Plant Omics Journal POJ

POJ 8(2):148-158 (2015) ISSN:1836-3644

Transcriptome-wide identification of genes related to fatty acid biosynthesis in the medicinal plant *Salvia miltiorrhiza*

Jianhui Zhang*, Haimei Chen*, Kai Wu, Chang Liu†

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, 151 Malianwa North Road, Haidian district, Beijing100193, P.R. China

***Contributed equally**
†Corresponding author: cliu@implad.ac.cn; cliu6688@yahoo.com

Abstract

Salvia miltiorrhiza is an important medicinal plant that is commonly used in the treatment of cardiovascular diseases and neuro-asthenic insomnia. How to increase the production of its active components is an active area of research. Since the fatty acid biosynthesis pathway is linked to the production of its active components, here we performed a systematic identification of the genes involved in fatty acid biosynthesis based on RNA-Seq data from three different types of tissues of *S. miltiorrhiza*. First, 76 fatty acid biosynthesis-related genes were identified based on sequence similarity to known genes, which were classified into 29 gene families. Second, multiple sequence alignment and phylogenetic analyses were performed for members of each protein family. One of these families, the Acyl Carrier Protein (ACP) family is the important cofactor protein involved in fatty acid biosynthesis, demonstrated significant variations among the sequences from different species. Third, qRT-PCR was used to validate the expression levels for 49 genes. It is found that the expression profiles obtained using qRT-PCR were consistent with those obtained from the RNA-Seq results for 36 genes (74%). Forth, hierarchical clustering analysis of the 76 genes based on their gene expression profiles across the tissues revealed eight major clusters. Last, detailed examination showed that 28 genes were significantly differentially expressed among the tissues. These tissue-specific genes might serve as the switches for the regulation of fatty acid biosynthesis in *S. miltiorrhiza*. In summary, the current study has identified and characterized 76 novels genes involved in the fatty acid biosynthesis pathway in *S. miltiorrhiza*, laying the foundation for their future detailed functional characterization.

Keywords: Acyl carrier protein; Fatty acid biosynthesis; RNA-Seq; S*alvia miltiorrhiza*; Transcriptome.
Abbreviations: ACC_acetyl-CoA carboxylase, ACP_acyl carrier protein, BC_biotin carboxylase, BCCP_biotin carboxyl carrier protein, CT_carboxyl transferase , DHLAT_dihydrolipoyl acyltransferase, ENR_enoyl-ACP reductase , FAS_fatty acid synthase HAD_hydroxyacyl-ACP dehydratase, KAR_3-ketoacyl-ACP reductase, KAS_3-ketoacyl-ACP synthases , KASI_KAS isoform I KASII_KAS isoform II, KASIII_KAS isoform III, LACS_long-chainacyl-CoA synthetase , LPD_dihydrolipoamide dehydrogenase LS_lipoate synthase, LT_lipoyltransferase, MCMT_malonyl-CoA_acyl carrier protein malonyltransferase, PDHa_pyruvate dehydrogenase a subunit, PDHb_pyruvate dehydrogenase b subunit, PDHC_pyruvate dehydrogenase complex, qRT-PCR_ quantitative reverse transcriptase polymerase chain reaction, RNA-Seq_RNA Sequencing, SAD_ stearoyl-ACP desaturase Sm_*Salvia miltiorrhiza*, TGD1_trigalactosyldiacylglycerol1.

Introduction

Fatty acids are an important source of reserve energy and are essential cellular components, such as cell membrane lipids (Ohlrogge and Jaworski, 1997; Yuan et al., 2012). In addition, fatty acids are precursors for compounds that play significant roles in pathogen defense, which involves modulate basal, effector-triggered, and systemic immunity (Kachroo and Kachroo, 2009). Because of the important roles of fatty acids in plant development and the potential value of using metabolic engineering technologies to produce fatty acids in abundance, identification and characterization of genes that encode enzymes involved in fatty acid biosynthesis have been actively pursued in various plant species, such as *Arabidopsis thaliana* (Li-Beisson et al., 2013), *Glycine max* L. (Chen et al., 2012), and *Brassica oleracea* (Barker et al., 2007).

In nature, fatty acid biosynthesis consists of a conserved set of reactions, which can be grouped into two distinct types, type I and type II (Chan and Vogel, 2010). Type I is primarily present in eukaryotes. Most bacteria and eukaryotic organelles, such as mitochondria and plant plastids, exhibit type II fatty acid biosynthesis (White et al., 2005). At the beginning of the pathway, PDHC catalyzes oxidative decarboxylation of pyruvate to produce acetyl-CoA (Johnston et al., 1997). Then malonyl-CoA is formed from acetyl-CoA and bicarbonate by ACC. This step is the rate-limiting step in fatty acid biosynthesis (Konishi et al., 1996). Malonyl-CoA is loaded onto the ACP and enters a series of condensation reactions. Reactions are catalyzed by series of condensing enzymes including KASIII, KAR, HAD, and ENR (Mou et al., 2000). Subsequent condensations are repeated seven times with KASI that elongates fatty acid chain from C4 to C16). Final elongation, which forms C18 from C16, is catalyzed by KASII (White et al., 2005; Pidkowich et al., 2007). Production of 16- or 18-carbon fatty acids is catalyzed by FAS, an easily dissociable multi-subunit complex consisting of mono-functional enzymes (Brown et al., 2006). Among these proteins, the ACP cofactor protein, which covalently binds all fatty acyl intermediates, acts as a

metabolic scaffold (Bonaventure et al., 2003). The substrate is transferred by a thioester linkage to a conserved serine on the ACP from one enzyme to the next during fatty acid biosynthesis (Blatti et al., 2013).

S. miltiorrhiza Bunge (Danshen in Chinese) is a well-known medicinal plant belonging to the Labiatae family. Its produces a large variety of secondary metabolites, such as phenolic acid and tanshinones, which have been used extensively in the treatment of cardiovascular, cerebrovascular, hyperlipidemia, and acute ischemic stroke diseases for thousands of years (Wenping et al., 2011; Ma et al., 2012). The biosynthesis of fatty acid is tightly linked to the biosyntheses of secondary metabolites, including flavonoids, cuticular waxes, and other compounds (Nikolau et al., 2003; Lu et al., 2011). Elucidation of the fatty acid biosynthesis pathway will facilitate the study of secondary metabolism in *S. miltiorrhiza*. To date, only one gene involved in fatty acid biosynthesis, the carboxyl transferase b subunit (CTb), has been isolated through the *S. miltiorrhiza* chloroplast genome sequencing project (Qian et al., 2013; Chen et al., 2014b).

In this study, we performed a systematic identification of genes related to fatty acid synthesis in *S. miltiorrhiza* using RNA-Seq data generated from flower, root, and leaf tissues. A total of 76 genes were identified based on their sequence similarity to genes previous known to be involved in fatty acids biosynthesis. The genes were subjected to detailed sequence and phylogenetic analyses. Furthermore, we have examined their expression profiles across the three tissues in detail. Taking together, results obtained from this study have allowed us to reconstruct the fatty acids biosynthesis pathway in *S. miltiorrhiza* for the first time. Several genes were found to be potential switches for the regulation of fatty acids biosynthesis. This information has laid the foundation for future studies attempting to increase the production of secondary metabolites in *S. miltiorrhiza*.

Results

Identification and sequence analyses of genes related to fatty acid biosynthesis in S. miltiorrhiza

The RNA-Seq experiments using the leaf, flower and root tissues have been described previously (Chen et al., 2014a). To identify fatty acid biosynthesis gene from these RNA-Seq data, proteins related to fatty acid biosynthesis in *A. thaliana* were downloaded to construct a query dataset. The GenBank accession numbers for the sequences used in the query dataset are listed in Table 1. These sequences were then used to scan the transcripts identified from the RNA-Seq data. A total of 76 novel fatty acid synthesis-related genes were identified given E-value <= 1e-5, which belonged to 29 protein families. The GenBank accession numbers for the sequences are listed in Table 1. All of these genes are novel and they were named based on their homologs in *A. thaliana*. These genes included 62 full-length and 14 partial proteins and were involved in all steps of the fatty acid biosynthesis. The protein sequences of these genes were subjected to sequence analyses and showed different lengths, isoelectric points (pI), molecular weights, and subcellular localizations (Table 1). The protein sequences were then searched against Pfam database to identify conserved protein domains (Table S1). It shows that *S. miltiorrhiza* and *A. thaliana* proteins belonging to the same gene family tend to share the same protein domains, which were Hidden Markov Model (HMM) constructed with alignment of conserved protein sequences.

Multiple sequence alignment and phylogenetic analysis of fatty acid biosynthesis related proteins

To determine the degree of conservation for these genes involved in fatty acid biosynthsis. The identified protein sequences were subjected to multiple sequence alignment and phylogenetic analysis with their homologous sequences from various species, such as *A. thaliana*, *G. max L.*, *Oryza sativa*, and *B. oleracea*. Consistent with the general notion, the sequences were found to be highly similar to those from various species. From the beginning of the pathway, the PDHC catalyzed the oxidative decarboxylation of pyruvate to produce acetyl-CoA (Johnston et al., 1997). The deduced amino acid sequences of SmPDHa are most similar to those of *Solanum tuberosum* in the phylogenetic tree (Fig S1). Two genes of PDHb showed higher similarity to the genes of *A. thaliana* than other species (Fig S2). Both SmDHLAT1 and SmDHLAT2 were clustered together with those from *A. thaliana* with a bootstrap score of 99 (Fig S3). Four SmLPD proteins were identified. The subcellular location of two of these proteins was in the plastid, and the other two proteins were located in the mitochondrion, similar to those of AtLPD (Fig S4). The rate-limiting step in fatty acid biosynthesis is the formation of malonyl-CoA from acetyl-CoA and bicarbonate, catalyzed by ACC (Konishi et al., 1996). In plants, plastid ACC produces malonyl-CoA for the biosynthesis of long-chain fatty acids, whereas cytoplasmic ACC (shown in Fig S5) is important for the secondary metabolism (Nikolau et al., 2003; Lu et al., 2011). Plastid ACC in dicots is a four-subunit enzyme. Three subunits, BC, BCCP, and CTa, are most similar to those in *A. thaliana* (Fig S6, S8 and S9). CTb (Fig S7) is the only component of plant lipid metabolism known to be encoded by the plastid genome, (Ohlrogge and Browse, 1995) and is most similar to a protein from *Nicotiana tabacum*. In Fig S10, SmHACPS is adjacent to AtHACPS in the phylogenetic tree, with a bootstrap score of 100. Details regarding the multiple sequence alignments and phylogenetic trees for other gene families are shown in Fig S11 to Fig S28.

Given the important role ACP proteins playing in the metabolism across all life forms, the multiple sequence alignment of 15 full-length proteins, including eight SmACP and its homologs from seven other species is shown in Fig 1A. These proteins were further grouped into 5 subfamilies. The grouping of SmACP with ACP proteins from other species (e.g. subfamily 1) suggests that these subfamilies be created before speciation process that leads to the creation of *S. miltiorrhiza*. Protein sequences of ACP from various organisms vary significantly, but its modification structure is highly conserved with an Asp-Ser-Leu motif (Misra et al., 2009). The N terminal region of ACP shows lower sequence similarity than its C terminal region. These fifteen sequences of ACP were used to construct a phylogenetic tree (Fig 1B). The SmACP gene family could be classified into 5 subfamilies based on the topology of the tree, with bootstrap support of less than 60 and showed higher degree of variations.

Quantification and validation of gene expression

The gene expression levels of the 76 genes across the three tissues were calculated using the RSEM program (Li and Dewey, 2011) as part of the Trinity analysis pipeline (Haas et al., 2013). The results were represented by fragments per kilobase per million fragments mapped (FPKM) (Table S2). To validate the expression levels of genes obtained from the RNA-Seq experiments, we performed qRT-PCR on 49 selected genes. Among them, 29 genes were used to validate

Table 1. Molecular characteristics of 76 genes potentially involved in fatty acid biosynthesis.

S. miltiorrhiza									*A. thaliana*	
Gene Family Name	Gene Name	Full or Partial	Accession No.	Len	Loc	Sequence ID	PI	MW (kDa)	Gene Name	Gene Locus
PDHa	SmPDHa1	Full	KF887930	390	M	flr_comp62422_c0_seq2	7.8	42.9	AtPDHa	AT1G01090.1
PDHb	SmPDHb1	Full	KF887928	395	C	flr_comp61712_c0_seq1	6.29	42.87	AtPDHb1	AT1G30120.1
	SmPDHb2	Full	KJ784450	402	C	flr_comp56113_c0_seq1	6.84	43.88	AtPDHb2	AT2G34590.1
DHLAT	SmDHLAT1	Full	KF887922	470	C	flr_comp64645_c0_seq1	8.37	48.91	AtDHLAT1	AT1G34430.1
	SmDHLAT2	Full	KJ784428	463	C	flr_comp68098_c0_seq1	8.8	48.4	AtDHLAT2	AT3G25860.1
LPD	SmLPD1	Full	KF887936	567	C	flr_comp61449_c0_seq1	6.59	60.39	cpAtLPD1	AT3G16950.2
	SmLPD2	Full	KJ784445	507	M	flr_comp37506_c0_seq1	7.45	53.56	cpAtLPD2	AT4G16155.1
	SmLPD3	Partial	KJ784446	347	*	flr_comp43854_c0_seq1	4.88	36.95	mtAtLPD1	AT1G48030.1
	SmLPD4	Full	KJ784447	570	C	flr_comp57518_c0_seq1	8	60.14	mtAtLPD2	AT3G17240.1
CTa	SmCTa1	Full	KF887929	751	C	flr_comp69151_c0_seq9	6.94	83.72	AtCTa	AT2G38040.1
	SmCTa2	Full	KJ784419	727	*	flr_comp67507_c0_seq4	6.74	81.56		
CTb	SmCTb1	Full	KF887943	495	M	flr_comp58228_c0_seq1	5.36	55.81	AtCTb	ATCG00500.1
BC	SmBC1	Full	KF887921	536	C	flr_comp50492_c0_seq1	6.65	58.43	AtBC	AT5G35360.3
BCCP	SmBCCP1	Full	KF887920	253	C	flr_comp60326_c0_seq1	8.2	27.03	AtBCCP1	AT5G15530.1
	SmBCCP2	Full	KJ784427	285	C	flr_comp50323_c0_seq1	8.81	30.08	AtBCCP2	AT5G16390.1
ACC	SmACC1	Full	KF876015	225	M	flr_comp65559_c0_seq1	6.36	252.39	AtACC1	AT1G36160.1
									AtACC2	AT1G36180.1
HACPS	SmHACPS1	Full	KF887926	319	M	flr_comp62467_c0_seq24	9.13	36.24	AtHACPS1	AT2G02770.1
									AtHACPS2	AT3G11470.2
ACP	SmACP1	Full	KF887919	129	M	flr_comp36339_c0_seq1	5.63	14.36	AtACP1	AT3G05020.1
	SmACP2	Full	KJ784420	119	M	flr_comp51202_c0_seq1	5.17	13.35	AtACP2	AT1G54580.1
	SmACP3	Full	KJ784421	126	M	flr_comp59763_c0_seq1	4.93	14.03	AtACP3	AT1G54630.1
	SmACP4	Full	KJ784422	114	C	flr_comp58141_c1_seq1	4.94	12.78	AtACP4	AT4G25050.1
	SmACP5	Full	KJ784423	128	M	flr_comp36630_c0_seq1	4.78	14.23	AtACP5	AT5G27200.1
	SmACP6	Full	KJ784424	135	C	flr_comp49660_c1_seq1	4.97	14.64	AtACP6	AT1G65290.1
	SmACP7	Full	KJ784425	136	C	flr_comp70287_c0_seq1	4.6	14.96	AtACP7	AT2G44620.1
	SmACP8	Full	KJ784426	134	M	flr_comp68762_c2_seq1	8.75	22.83	AtACP8	AT5G47630.1
MCMT	SmMCMT1	Full	KF887939	407	C	flr_comp60789_c0_seq1	7.17	43.25	AtMCMT	AT2G30200.1
KASIII	SmKASIII1	Full	KF887934	400	M	flr_comp69608_c2_seq20	5.12	42.46	AtKASIII	AT1G62640.1
	SmKASIII2	Partial	KJ784437	311	*	flr_comp70506_c0_seq1	4.91	33.06	AtKAR1	AT1G24360.1
KAR	SmKAR1	Full	KF887931	318	C	flr_comp50364_c0_seq1	9.15	32.85	AtKAR2	AT1G62610.4
	SmKAR2	Partial	KJ784431	203	M	flr_comp10032_c0_seq1	5.8	21.3	AtKAR3	AT1G63380.1
	SmKAR3	Partial	KJ784432	123	*	flr_comp27087_c1_seq1	5.91	13.43	AtKAR4	AT3G46170.1
	SmKAR4	Partial	KJ784433	206	*	flr_comp53912_c1_seq2	5.36	21.72	AtKAR5	AT3G55290.1
									AtKAR6	AT3G55310.1
HAD	SmHAD1	Full	KF887927	228	C	flr_comp58300_c0_seq1	9.89	24.66	AtHAD1	AT2G22230.1
									AtHAD2	AT5G10160.1
									AtHAD3	AT5G60340.1
ENR	SmENR1	Full	KF887923	393	C	flr_comp70424_c0_seq1	8.72	41.56	AtER1	AT2G05990.1
	SmENR2	Full	KJ784429	391	C	flr_comp61629_c0_seq2	9.63	41.32	AtER2	AT3G45770.1
	SmENR3	Full	KJ784430	377	M	flr_comp70700_c0_seq1	9.82	41.44		
KASI	SmKASI1	Full	KF887932	473	C	flr_comp62041_c0_seq4	7.04	50.17	AtKASI	AT5G46290.3
	SmKASI2	Full	KJ784434	491	C	flr_comp63503_c0_seq1	8.32	51.97		
LS	SmLS1	Full	KF887937	384	M	flr_comp66287_c0_seq1	8.37	42.67	AtLS1	AT2G20860.1
	SmLS2	Full	KJ784448	356	C	flr_comp52566_c0_seq1	7.38	39.09	AtLS2	AT5G08415.1
LT	SmLT1	Full	KF887938	246	*	flr_comp69663_c1_seq8	9.18	27.72	AtLT1	AT4G31050.1
	SmLT2	Full	KJ784449	288	C	flr_comp67195_c0_seq1	7.03	32.8	AtLT2	AT1G04640.1

FATa	SmFATa1	Full	KF887924	370	C	flr_comp59239_c0_seq1	7.69	41.42	AtFATa1	AT3G25110.1
									AtFATa2	AT4G13050.1
FATb	SmFATb1	Full	KF887925	422	C	flr_comp38957_c0_seq1	7.26	46.64	AtFATb	AT1G08510.1
KASII	SmKASII1	Full	KF887933	534	C	flr_comp65911_c3_seq3	7.8	57.42	AtKASII	AT1G74960.2
	SmKASII2	Full	KJ784435	493	M	flr_comp69741_c0_seq8	7.35	52.71		
	SmKASII3	Partial	KJ784436	184	*	flr_comp55243_c0_seq1	4.48	19.6		
SAD	SmSAD1	Full	KF887940	396	C	flr_comp70326_c0_seq1	6.65	45.08	AtSAD	AT5G16240.1
	SmSAD2	Partial	KJ784451	229	*	flr_comp57018_c1_seq1	7.21	26.03		
	SmSAD3	Partial	KJ784452	341	M	flr_comp64244_c0_seq1	6.02	39		
LACS	SmLACS1	Full	KF887935	697	*	flr_comp67818_c0_seq1	6.91	77.21	AtLACS1	AT1G49430.1
	SmLACS2	Full	KJ784439	699	*	flr_comp58598_c0_seq2	7.27	76.64	AtLACS2	AT1G64400.1
	SmLACS3	Full	KJ784440	671	*	flr_comp59433_c1_seq3	7.62	74.43	AtLACS3	AT1G77590.1
	SmLACS4	Full	KJ784441	659	*	flr_comp62227_c0_seq2	6.77	73.7	AtLACS4	AT2G04350.1
	SmLACS5	Full	KJ784442	658	*	flr_comp66306_c0_seq1	6.94	73.26	AtLACS5	AT2G47240.1
	SmLACS6	Full	KJ784443	662	*	flr_comp67495_c0_seq4	6.47	74.95	AtLACS6	AT3G05970.1
	SmLACS7	Full	KJ784444	659	*	flr_comp67727_c0_seq1	6.59	73.26	AtLACS7	AT4G11030.1
	SmLACS8	Full	KJ784464	718	*	flr_comp69395_c0_seq1	8.29	78.26	AtLACS8	AT4G23850.1
									AtLACS9	AT5G27600.1
ABCAT	SmABCAT1	Full	KF887918	133	M	flr_comp69113_c0_seq5	9.27	150.17	AtABCAT1	AT4G39850.3
	SmABCAT2	Full	KJ784418	747	*	flr_comp60559_c0_seq1	8.23	83.99	AtABCAT2	AT1G54350.1
AAS	SmAAS1	Full	KF887917	724	C	flr_comp69583_c1_seq36	8	81.58	AtAAS	AT4G14070.1
	SmAAS2	Partial	KJ784438	334	*	flr_comp31252_c0_seq1	4.57	36.83		
TGD	SmTGD1	Full	KF887942	283	*	flr_comp68050_c0_seq9	5.34	30.2	AtTGD1	AT1G19800.2
GLB	SmGLB1	Full	KF887941	206	C	flr_comp40972_c0_seq2	9.68	22.02	AtGLB1	AT4G01900.1
WRI	SmWRI1	Full	KF887944	385	C	flr_comp63546_c0_seq4	5.35	42.87	AtWRI1	AT3G54320.1
	SmWRI2	Full	KJ784453	330	C	flr_comp47859_c0_seq5	8.43	37.51		
	SmWRI3	Partial	KJ784454	192	*	flr_comp53299_c0_seq1	9.77	21.35		
	SmWRI4	Full	KJ784455	461	*	flr_comp54654_c0_seq2	6.66	51.69		
	SmWRI5	Full	KJ784456	309	C	flr_comp56825_c0_seq4	7.58	34.98		
	SmWRI6	Partial	KJ784457	400	C	flr_comp61702_c0_seq5	7.24	45.1		
	SmWRI7	Partial	KJ784458	144	*	flr_comp62865_c1_seq2	9.42	16.38		
	SmWRI8	Partial	KJ784459	228	*	flr_comp63320_c0_seq3	9.76	25.22		
	SmWRI9	Full	KJ784460	414	C	flr_comp66193_c0_seq2	8.83	46.19		
	SmWRI10	Full	KJ784461	471	*	flr_comp66213_c2_seq20	8.53	51.4		
	SmWRI11	Partial	KJ784462	226	*	flr_comp67126_c0_seq2	10.2	25.05		
	SmWRI12	Full	KJ784463	610	*	flr_comp67856_c0_seq4	7.47	67.19		

Len: length; Loc: localization; PI: isoelectric point; MW: molecular weight; M: mitochondrion; C: chloroplast; *: unknown.

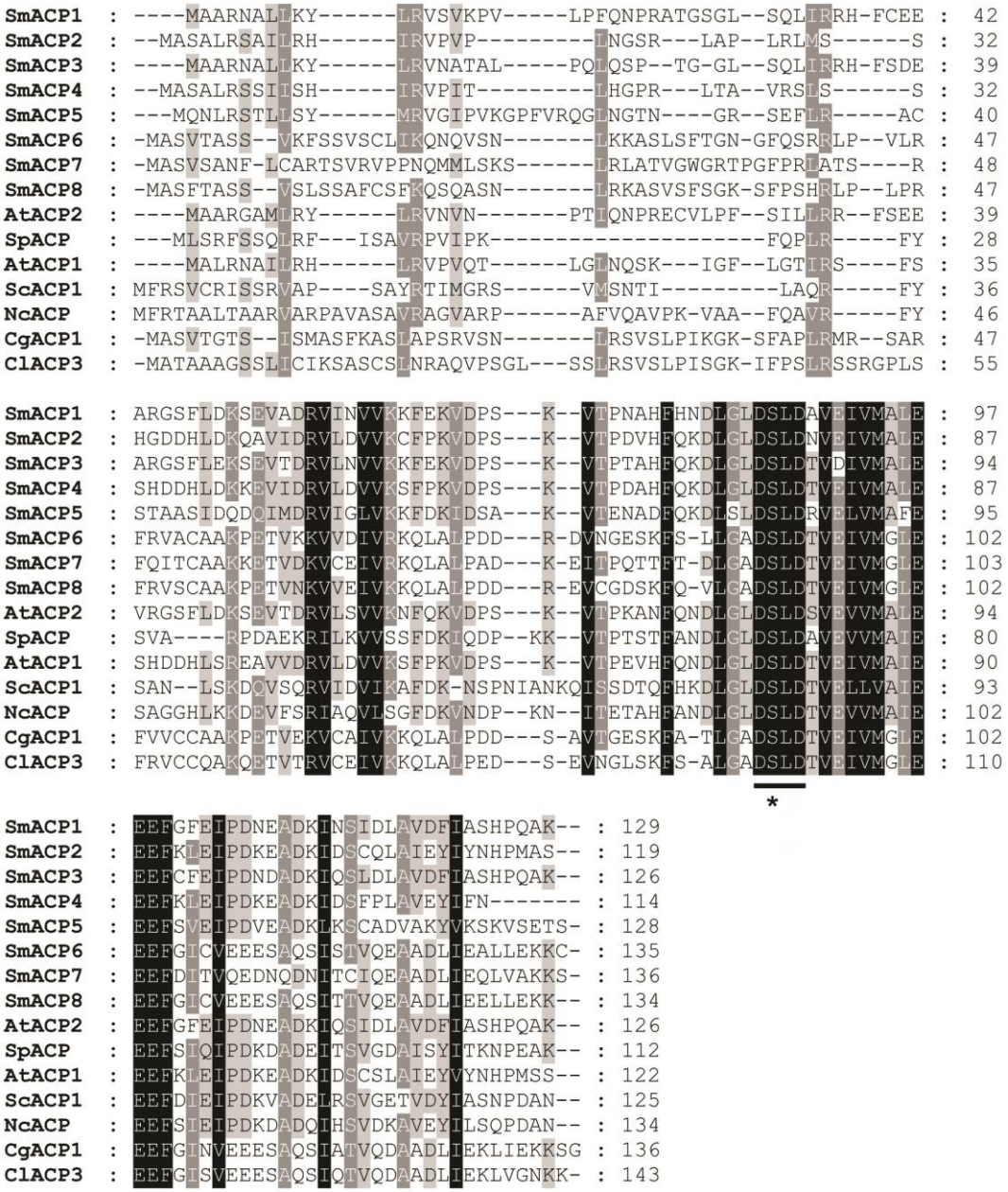

Fig 1. Multiple sequence alignment and phylogenetic tree for ACP proteins from *S. miltiorrhiza* and various species. (A) Multiple sequence alignment of ACPs. The alignments were generated using the ClustalW program. Black indicates columns in which all amino acids are identical. Dark grey and light grey indicate columns in which more than 80% and 60% amino acids are identical, respectively. Amino acids marked with a thick line represent the conserved motif. '*' indicate the Ser residue bound with 4'-phosphopantetheine. The species name and accession number for each ACP protein are enclosed in parentheses in the following text: SmACP1 (KF887919); SmACP2 (KJ784420); SmACP3 (KJ784421); SmACP4 (KJ784422); SmACP5 (KJ784423); SmACP6 (KJ784424); SmACP7 (KJ784425); SmACP8 (KJ784426); AtACP1 (*A. thaliana*, AT2G44620); AtACP2 (AT1G65290); SpACP (*Schizosaccharomyces pombe* 972h, Q10217); ScACP1 (*Saccharomyces cerevisiae* S288c, P32463); NcACP (*Neurosporacrassa* OR74A, P11943); CgACP1 (*Casuarinaglauca* (swamp oak), P93092); and ClACP3 (*Cuphealanceolata*, P52413). (B) Phylogenetic relationships of ACPs from *S. miltiorrhiza* and various species. The Neighbor-Joining trees were constructed using the MEGA6 package.

RNA-Seq data randomly, and the remaining 20 genes were used to validate the expression level of the genes that are significantly differentially expressed between at least two of three tissues. Primers used in this analysis are listed in Table S3. The mean expression profiles of three tissue types for each gene were arranged in order. If the order of expression profiles obtained from the RNA-Seq results was the same as that from the qRT-PCR results, the RNA-Seq results was considered to be validated. Then, the Pearson correlation coefficients of the tissue expression profiles of each gene obtained by RNA-Seq and qRT-PCR were calculated. The results of the RNA-Seq and qRT-PCR experiments were

considered consistent if the correlation coefficients of the corresponding gene expression profiles have a Pearson correlation coefficient >= 0.9. The results are shown in Table S4, and 24 gene pairs had r>=0.9. These data suggested that the expression patterns deduced from the FPKM values in our transcriptome analyses were reliable and can be used in downstream gene expression analyses.

The first group contained 29 genes that were randomly selected one per protein family. In Fig 2, 21 of 29 (72.4%) genes showed similar relative expression profiles between the RNA-Seq results and the qRT-PCR results, thereby suggesting that the expression patterns deduced from our transcriptome analyses are reliable. The second group contained 28 genes that were significantly differentially expressed (log2 ratio >= 2 and FDR significance score <0.001) across the three tissue types. Eight genes were shared by the two groups. In Fig 2, for genes belonging to the second group, 15 of 20 genes showed similar relative expression profiles between the RNA-Seq results and the qRT-PCR results.

It should be pointed out that all genes studied were found to be expressed in at least one tissue. Among the differentially expressed genes, those that played a role in the flow of carbon from sucrose to pyruvate exhibited differential expression profiles between root and leaf tissues. Genes that encoded enzymes of PDHC were more highly expressed in the leaf than in the root. Genes involved in fatty acid export, such as LACS and ABCAT, were also expressed at much higher levels in the root than in the leaf. However, tissue-specific expression patterns were found for members of the same gene family, thereby indicating the roles of these genes in different tissues.

Analysis of co-expressed genes in flower, leaf, and root tissues

To determine if functionally related genes were co-expressed, we clustered the genes based on their expression patterns across the three tissues (Fig 3). Eight main clusters designated as P1 to P8, respectively, were readily discernible. The P1 cluster comprised 12 genes, which share similar gene expression profiles across flower, leaf and root tissues, suggesting that this cluster of genes was constitutively expressed in a manner similar to the housekeeping genes. Genes in the clusters P2 and P3 showed the lowest expression levels in the leaf (10 genes) and flower (5 genes) tissues, respectively. In the cluster P4, four genes SmENR1, SmENR3, SmWRI5 and SmWRI7 were expressed with the highest level in the flower tissue and their expression levels were significantly different from those in the other two tissue types. In the cluster P5, two genes SmHACPS1 and SmMCMT1, which were both responsible for the biosynthesis of malonyl-CoA, showed similar expression profiles and had their highest expression levels in the flower.

In the cluster P6, four genes SmACP6, SmKAS2, SmLACS4 and SmLACS8 had highest expression levels in the flower tissue, similar to those found in the cluster P5. However, the expression among the three tissues for genes in the cluster P6 appeared to be more dramatic than those in the cluster P5.

All genes in the cluster P7 showed the highest expression levels in the leaf (21 genes) and the lowest expression levels in the root tissues. The genes in the cluster P8 were found to express highest in the leaf tissue, similar to those found in genes from the cluster P7. However, the expression levels of these genes were found significantly lower in the root tissue. This might reflect the fact that fatty acids were least produced in the root tissue.

Identification of potential regulatory switches

Considering the importance of fatty acid biosynthesis in cells, it would be expected that the cells would control the fatty acid biosynthesis through the expression regulation of genes involved in the pathway. The clustering analyses have already confirmed this hypothesis. The differential expressions of several genes are particularly dramatic. For example, SmWRI11 in cluster P3 was significantly up-regulated in the Root tissue and down-regulated in the flower tissue. In contrast, all genes in the cluster P6: SmACP8, SmKAS12, SmLACS6 and SmLACS4 were significantly up-regulated in the flower tissue, but down-regulated in the root tissue except for SmLACS4, which was down-regulated in the leaf tissue. Lastly, genes in the clusters P8 were significantly up-regulated in the leaf tissue, but down-regulated in the root tissue. These genes thus seem to represent the primary regulatory switch for fatty acid biosynthesis in various tissues, and further exploitation of their regulation is warranted.

Discussion

Previously, we have carried out transcriptomic analyses of samples from three different tissue types of *S. miltiorrhiza*, leaf, flower and root (Chen et al., 2014a). In the present study, the same dataset have been mined to identify genes involved in fatty acid biosynthesis. The identification of 76 genes with high similarity to those genes involved in the fatty acid biosynthesis in *A. thaliana* allowed us to construct the complete fatty acid biosynthesis pathway of *S. miltiorrhiza* for the first time.

ACP is the central cofactor protein for fatty acid synthesis

Fatty acid biosynthesis represents a central, conserved process in organism and a wide range of cellular processes depend on fatty acids. The fatty acid synthesis system provides acyl chains for many other processes. ACP, as the central cofactor protein for fatty acid synthesis, supplies acyl chains for the synthesis of other lipids, such as lipid A, and lipoic acid (Issartel et al., 1991; Jordan and Cronan, 1997). Moreover, ACP is utilized to produce the lipopeptide antibiotic, daptomycin, and the iron-carrying siderophore, enterobactin (Gokhale et al., 2007; Koglin and Walsh, 2009). The multiple sequence alignment of SmACP family shows that the ASL motif is highly conserved in Fig 1A. The determination of the three-dimensional interactive structure of acyl-ACP is a research hotspot, demonstrating the functional importance of ACP family (Chan and Vogel, 2010).

Transcriptional factors involved in the fatty acid biosynthesis

The fatty acid biosynthesis was regulated by several transcriptional factors. These transcriptional factors execute their functions by regulating mRNA turnover to control the changes in mRNA profiles. In our study, homologs of WRI1 gene family were found to express higher in the leaf and flower tissue than in the root tissue (shown in Fig 2). The PII/AtGLB1 protein is another transcriptional factor involved in the fatty acid biosynthesis. Previously, it was shown to interact with BCCP subunits to control the enzymatic activity of ACC (Feria Bourrellier et al., 2010).

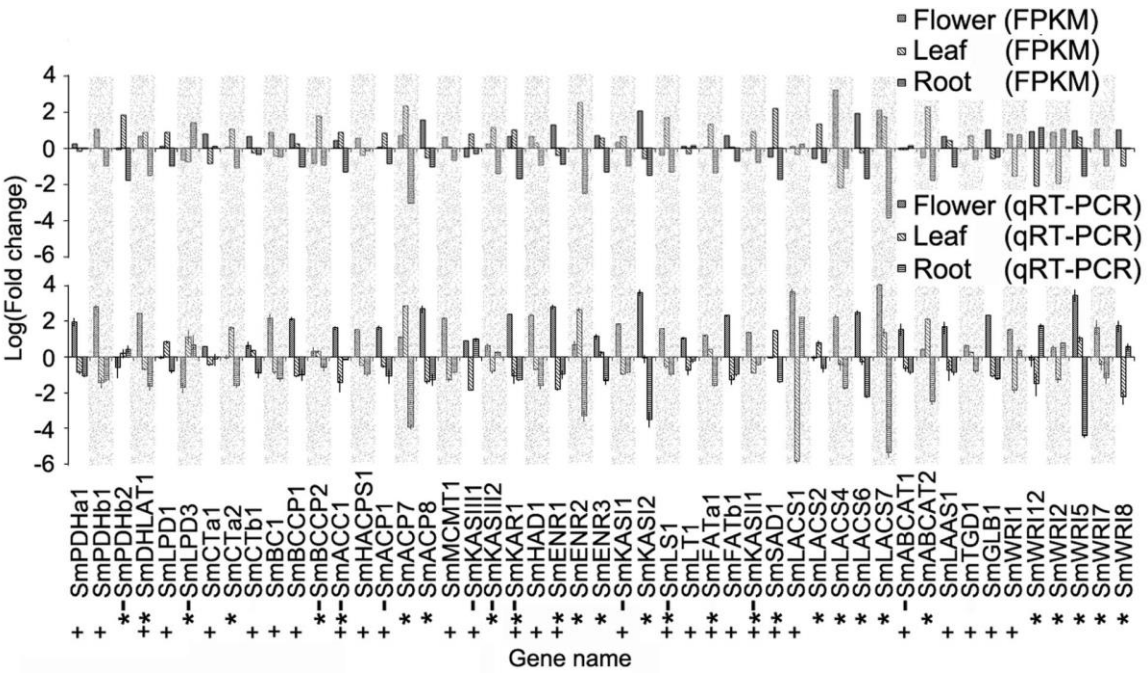

Fig 2. Comparison of gene expression levels determined by RNA-Seq and qRT-PCR experiments. Two groups of genes were selected to validate the RNA-Seq results by qRT-PCR. The first group comprised 29 genes that were randomly selected, one per family ('+'). The second group contains 28 genes ('*') that were found significantly differentially expressed across the three tissue types. Eight genes were shared by the two groups. The expression levels were represented as Log(Expression levels in particular tissue/Average expression level across the three tissues). Error bars represent standard deviation (STD). '|' indicates the genes, whose expression levels determined by RNA-Seq experiments were not consistent with those determined by qRT-PCR experiments.

In our analysis, we found that the homolog of GBL1 (SmGLB1) has the highest expression level in the leaf (shown in Fig 2), while the homolog of BCCP (SmBCCP) has the highest expression level in the flower. Further study is needed to investigate this discrepancy in their tissue-specific expression.

Co-expression of genes involved in the fatty acid biosynthesis

Genes belonging to the same pathway usually cooperate with each other to exercise their biological functions (Wenping et al., 2011). A previous study showed detailed expression patterns for genes involved in plant metabolic processes, such as fatty acid biosynthesis, by DNA microarray (Schmid et al., 2005). They provide an expression map that covers a wide range of developmental stages in the reference plant *A. thaliana*. Another study indicated that a number of genes encoding core fatty acid synthesis enzymes were likely to be co-regulated (Mentzen et al., 2008). Their study reveals that the module contains genes for the final reaction in biotin biosynthesis and for lipoic acid biosynthesis, obligate cofactors of acetyl-CoA carboxylase and pyruvate dehydrogenase, respectively. Biosynthesis and accumulation of secondary metabolites are often tissue-specific, and related genes of enzymes and regulators also show organ- or tissue-specific expression patterns (Yang et al., 2013).

In our study, we utilized the hierarchical clustering method to systematically study the co-expression of genes involved in the fatty acid biosynthesis. Eight clusters were identified, each demonstrated distinct expression profiles (shown in Fig 3) . In several cases, functional related genes showed similar expression profiles. For example, two genes found in cluster P4, SmWRI5 and SmWRI7, are homologs of the Transcription factor, WRINKLED1 (WRI1). WRI1 belongs to the APETALA2-ethylene responsive element-binding protein (AP2-EREBP), and has been found to directly control the transcriptional activation of the fatty acid biosynthetic pathway (Maeo et al., 2009) (Cernac and Benning, 2004; Baud et al., 2007). It is possible that SmWRI5 and SmWRI7 regulate the expression of SmENR1, SmENR3 in flower. For another example, four proteins in the cluster P5, SmBC, SmCTa, SmCTb, and SmBCCP, are the subunits of one complex plastid ACC. Thus, it is of no surprise to see that they had similar expression profiles. Another group includes four proteins, cytoplasmic ACC, its regulator SmGLB1, and the downstream genes SmHACPS1 and SmMCMT1. These four proteins showed similar expression profiles, supporting the notion that this is the critical step in fatty acid biosynthesis. The expression levels of these genes were highest in the flower tissue, thereby suggesting that the flower may be the tissue producing the largest amount of fatty acids among the three tissues under study. The last example is found in the cluster P7. PDHC is important for the flow of carbon from sucrose to pyruvate in the chloroplast. Since the chloroplast is most abundant in the leaf, it is reasonable to expect that PDHC is expressed highest in the leaf as well. PDHC comprised four subunits: PDHa, PDHb, DHLAT, and LPD. Among genes in the P7 cluster, four genes encode the subunits of enzymes SmPDHb2, SmDHLAT1, SmDHLAT2 and SmLPD1. The highest expression levels of these genes were indeed found to be in the leaf.

Co-expression analysis also identified genes that are not apparently related. For example, there are two gene families, one included SmLS1, SmLS2, and SmLT2 and the other included SmKASIII2, SmKASI1, SmKAR1, and SmENR2. Genes from these two families were also found co-expressed

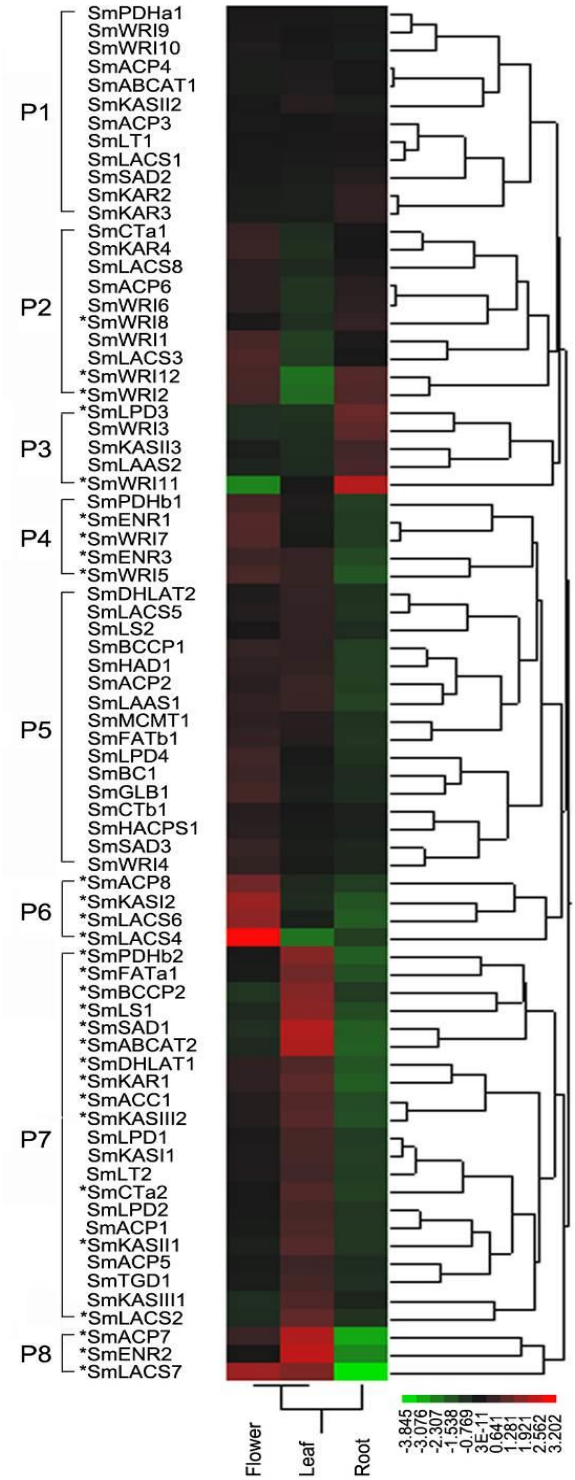

Fig 3. Hierarchical clustering of identified genes based on their tissue expression profiles.

For data preparation, the expression levels represented by FPKM values were floored with "1", log transformed, and mean-centered as described in the method section. Hierarchical clustering was conducted using JMP software. The heat map reflects the relative expression level of genes in the three tissues compared with their mean expression levels across the three tissues. Eight clusters (P1 to P8) are enclosed in square brackets. The color scale corresponds to log (FPKM+1)/(mean expression across the three tissue types). The green color indicates down-regulated expression, and the red color indicates up-regulated expression. Tissue names are shown below the heat map. '*' indicate that the expression levels of this gene were significantly different between at least two tissues ($P < 0.05$).

and had the highest expression levels in the leaf tissue. However, these genes were not directly functionally related. The biological consequence and molecular mechanisms underlying this discovery require further investigation.

Limitations of current work and future plan

The current study represents the first report of the most comprehensive set of fatty acid biosynthesis genes in *S. miltiorrhiza*. However, because of the unavailability of the genome sequence of *S. miltiorrhiza*, all fatty acid biosynthesis related genes in *S. miltiorrhiza* were not identified in the current study. Consequently, the aim of future studies should be to identify all genes related to fatty acid biosynthesis in *S. miltiorrhiza* with full-length coding sequences. Once the complete reservoir of sequences in *S. miltiorrhiza* has been obtained and the full–length sequences have been confirmed, functional characterization of these genes will be the next logical step. The identification of genes involved in fatty acid biosynthesis pathway not only facilitates foundational studies but will also allow researchers to improve the fatty acid production of *S. miltiorrhiza* through metabolic engineering.

Materials and Methods

Plant materials

Two-year-old *S. miltiorrhiza* Bunge cv. '99-3' plants were grown in the field of the Institute of Medicinal Plant Development, Beijing, China. Fresh leaves, roots, and flowers were collected from three individual plants and immediately stored at -80 °C until use. Total RNA was extracted from each tissue sample using an RNAprep pure plant kit (Tiangen, China).

Identification of genes related to fatty-acid synthesis

The RNA-Seq data could be obtained from NCBI using the accession numbers SRR1043988, SRR1045051, and SRR1020591. The experimental procedure for the production of the RNA-Seq data and the assembly of the RNA-Seq reads into a unigene dataset had been described elsewhere (Chen et al., 2014a). To identify genes related to fatty-acid synthesis from the unigene dataset, we retrieved sequences for *A. thaliana* proteins that are involved in fatty acid biosynthesis from the *A. thaliana* Acyl-Lipid Metabolism Website (ARALIP) as query sequences (Samuels et al., 2013). The query dataset was used to scan the unigene dataset using BLASTX program, with an E-value <= 1e−5 as cutoff. A total of 3348 matches were identified including 975 unique transcripts, some of which might belong to the same unigene. The best hit sequence (E-value <= 1e-5) for each query sequence was selected using custom Perl scripts. As many of these best-hit transcripts represent alternative sliced transcripts for the same unigene, the longest transcript was selected as the representative transcript for each unigene, resulting 391 representative transcripts. These transcripts were then subjected to GetORF analysis (Emboss v. 6.2). For a sequence with multiple open reading frames, the longest ORFs were selected and translated into protein sequences, which were used to search against Swissprot database

(version number 2013_11) using BLASTP program with E-value <= 1e-5 and identity >= 40%. The top 10 hit sequences were retrieved and aligned with the query protein sequence. After manual examination of the multiple sequence alignment, full-length fatty acid biosynthesis-related gene proteins were identified based on the criteria that the corresponding protein starts with amino acid "M" and stops at a stop codon, and the length of the protein is similar to those of its homologs.

Analysis of sequence features

The theoretical isoelectric points and molecular weights were predicted using the Compute pI/MW tool from the EMBOSS explorer (http://emboss.bioinformatics.nl/) (Rice et al., 2000). Localizations of deduced proteins were predicted using the TargetP 1.1 server (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al., 2000). Protein sequences were searched against Pfam database to determine their domain organization (Finn et al., 2013). Alignments of each *S. miltiorrhiza* gene and its top 10 homologs were generated using the Clustal W program (Thompson et al., 1994). Phylogenetic analyses were conducted using MEGA version 6 (Tamura et al., 2013). The neighbor-joining phylogenetic tree was constructed using the full-length protein sequences with default parameters. Bootstrap analysis with 1,000 replicates was used to establish the confidence level of tree branches.

Expression quantification for genes related to fatty acid biosynthesis using RNA-Seq data

Expression levels of fatty acid biosynthesis-related unigenes were calculated using the RESM program implemented in the Trinity program pipeline (Haas et al., 2013) and were represented as numbers of FPKM. Unigenes having an FPKM ≤1 were considered unexpressed. The FPKM values were added with 1, log transformed and centered with the mean expression levels at three different tissues, and subjected to hierarchical analysis using JMP (Version 10, SAS, North Carolina).

Validation of RNA-Seq results using qRT-PCR

For each tissue type, RNA samples extracted from the corresponding tissues of three individual plants were used as biological replicates for qRT-PCR analyses. Each biological replicate had three technical replicates. For each gene, the mean expression profiles of all biological and technical replicates for each tissue type were calculated. Reverse transcription was performed using a PrimeScript RT-PCR Kit (Takara). The reverse transcription product was amplified in triplicate reactions with an UltraSYBR Mixture kit (Beijing CoWin Biotech, China) using an ABI 7500 Fast instrument (Applied Biosystems, USA). Gene-specific primers were designed using PrimerQuest (http://www.idtdna.com/Primerquest/Home/Index). Melting curve analyses were performed to verify the specificity of the amplification. An actin gene (Accession no.: ADK11998) was chosen as the endogenous control. The relative expression profiles were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Statistical analysis

The statistical significance of differentially expressed genes were determined by EdgeR as implemented in the Trinity analysis pipeline (Robinson et al., 2010). The statistical methods used to validate the expression levels obtained by RNA-Seq experiments by qRT-PCR experiments have been described in previous study (Chen et al., 2014a).

Conclusions

Through the analyses of RNA-Seq data from leaf, flower and root tissues of *S. miltiorrhiza*, a total of 76 candidate genes related to fatty acid biosynthesis in *S. miltiorrhiza* were identified. These genes were grouped into 29 gene families. Twenty-seven of these gene families encode enzymes covering all steps of the fatty acid biosynthesis pathways. The other two gene families encode transcript factors associated with the pathway. The expression levels of these genes were quantified and validated by qRT-PCR. These genes were then clustered into eight groups based on their expression profiles across the three tissue types. Differentially expressed genes were identified, which might represent potential regulatory switch for fatty acid biosynthesis. Data obtained from this study has laid the foundation for future functional study of the genes involved in the fatty acid synthesis pathway in *S. miltiorrhiza*.

Acknowledgements

We thank Dr. Bin Wu for helpful comments. This work was supported by funds from National Science Foundation of China granted to C. Liu (No. 81373912), H. Chen (No. 81202859), Program for Changjiang Scholars and Innovative Research Teams in Universities of Ministry of Education of China (No. IRT1150), and Program for Innovative Research Team in IMPLAD (PIRTI, IT1305). The grantors have no role in the study design, data collection, analysis, decision to publish, and preparation of the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

Barker GC, Larson TR, Graham IA, Lynn JR, King GJ (2007) Novel insights into seed fatty acid synthesis and modification pathways from genetic diversity and quantitative trait Loci analysis of the *Brassica* C genome. Plant Physiol. 144: 1827-1842.

Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B (2007) *WRINKLED1* specifies the regulatory action of *LEAFY COTYLEDON2* towards fatty acid metabolism during seed maturation in *Arabidopsis*. Plant J. 50: 825-838.

Blatti JL, Michaud J, Burkart MD (2013) Engineering fatty acid biosynthesis in microalgae for sustainable biodiesel. Curr Opin Chem Biol. 17: 496-505.

Bonaventure G, Salas JJ, Pollard MR, Ohlrogge JB (2003) Disruption of the FATB gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. Plant Cell. 15: 1020-1033.

Brown AP, Affleck V, Fawcett T, Slabas AR (2006) Tandem affinity purification tagging of fatty acid biosynthetic enzymes in *Synechocystis sp.* PCC6803 and *Arabidopsis thaliana*. J Exp Bot. 57: 1563-1571.

Cernac A, Benning C (2004) WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*. Plant J. 40: 575-585.

Chan DI, Vogel HJ (2010) Current understanding of fatty acid biosynthesis and the acyl carrier protein. Biochem J. 430: 1-19.

Chen H, Wang FW, Dong YY, Wang N, Sun YP, Li XY, Liu L, Fan XD, Yin HL, Jing YY, Zhang XY, Li YL, Chen G, Li HY (2012) Sequence mining and transcript profiling to explore differentially expressed genes associated with lipid biosynthesis during soybean seed development. BMC Plant Biol. 12: 122.

Chen H, Wu B, Nelson DR, Wu K, Liu C (2014a) Computational Identification and Systematic Classification of Novel Cytochrome P450 Genes in *Salvia miltiorrhiza*. PLoS One. 9: e115149.

Chen H, Zhang J, Yuan G, Liu C (2014b) Complex Interplay among DNA Modification, Noncoding RNA Expression and Protein-Coding RNA Expression in *Salvia miltiorrhiza* Chloroplast Genome. PLoS One. 9: e99314.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 300: 1005-1016.

Feria Bourrellier AB, Valot B, Guillot A, Ambard-Bretteville F, Vidal J, Hodges M (2010) Chloroplast acetyl-CoA carboxylase activity is 2-oxoglutarate-regulated by interaction of PII with the biotin carboxyl carrier subunit. Proc Natl Acad Sci U S A. 107: 502-507.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M (2013) Pfam: the protein families database. Nucleic Acids Res.

Gokhale RS, Sankaranarayanan R, Mohanty D (2007) Versatility of polyketide synthases in generating metabolic diversity. Curr Opin Struct Biol. 17: 736-743.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protocols. 8: 1494-1512.

Issartel JP, Koronakis V, Hughes C (1991) Activation of Escherichia coli prohaemolysin to the mature toxin by acyl carrier protein-dependent fatty acylation. Nature. 351: 759-761.

Johnston ML, Luethy MH, Miernyk JA, Randall DD (1997) Cloning and molecular analyses of the *Arabidopsis thaliana* plastid pyruvate dehydrogenase subunits. Biochim Biophys Acta. 1321: 200-206.

Jordan SW, Cronan JE, Jr. (1997) A new metabolic link. The acyl carrier protein of lipid synthesis donates lipoic acid to the pyruvate dehydrogenase complex in Escherichia coli and mitochondria. J Biol Chem. 272: 17903-17906.

Kachroo A, Kachroo P (2009) Fatty Acid-derived signals in plant defense. Annu Rev Phytopathol. 47: 153-176.

Koglin A, Walsh CT (2009) Structural insights into nonribosomal peptide enzymatic assembly lines. Nat Prod Rep. 26: 987-1000.

Konishi T, Shinohara K, Yamada K, Sasaki Y (1996) Acetyl-CoA carboxylase in higher plants: most plants other than gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. Plant Cell Physiol. 37: 117-122.

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, Debono A, Durrett TP, Franke RB, Graham IA, Katayama K, Kelly AA, Larson T, Markham JE, Miquel M, Molina I, Nishida I, Rowland O, Samuels L, Schmid KM, Wada H, Welti R, Xu C, Zallot R, Ohlrogge J (2013) Acyl-lipid metabolism. Arabidopsis Book. 11: e0161.

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12: 323.

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 25: 402-408.

Lu S, Zhao H, Parsons EP, Xu C, Kosma DK, Xu X, Chao D, Lohrey G, Bangarusamy DK, Wang G, Bressan RA, Jenks MA (2011) The glossyhead1 allele of ACC1 reveals a principal role for multidomain acetyl-coenzyme A carboxylase in the biosynthesis of cuticular waxes by *Arabidopsis*. Plant Physiol. 157: 1079-1092.

Ma Y, Yuan L, Wu B, Li X, Chen S, Lu S (2012) Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. J Exp Bot. 63: 2809-2823.

Maeo K, Tokuda T, Ayame A, Mitsui N, Kawai T, Tsukagoshi H, Ishiguro S, Nakamura K (2009) An AP2 - type transcription factor, *WRINKLED1*, of *Arabidopsis thaliana* binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. The Plant Journal. 60: 476-487.

Mentzen WI, Peng J, Ransom N, Nikolau BJ, Wurtele ES (2008) Articulation of three core metabolic processes in *Arabidopsis*: fatty acid biosynthesis, leucine catabolism and starch metabolism. BMC Plant Biol. 8: 76.

Misra A, Surolia N, Surolia A (2009) Catalysis and mechanism of malonyl transferase activity in type II fatty acid biosynthesis acyl carrier proteins. Mol Biosyst. 5: 651-659.

Mou Z, He Y, Dai Y, Liu X, Li J (2000) Deficiency in fatty acid synthase leads to premature cell death and dramatic alterations in plant morphology. Plant Cell. 12: 405-418.

Nikolau BJ, Ohlrogge JB, Wurtele ES (2003) Plant biotin-containing carboxylases. Arch Biochem Biophys. 414: 211-222.

Ohlrogge J, Browse J (1995) Lipid biosynthesis. Plant Cell. 7: 957-970.

Ohlrogge JB, Jaworski JG (1997) Regulation of Fatty Acid Synthesis. Annu Rev Plant Physiol Plant Mol Biol. 48: 109-136.

Pidkowich MS, Nguyen HT, Heilmann I, Ischebeck T, Shanklin J (2007) Modulating seed beta-ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil. Proc Natl Acad Sci U S A. 104: 4742-4747.

Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, Yao H, Sun C, Li X, Li C, Liu J, Xu H, Chen S (2013) The complete chloroplast genome sequence of the medicinal plant Salvia miltiorrhiza. PLoS One. 8: e57607.

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16: 276-277.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26: 139-140.

Samuels KMS, Wada H, Welti R, Xu C, Zallot R, Ohlrogge J (2013) Acyl-Lipid Metabolism. The Arabidopsis Book. 1: 70.

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nat Genet. 37: 501-506.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30: 2725-2729.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680.

Wenping H, Yuan Z, Jie S, Lijun Z, Zhezhi W (2011) *De novo* transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. Genomics. 98: 272-279.

White SW, Zheng J, Zhang YM, Rock (2005) The structural biology of type II fatty acid biosynthesis. Annu Rev Biochem. 74: 791-831.

Yang L, Ding G, Lin H, Cheng H, Kong Y, Wei Y, Fang X, Liu R, Wang L, Chen X, Yang C (2013) Transcriptome analysis of medicinal plant Salvia miltiorrhiza and identification of genes related to tanshinone biosynthesis. PLoS One. 8: e80464.

Yuan Y, Sachdeva M, Leeds JA, Meredith TC (2012) Fatty acid biosynthesis in Pseudomonas aeruginosa is initiated by the FabY class of beta-ketoacyl acyl carrier protein synthases. J Bacteriol. 194: 5171-5184.