to phenolic acid biosynthesis in different root tissue of *S. miltiorrhiza* by combining NGS and TGS technologies. Previous studies have only cloned a small number of full-length genes, such as *SmPAL1*, *SmC4H1*, *SmTAT1*, *SmHPPR1*, *SmRAS*, and *SmCYP98A78*, and identified their functions (Huang et al., 2008; Song and Wang, 2009, 2011; Xiao et al., 2009a, 2011; Di et al., 2013). Despite predicted locations and functions based on genome annotation, other full-length homologous genes and their functions have not yet



been identified (Wang et al., 2015). Among the 28 homologous genes identified as being involved in RA biosynthesis, the ability to detect 68% of full-length transcripts (15 full-length transcripts/22 expressed genes) and 27% of alternative splicing events at gene loci (4/15) indicates a significant advantage of hybrid sequencing in such discovery (Supplementary Table S2). Indeed, the availability of full-length transcripts will allow for establishing a metabolic engineering strategy with the aim of modulating the phenolic acid content, and the identification of alternative splicing events is beneficial for understanding the molecular mechanisms of phenolic acid biosynthesis in *S. miltiorrhiza*.

In line with our interest in phenolic acid biosynthesis, we found that not only the distribution of LAB but also the major expression of phenolic acid biosynthetic genes in the root occurred in the phloem and xylem (Figure 1). This agreement between phytochemical assay and gene expression in the root provided a basis for co-expression analysis. In addition, MeJA was found to dramatically promote the accumulation of

phenolic acids and the expression of key genes (Xiao et al., 2009b). Although many genes related to RA biosynthesis have been cloned and identified in other species, 28 homologous genes based on genome-wide identification generated more candidates to assist in fully elucidating the RA biosynthetic pathway in *S. miltiorrhiza*. *4CL1* and *4CL2* of phenylpropanoid pathway have been cloned and their functions identified *in vitro* (Zhao et al., 2006); however, the 4CL catalyzing 4-cinnamic acid to 4-coumaroyl-CoA *in vivo* remains unknown. *Sm4CL3*, *Sm4CL-like1*, and *Sm4CL-like4* are most likely to be involved in the synthesis of RA. The overexpression of *SmHPPR1* in tyrosine-derived pathway resulted in the accumulation of 4-hydroxyphenylpyruvic acid, the substrate of SmHPPR (Xiao et al., 2011). However, *SmHPPR3*, rather than *SmHPPR1*, might participate in RA biosynthesis. An additional step from 4-hydroxyphenyllactic acid to 3,4-dihydroxyphenyllactic acid was found in *S. miltiorrhiza* using a C¹³ tracer. As this step is likely to be catalyzed by an unknown CYP450 (Di et al., 2013), we then selected six CYP450s that were more significantly expressed



in the phloem and xylem than in the periderm (Supplementary Table S3). According to phylogenetic tree and qRT-PCR analyses, CYP98A75 and CYP98A76 likely participate in RA biosynthesis rather than as 4-coumaroylshikimate/quinate 3-hydroxylases in quinic acid and shikimic acid biosynthesis (Chen et al., 2014). A previous study proposed that a laccase was potentially involved in the oxidative dimerization of RA to synthesize LAB (Di et al., 2013). To explore the reactions that convert RA to LAB, five laccases were identified as exhibiting higher expression in the phloem and xylem than in the periderm (Supplementary Table S4). Furthermore, *SmLAC5* was considered to be the best candidate for LAB synthesis. Further studies of these candidate CYPs and laccases using transgenic methods and biochemical reactions may accurately elucidate the mechanism of phenolic acid biosynthesis.

The complexity of alternative splicing events plays a potentially important regulatory role in SA biosynthesis, and many studies focusing on alternative splicing events in *Arabidopsis* have been reported (Filichkin et al., 2010; Marquez et al., 2012). In contrast to humans, the most common type of alternative splicing event in plants appears to be intron retention (Au et al., 2013; Reddy et al., 2013). In fact, all of the identified alternative splicing events in *S. miltiorrhiza* SA biosynthesis showed intron retention. A recent study reported that most of the intron retention isoforms in *Arabidopsis* are predicted to be targets of nonsense-mediated decay (NMD) to regulate mRNA stability. Except for *Sm4CL-like7*, the alternative splicing events related to SA biosynthesis in *S. miltiorrhiza* expressed one predominant isoform (Supplementary Table S5).

Low-expression isoforms have been described as splicing errors, which commonly trigger NMD to maintain mRNA stability (Filichkin et al., 2010; Zhang et al., 2010; Marquez et al., 2012; Reddy et al., 2013; Walters et al., 2013; Shen et al., 2014). In addition, the highly expressed genes *Sm4CL-like7-iso2* and *Sm4CL-like7-iso3*, which contain PTCs downstream of the splice junctions, might be subjected to NMD to eliminate incomplete transcripts (Supplementary Figure S2). Another prediction about these two PTC isoforms of *Sm4CL-like7* is that small interfering peptides with absent functional domains could form non-functional dimers that compete with and negatively regulate functional proteins. Our results clearly detected and predicted alternative splicing events related to SA biosynthesis, though the actual functions of the alternative splicing isoforms remain unknown. Thus, the systematic identification of co-expression, full-length transcripts and alternative splicing events related to the biosynthesis of lipophilic diterpenoid pigments (Xu et al., 2015) and hydrophilic phenolic acids in various root tissues of *S. miltiorrhiza* could better resolve the biology of the synthesis of such natural products.

In summary, we localized SA metabolism in the medicinal plant *S. miltiorrhiza* to the root phloem and xylem. We then identified full-length transcripts, encoding isoforms as well as alternative splicing events in SA biosynthesis and systematically analyzed six candidate CYP450s and five candidate laccases related to SA biosynthesis in *S. miltiorrhiza* using hybrid sequencing. Furthermore, our study provides a model for analyzing the full-length transcriptome and the biosynthesis of active constituents in other medicinal plants.

AUTHOR CONTRIBUTIONS

ZX, JS, and SC designed and coordinated the study. ZX, HL, XZ, and AJ performed experiments and analyzed the data. ZX, JS, HL, and SC wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant no. 81573398), the Major Scientific

and Technological Special Project for ‘Significant New Drugs Creation’ (Grant no. 2014ZX09304307001), the Program for Innovative Research Team at the Institute of Medicinal Plant Development (Grant no. IT1304), and the National Science-technology Support Plan of China (Grant no. 2012BAI29B01).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00100>

REFERENCES

- Au, K. F., Sebastiano, V., Afshar, P. T., Durruthy, J. D., Lee, L., Williams, B. A., et al. (2013). Characterization of the human ESC transcriptome by hybrid sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 110, E4821–E4830. doi: 10.1073/pnas.1320101110
- Au, K. F., Underwood, J. G., Lee, L., and Wong, W. H. (2012). Improving PacBio long read accuracy by short read alignment. *PLoS ONE* 7:e46679. doi: 10.1371/journal.pone.0046679
- Bloch, S. E., and Schmidt-Dannert, C. (2014). Construction of a chimeric biosynthetic pathway for the de novo biosynthesis of rosmarinic acid in *Escherichia coli*. *Chembiochem* 15, 2393–2401. doi: 10.1002/cbic.201402275
- Chen, H., Wu, B., Nelson, D. R., Wu, K., and Liu, C. (2014). Computational identification and systematic classification of novel cytochrome P450 Genes in *Salvia miltiorrhiza*. *PLoS ONE* 9:e115149. doi: 10.1371/journal.pone.0115149
- Cheng, T. O. (2006). Danshen: a popular chinese cardiac herbal drug. *J. Am. Coll. Cardiol.* 47:1498. doi: 10.1016/j.jacc.2006.01.001
- Di, P., Zhang, L., Chen, J., Tan, H., Xiao, Y., Dong, X., et al. (2013). 13C tracer reveals phenolic acids biosynthesis in hairy root cultures of *Salvia miltiorrhiza*. *ACS Chem. Biol.* 8, 1537–1548. doi: 10.1021/cb3006962
- Dong, Y., Morris-Natschke, S. L., and Lee, K. H. (2011). Biosynthesis, total syntheses, and antitumor activity of tanshinones and their analogs as potential therapeutic agents. *Nat. Prod. Rep.* 28, 529–542. doi: 10.1039/c0np00035c
- Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5:R80. doi: 10.1186/gb-2004-5-10-r80
- Giardina, P., Faraco, V., Pezzella, C., Piscitelli, A., Vanhulle, S., and Sannia, G. (2010). Laccases: a never-ending story. *Cell Mol. Life Sci.* 67, 369–385. doi: 10.1007/s00018-009-0169-1
- Guo, H., Zhu, N., Deyholos, M. K., Liu, J., Zhang, X., and Dong, J. (2015). Calcium mobilization in salicylic acid-induced *Salvia miltiorrhiza* cell cultures and its effect on the accumulation of rosmarinic acid. *Appl. Biochem. Biotechnol.* 175, 2689–2702. doi: 10.1007/s12010-014-1459-0
- Guo, J., Zhou, Y. J., Hillwig, M. L., Shen, Y., Yang, L., Wang, Y., et al. (2013). CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12108–12113. doi: 10.1073/pnas.1218061110
- Hou, X., Shao, F., Ma, Y., and Lu, S. (2013). The phenylalanine ammonia-lyase gene family in *Salvia miltiorrhiza*: genome-wide characterization, molecular cloning and expression analysis. *Mol. Biol. Rep.* 40, 4301–4310.
- Hua, W., Zhang, Y., Song, J., Zhao, L., and Wang, Z. (2011). De novo transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. *Genomics* 98, 272–279. doi: 10.1016/j.ygeno.2011.03.012
- Huang, B., Yi, B., Duan, Y., Sun, L., Yu, X., Guo, J., et al. (2008). Characterization and expression profiling of tyrosine aminotransferase gene from *Salvia miltiorrhiza* (Dan-shen) in rosmarinic acid biosynthesis pathway. *Mol. Biol. Rep.* 35, 601–612. doi: 10.1007/s11033-007-9130-2
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., et al. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* 30, 693–700. doi: 10.1038/nbt.2280
- Lee, E., Helt, G. A., Reese, J. T., Munoz-Torres, M. C., Childers, C. P., Buels, R. M., et al. (2013). Web Apollo: a web-based genomic annotation editing platform. *Genome Biol.* 14:R93. doi: 10.1186/gb-2013-14-8-r93
- Lei, X., Chen, J., Liu, C. X., Lin, J., Lou, J., and Shang, H. C. (2014). Status and thoughts of Chinese patent medicines seeking approval in the US market. *Chin. J. Integr. Med.* 20, 403–408. doi: 10.1007/s11655-014-1936-0
- Li, D., Shao, F., and Lu, S. (2015). Identification and characterization of mRNA-like noncoding RNAs in *Salvia miltiorrhiza*. *Planta* 241, 1131–1143. doi: 10.1007/s00425-015-2246-z
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Luo, H., Zhu, Y., Song, J., Xu, L., Sun, C., Zhang, X., et al. (2014). Transcriptomic data mining of *Salvia miltiorrhiza* in response to methyl jasmonate to examine the mechanism of bioactive compound biosynthesis and regulation. *Physiol. Plant.* 152, 241–255. doi: 10.1111/ppl.12193
- Marquez, Y., Brown, J. W., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* 22, 1184–1195. doi: 10.1101/gr.134106.111
- Petersen, M., Abdullah, Y., Benner, J., Eberle, D., Gehlen, K., Huchlerig, S., et al. (2009). Evolution of rosmarinic acid biosynthesis. *Phytochemistry* 70, 1663–1679. doi: 10.1016/j.phytochem.2009.05.010
- Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J., Pang, X., et al. (2013). The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE* 8:e57607. doi: 10.1371/journal.pone.0057607
- Reddy, A. S., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi: 10.1105/tpc.113.117523
- Roberts, R. J., Carneiro, M. O., and Schatz, M. C. (2013). The advantages of SMRT sequencing. *Genome Biol.* 14, 405–408. doi: 10.1186/gb-2013-14-6-405
- Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., et al. (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26, 996–1008. doi: 10.1105/tpc.114.122739
- Song, J., and Wang, Z. (2009). Molecular cloning, expression and characterization of a phenylalanine ammonia-lyase gene (SmPAL1) from *Salvia miltiorrhiza*. *Mol. Biol. Rep.* 36, 939–952. doi: 10.1007/s11033-008-9266-8
- Song, J., and Wang, Z. (2011). RNAi-mediated suppression of the phenylalanine ammonia-lyase gene in *Salvia miltiorrhiza* causes abnormal phenotypes and a reduction in rosmarinic acid biosynthesis. *J. Plant Res.* 124, 183–192. doi: 10.1007/s10265-010-0350-5
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tan, Y., Wang, K. Y., Wang, N., Li, G., and Liu, D. (2014). Ectopic expression of human acidic fibroblast growth factor 1 in the medicinal plant, *Salvia miltiorrhiza*, accelerates the healing of burn wounds. *BMC Biotechnol.* 14:74. doi: 10.1186/1472-6750-14-74
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq

- experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- Walters, B., Lum, G., Sablok, G., and Min, X. J. (2013). Genome-wide landscape of alternative splicing events in *Brachypodium distachyon*. *DNA Res.* 20, 163–171. doi: 10.1093/dnares/dss041
- Wang, B., Sun, W., Li, Q., Li, Y., Luo, H., Song, J., et al. (2015). Genome-wide identification of phenolic acid biosynthetic genes in *Salvia miltiorrhiza*. *Planta* 241, 711–725. doi: 10.1007/s00425-014-2212-1
- Wang, X., Morris-Natschke, S. L., and Lee, K. H. (2007). New developments in the chemistry and biology of the bioactive constituents of Tanshen. *Med. Res. Rev.* 27, 133–148. doi: 10.1002/med.20077
- Xiao, Y., Di, P., Chen, J., Liu, Y., Chen, W., and Zhang, L. (2009a). Characterization and expression profiling of 4-hydroxyphenylpyruvate dioxygenase gene (Smhppd) from *Salvia miltiorrhiza* hairy root cultures. *Mol. Biol. Rep.* 36, 2019–2029. doi: 10.1007/s11033-008-9413-2
- Xiao, Y., Gao, S., Di, P., Chen, J., Chen, W., and Zhang, L. (2009b). Methyl jasmonate dramatically enhances the accumulation of phenolic acids in *Salvia miltiorrhiza* hairy root cultures. *Physiol. Plant.* 137, 1–9. doi: 10.1111/j.1399-3054.2009.01257.x
- Xiao, Y., Zhang, L., Gao, S., Saechao, S., Di, P., Chen, J., et al. (2011). The c4h, tat, hppr and hppd genes prompted engineering of rosmarinic acid biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *PLoS ONE* 6:e29713. doi: 10.1371/journal.pone.0029713
- Xu, Z., Peters, R. J., Weirather, J., Luo, H., Liao, B., Zhang, X., et al. (2015). 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. doi: 10.1111/tj.12865
- Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., et al. (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* 20, 646–654. doi: 10.1101/gr.100677.109
- Zhang, S., Ma, P., Yang, D., Li, W., Liang, Z., Liu, Y., et al. (2013). Cloning and characterization of a putative R2R3 MYB transcriptional repressor of the rosmarinic acid biosynthetic pathway from *Salvia miltiorrhiza*. *PLoS ONE* 8:e73259. doi: 10.1371/journal.pone.0073259
- Zhang, X., Luo, H., Xu, Z., Zhu, Y., Ji, A., Song, J., et al. (2015). Genome-wide characterisation and analysis of bHLH transcription factors related to tanshinone biosynthesis in *Salvia miltiorrhiza*. *Sci. Rep.* 5:11244. doi: 10.1038/srep11244
- Zhao, S. J., Hu, Z. B., Liu, D., and Leung, F. C. (2006). Two divergent members of 4-coumarate: coenzyme a ligase from *Salvia miltiorrhiza* bunge: cDNA cloning and functional study. *J. Integr. Plant Biol.* 48, 1355–1364. doi: 10.1111/j.1744-7909.2006.00302.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Xu, Luo, Ji, Zhang, Song and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.