Integrated ~~Analysis~~analysis of lncRNAs, mRNAs, and TFs to ~~Identify Network Module Underlying Diterpenoid Pathway~~identify network modules underlying diterpenoid biosynthetic pathway in *Salvia miltiorrhiza*

Lin Wang[1,2], Peijin Zou[1,2], Fang Liu[1,2], Rui Liu[1,2], Zhu-Yun Yan[1,2] and Xin Chen[1,2]

[1] School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China
[2] Key Laboratory of Characteristic Chinese Medicinal Resources in Southwest, Chengdu, Sichuan, China

Corresponding Author:
Xin Chen[1,2]

Email address: chenxin@cdutcm.edu.cn

Abstract

With the emergence and development of high-throughput sequencing technology, a large number of long non-coding RNAs (lncRNAs) ~~have been~~were found to be involved in a variety of biological processes, including the synthesis of secondary metabolites. LncRNA may function by interacting with different biomolecules. The construction of interacting biomolecular networks in favor of understanding their functions. *Salvia miltiorrhiza* Bunge is an important medicinal plant in China, and its diterpene tanshinone is one of its main medicinal components. In this study, the transcriptome ~~sequencing~~ data of *S. miltiorrhiza* at the stage of post-anthesis tanshinones accumulation were used for analysis. According to NR, Swiss-Prot, KEGG, and KOG databases, transcripts without coding ability were screened out, and ~~then~~ the transcripts with sequence length greater than 200 nucleotides (nt) and open reading frame (ORF) < 100 amino acids (aa) were retained. ~~Afterwards~~Afterward, the coding potential ~~value~~values of each transcript ~~was~~were calculated using CPC2 and PLEK software, and scores ~~>~~≥ 0 were discarded. Finally, 6,651 lncRNAs were selected. In transcriptome annotation, we obtained ~~66~~57 diterpenoid biosynthetic pathway genes, including ~~32~~46 diterpene pathway genes and ~~34~~11 transcription factors (TFs), belonging to five TF families: bHLH, ERF, GRAS, MYB, and WRKY. The candidate lncRNAs, and diterpenoid biosynthetic pathway genes ~~and these TFs~~ carried out genomic location analysis and co-expression network construction. ~~In the location analysis, 199 lncRNA-mRNA/TF pairs with adjacent positions were~~We obtained~~, in which~~ 23 candidate lncRNA-mRNA/TF ~~pairs with~~

Formatado: Fonte: Times, 12 pt

co-location and co-expression ~~correlation greater than 0.4 were obtained.~~pairs. Two candidate lncRNA-mRNA/TF pairs were found in the co-expression network, ~~and their~~which included hub genes~~ were:~~ *Sm0012648* and *Sm0037093*~~, respectively.~~. In order to further explore the relationship between these genes, *S. miltiorrhiza* was induced by methyl jasmonate (MeJA). By MeJA-induced time-series expression analysis, 13 genes expressed differentially, and 3 lncRNA-mRNA and/or TF network modules were screened in candidate lncRNA-mRNA/TF pairs. This study revealed the relationship among lncRNAs, mRNAs, and TFs, and provided ~~a~~ new insight into the ~~relationship between lncRNA and the~~ biosynthetic pathway of *S. miltiorrhiza* diterpenoids.

Introduction

~~LncRNA is~~LncRNAs are defined as ~~an~~ RNA ~~that has an absence of~~transcripts with little or no potential for protein-coding capacity, and ~~greater than~~with at least 200 nt in size (Ponting, Oliver & Reik, 2009; Nagano & Fraser, 2011~~);~~ Palazzo & Koonin, 2020). They are~~. Is~~ usually transcribed by RNA polymerase II (Rinn & Chang, 2020). The functional forms of lncRNA are various and very complicated. ~~LncRNA can be used~~LncRNAs have been documented to serve as ~~a~~ molecular ~~scaffold, a~~scaffolds, guide ~~molecule, a~~molecules, molecular ~~sponge~~sponges and decoys, ~~a precursor~~precursors of microRNAs (miRNAs) and other small RNAs, or as miRNA target mimics (TMs) to regulate gene expression at multiple levels (epigenetic regulation, transcriptional regulation, and posttranscriptional regulation, etc.) (Franco-Zorrilla et al., 2007; ~~(~~Mercer, Dinger & Mattick, 2009; Wang & Chekanova, 2017).

The pharmacological activity of ~~plant are~~the plants is related to their secondary metabolites (Li et al., 2020). More and more studies ~~had~~have shown that lncRNA can regulate the secondary metabolism of organisms to some extent. For example, mLncR8 putatively ~~regulates~~regulated terpenoid biosynthesis, and mLncR31 ~~is~~was involved in the biosynthesis of the isoprenoid side chain of ubiquinone and plastoquinone in *Digitalis purpurea* (Wu et al., ~~2012)~~2012a). LncRNAs might ~~regulation of~~regulate genes in the phenylpropanoid pathway of *Populus tomentosa* (Zhou et al., 2017). ~~LncRNAs might involved in lignin biosynthesis pathway in *Populus* (Quan et al., 2019).~~ LncRNAs were also found to be involved in rubber biosynthesis in *Eucommia ulmoides* (Liu et al., 2018). LncRNAs were possibly involved in the different fatty acid synthesis and lipid metabolism through post-transcriptional regulation in tree peony seeds (Yin et al., 2018). ~~LncRNAs, MLNC3.2 and MLNC4.6 function as endogenous target mimics (eTMs) for miR156a and prevent cleavage of *SPL2-like* and *SPL33* transcription factors by miR156a during light-induced anthocyanin biosynthesis in apple fruit (Yang et al., 2019)~~LncRNAs might involve in lignin biosynthesis pathway in *Populus* (Quan et al., 2019).

~~The researchs~~Studies have suggested that ~~many~~ lncRNAs can act as local regulators, ~~that~~and lncRNA expression is correlated with the expression of nearby genes (Guil & Esteller, 2012; ~~(~~Engreitz et al., 2016). In view of the fact that transcription regulatory elements such as

Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Times, 12 pt
Formatado: Fonte: Times, 12 pt
Formatado: Fonte: Times, 12 pt
Formatado: Cor da fonte: Automática
Formatado: Normal
Formatado: Cor da fonte: Automática

enhancers and promoters can transcribe bi-directionally (Sigova et al., 2015; Kopp & Mendell, 2018), ~~the~~ lncRNAs can regulate the expression of upstream and downstream genes. ~~Theoretically,~~ Co-expressed genes ~~belonging to~~are usually members of the same protein complex or metabolic pathway ~~also can be recovered as a part of~~ and are functionally controlled by the same transcriptional regulatory program, and the genes or proteins within the co-expression network may have the same expression patterns (Liao et al., 2011; ~~expressed modules across a vast array of conditions~~ (Rao & Dixon, 2019). ~~And lncRNA~~LncRNAs can also act on TFs, eg. lncRNA can be used as ~~the~~ a transcription factor binding ~~of TFBS~~site to regulate its expression (Yu et al., 2019)~~. Therefore, the location of lncRNA, the pathway genes and TF of co-expression pattern can be combined to explore their functions~~.

RNA localization is the main mechanism to control cell function, and RNA subcellular localization can help us to better understand RNA function (Lécuyer et al., 2007). The function of lncRNA mainly depends on RNA physical interaction, and the local concentration and subcellular localization of lncRNAs determine their molecular interaction network (Carlevaro-Fita & Johnson, 2019). Most lncRNAs have significant tissue-specific expression, and their expression may be limited to specific cell types or induced by specific signals (Liu et al., 2012). Subcellular localization plays an important role in studying the function and mechanism of lncRNAs (Chen, 2016).

Previous studies have been ~~done about~~made on lncRNAs in *S. miltiorrhiza*. In the study of ~~mlncRNAs~~lncRNAs in *S. miltiorrhiza* (Li, Shao & Lu, 2015), ~~mlncRNAs~~lncRNAs responded to Ag^+ solution and Yeast Extract (YE) were identified in the root of *S. miltiorrhiza*, and differentially expressed in leaves under MeJA treatment, besides indicated that some ~~mlncRNAs~~lncRNAs were differentially expressed in different tissues of roots, flowers, and leaves. Jiang *et al*. ~~(Jiang et al., 2021)~~ identified the differential expression of natural antisense transcripts (NAT) with ~~ployA~~polyA tail in different tissues of *S. miltiorrhiza*, and found *cis*-NAT of *SmKSL1*, they showed a high co-expression relationship and might participate in tanshinones synthesis in a *cis*-regulation (Jiang et al., 2021). However, although lncRNAs reportedly play important roles in *S. miltiorrhiza*, knowledge of the functions of lncRNAs in the diterpenoid biosynthetic pathway of *S. miltiorrhiza* is limited.

The medicinal plant *S. miltiorrhiza* produces a variety of diterpenoids (Ma et al., 2015). Wherein the tanshinones are the main bioactive compounds of *S. miltiorrhiza*, which mainly accumulated in the roots of *S. miltiorrhiza* (Xu et al., 2015; Chang et al., 2019). Another ~~diterpenoids~~diterpenoid in *S. miltiorrhiza* is the plant hormone gibberellin (GA), which is one of the five classic plant hormones (Brockdorff, 1998). They are ~~all produced by~~generally derived from the ~~generic diterpenoid precursor~~linear primary metabolite geranylgeranyl diphosphate (GGPP) ~~under the action of genes~~by diterpene synthases (diTPS). These enzymes undergo complex electrophilic cyclizations and/or rearrangements leading to diverse backbone structures

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Times, 12 pt

Formatado: Cor da fonte: Preto

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Itálico

(Gong, Zeng & Chen ~~TFs (including bHLH, GRASR, 2R3-MYB, WRKY and AP2/ERF, etc (Li & Lu~~, 2014~~); Zhang et al., 2015; Li et al., 2015, 2019; Ji et al., 2016).) in different branching pathways~~.

MeJA is a hormone involved in plant signal transduction, which is considered to play an indispensable role as a second messenger in the induction process leading to the accumulation of secondary metabolites. Therefore, it is often used as an inducer to explore the regulation mechanism of biosynthesis (Gundlach et al., 1992; Wasternack, 2007). MeJA induced the synthesis of defensive compounds and the accumulation of active compounds, which is an effective inducer of tanshinones (Gao et al., 2009; Liang et al., 2012) and phenolic acids (Xiao et al., 2011) in *S. miltiorrhiza*. MeJA is used for genome-wide identification and characterization of novel diterpene biosynthesis genes in *S. miltiorrhiza* (Ma et al., 2012).

This study aims to explore the lncRNA-mRNA/TF network modules related to the biosynthesis of diterpenoids in *S. miltiorrhiza*. Thus, based on the transcriptome data of the tanshinones accumulation stage of *S. miltiorrhiza*, the candidate lncRNAs were identified according to the definition and characteristics of lncRNAs. The expression correlation and genomic location of diterpenoids biosynthesis genes/TFs and candidate lncRNAs were analyzed to obtain lncRNA-mRNA/TF pairs. By constructing a co-expression network, the hub genes were obtained. Further, under the induction of MeJA, the time-series expression patterns of the candidate lncRNA-mRNA/TF pairs were analyzed, obtaining differentially expressed lncRNAs and mRNAs/TFs in response to MeJA induction and describing the dynamic expression patterns of them under MeJA induction, thereby constructing lncRNA-mRNA and/or TF correlation modules.

~~Materials~~Material & Methods

S. miltiorrhiza genomic and transcriptome data ~~(Data collection)~~

In this study, the *S. miltiorrhiza* genome datas were downloaded from the *S. miltiorrhiza* Genomics Database (http://www.ndctcm.org/shujukujieshao/2015-04-23/27.html (Xu et al., 2016) and NCBI (National Center for Biotechnology Information) BioProject: PRJNA682867 (Ma et al., 2021)~~). We gathered 24 *S. miltiorrhiza* illumina RNAseq datasets that were sampled from root tissues during tanshinones accumulation stage (Zhou et al.). Transcription datas of *S. miltiorrhiza* were from NCBI BioProject (Zhou et al.), and the accession number was PRJNA712174.~~.

The RNA-seq data were obtained from our previously published study (Zhou et al., 2021) with an accession number assigned to PRJNA712174. These data were gathered from a total of 24 RNA-seq experiments of *S. miltiorrhiza* root tissues during the tanshinones accumulation stage and each with three biological replicates. Transcriptome *de novo* assembly was carried out with the short reads assembling program Trinity (Grabherr et al., 2011). Gene quantification (read

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: 12 pt

count and normalized expression value as Fragments Per Kilobase Million—FPKM) was obtained using RSEM (Li & Dewey, 2011) with default parameters.

Pipeline for ~~lncRNAs~~lncRNA identification

In order to identify ~~reliable~~ lncRNAs, we conducted screening according to the following steps. The first step was to ~~fiter~~filter the coding transcripts. Coding transcripts were annotated based on the following databases: Nr (NCBI non-redundant protein sequences, https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/); Swiss-Prot (A most complete annotated and refined protein database, https://www.uniprot.org/); KEGG (Kyoto Encyclopedia of Genes and Genomes, ~~http://www.kegg.jp/)~~http://www.kegg.jp/) and COG/KOG (Clusters of Orthologous Groups of proteins, eukaryotes are generally called KOG databases, https://www.ncbi.nlm.nih.gov/KOG/). To rule out the possibility that transcripts had protein-coding capability, we then used the following six filters to shortlist the bona fide lncRNAs from the obtained candidate transcripts: (1) transcripts with annotation information in the above database were removed; (2) transcripts shorter than 200 nt with an ORF longer than 100 aa were discarded, found and extracted ORFs on getorf (http://emboss.bioinformatics.nl/cgi-bin/emboss/getorf), selection of ORF length < 100 aa by R (Team, 2010) ~~(version 4.2.1)~~(version 4.2.1) script; (3) transcripts were searched against the Pfam database (Punta et al., 2012) (http://pfam.xfam.org/) ~~(http://pfam.xfam.org/)~~ by HMMER to remove transcripts possible containing known protein domain; (4) the protein coding-potential of each transcript was calculated using PLEK (Li, Zhang & Zhou, 2014) and Coding Potential Calculator 2 (CPC2, ~~http://cpc2.cbi.pku.edu.cn)~~http://cpc2.cbi.pku.edu.cn) (Kang et al., 2017) those with PLEK and CPC2 scores ~~>~~≥ 0 were discarded; (5) the remaining transcripts were searched against the Rfam database ~~(Mistry et al., 2021) (http://pfam.xfam.org/)~~(available at http://rfam.xfam.org) to identify RNA family. LncRNAs were expressed at low levels, and their expression became detectable only in a few tissues, mutants, and/or in plants subjected to certain treatments (Liu et al., 2012). FPKM value less than 0.05 was used as the standard for low expression levels (Li et al., 2016). The transcripts that remained were regarded as expressed candidate lncRNAs. ~~For the purpose of screening~~To screen out the lncRNAs with regulatory functions, ~~the lncRNAs with low expression levels,~~ tRNA and rRNA in RNA family analysis were removed~~, gene expression levels were estimated by RSEM (http://deweylab.github.io/RSEM/)~~.

Characteristics and conservation analysis of *S. miltiorrhiza* lncRNAs

To gain more understanding of the lncRNA ~~To find out the difference~~ in ~~sequence between lncRNA and mRNA,~~*S. miltiorrhiza* we compared ~~and analyzed~~ several different features of the ~~sequence length and GC content of candidate~~ lncRNAs and mRNAs ~~in transcriptome data using~~ : GC content and transcript length. GC content and transcripts length were determined by R ~~and~~ (version 4.2.1), and statistics by Excel.

Comentado [d1]: which package?

Formatado: Fonte: Times, 12 pt

Formatado: Cor da fonte: Preto

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Times, 12 pt, Não Negrito

~~The genomic locus, exon sequence and promoter region of lncRNA also exhibit homology conservation, and these lncRNA~~LncRNA with homology conservation ~~are~~is more likely to have function (Guttman et al., 2009). ~~Therefore, we performed conservative analysis on lncRNAs from different species and between species to identify~~To explore lncRNA conservation, all the ~~most functional lncRNAs. The lncRNA homologs in related species and distantly related specie genomes were investigated based on the primary sequence similarity. The genomes of~~ lncRNA sequences identified here were aligned with BLAST+ (Camacho et al., 2009) (blast-2.11.0+) against the genome sequences of *Lamiaceae* family: *Salvia splendens*, *Salvia hispanica*, *Mentha longifolia*, *Scutellaria baicalensis*, *Pogostemon cablin*, *Sesamum indicum*, of which *Salvia splendens* and *Salvia hispanica* both belong to *Salvia* genus, *Solanaceae* family: *Nicotiana tabacum*, *Brassicaceae* family: *Brassica napus* and *Arabidopsis thaliana* ~~were~~ , *Selaginellaceae* family: *Selaginella moellendorffii*. With a cutoff E-value < 1e-10. The genomes were downloaded from NCBI databases GCF_004379255.1 (SspV2), GCF_023119035.1 (UniMelb_Shisp_WGS_1.0), GCA_001642375.2 (Mlong_CMEN585_v), GCA_005771605.1 (ASM577160v1), GCA_023678885.1 (GZUCM_PCab_1.0), GCF_000512975.1 (S_indicum_v1.0~~) and GCF~~), GCF_000715135.1 (Ntab-TN90), GCF_020379485.1 (Da-Ae), GCF_000001735.4 (TAIR10.1~~), respectively. The lncRNA sequences were aligned against the whole genome sequences of these plants with BLAST+ (Camacho et al., 2009) (blast-2.11.0+). We analyzed lncRNA conservatism by preserving E-value < 1e-10 and~~) and GCF_000143415.4 (v1.0). The lncRNAs that had coverage of > 20% of matched regions ~~in the blast comparison results matched to *S. miltiorrhiza* genome.~~

~~Precursors of miRNA and miRNA targets prediction in *S. miltiorrhiza* lncRNAs~~

~~To explore whether the candidate lncRNAs can be used as the precursor of miRNA, the known miRNA miR12112 of *S. miltiorrhiza* in miRBase (Kozomara, Birgaoanu & Griffiths-Jones, 2019) (Release 22.1, http://www.mirbase.org/) was compared with the sequence of the candidate lncRNAs (NCBI BLAST Web, Http://www.ncbi.nlm.nih.gov/blast), then predicted the secondary structure, and predicted the secondary structure of lncRNAs by using the RNAfold (Gruber et al., 2008) web server, which can be found in Vienna RNA website http://rna.tbi.univie.ac.at/. LncRNAs with classical stem-loop hairpin are considered as the putative precursors of miRNA (Zhou et al., 2017). In addition, lncRNAs can be used as eTMs of miRNA, which can interfere with the normal binding of miRNA and its target gene (Wu et al., 2013). To search for potential miRNA targets in *S. miltiorrhiza* lncRNAs, the sequences of the known miRNAs of *S. miltiorrhiza* in the miRBase and the candidate lncRNA transcript sequences~~ were ~~submitted to the psRNATarget (Dai, Zhuang & Zhao, 2018) webserver and utilized with the current default settings (2017 release). The new release has an improved scoring schema V2 to cover more validated miRNA-mRNA interactions without significantly increasing the final prediction output. Targets with an E-value less than 5.0 were retained as potential miRNA targets~~defined as conserved lncRNAs.

Formatado: Normal

Prediction lncRNAs related to the ~~diterpene~~diterpenoid biosynthesis of *S. miltiorrhiza*

~~The lncRNA can regulate the expression of adjacent genes and thus to carry out functional research (Kopp & Mendell, 2018). Study based on previous studies in *Brassica rapa* (Huang et al., 2018) and rice (Wang et al., 2021). The whole genome of *S.*~~ The synthesis pathway of diterpenoids in *S. miltiorrhiza* included several stages. The first stage leads to the synthesis of the universal isoprene precursor isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) through the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway and/or the mevalonate (MVA) pathway. In the second stage, the intermediate diphosphate precursors, including geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and GGPP are synthesized under the catalysis of isoprenyl diphosphate synthases (IDSs), including geranyl diphosphate synthase (GPPS), farnesyl diphosphate synthase (FPPS), and geranylgeranyl diphosphate synthase (GGPPS). The last stage involves the formation of diverse diterpenoids under the catalysis of terpene synthases (TPSs), such as copalyl diphosphate synthase (CPS) and kaurene synthase (KS) catalyze the formation of miltiradiene (Kai et al., 2010, 2011; Lu, 2021), *ent*-copalyl diphosphate synthase (*ent*-CPS) (Shimane et al., 2014) and *ent*-kaurene synthase (*ent*-KS) are involved in the conversion of GGPP to the tetracyclic hydrocarbon intermediate *ent*-kaurene (Yamaguchi, 2008; Shimane et al., 2014). Then tanshinones and GAs are formed by cytochrome P450 monooxygenases (P450s) and 2-oxoglutarate-dependent dioxygenases (2ODDs) modification. Our research focused on the key enzymes downstream of IPP (Fig. S1).

The whole genomes of *S. miltiorrhiza* from Xu *et al.* (2016), available at http://www.ndctcm.org/shujukujieshao/2015-04-23/27.html (Xu et al., ~~2016)~~2016) and *S.* ~~(available in http://www.ndctcm.org/shujukujieshao/2015-04-23/27.html) and *S.*~~ *miltiorrhiza* genome GCA_016432925.1 (NRC_Smil_1.0). BLAST+ (BLAST-2.11.0+) ~~were~~was used for the analysis of genomic location analysis of candidate lncRNAs~~. In the present study~~ with E value < 1e-5. Study based on previous studies in *Brassica rapa* (Huang et al., 2018) and rice (Wang et al., 2021), we searched the adjacent lncRNAs and mRNAs/TFs in the 100 kilobases (kb) upstream and downstream regions of the genome.

~~Co-expressed genes are usually members of the same protein complex or metabolic pathway and are functionally controlled by the same transcriptional regulatory program, and the genes or proteins within the co-expression network may have the same expression patterns (Liao et al., 2011; Rao & Dixon, 2019).~~ We used the guide-gene approach (Aoki, Ogata & Shibata, 2007). ~~Firstly, the low-expression transcript was filtered by the customized R script, correlation analysis was performed using the FPKM matrix, and the pearson~~Pearson correlation coefficient (PCC) was calculated to quantify the correlation among the different predictors. PCC between the candidate lncRNAs and the ~~diterpene~~diterpenoid biosynthetic pathway genes and TFs were calculated based on their expression levels using the psych (Revelle, 2022) package in R. And

Formatado: Fonte: 12 pt, Não Negrito, Realce

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Times, 12 pt

|PCC| ≥≥ 0.4 and p ≤ 0.05 were used as the threshold value for the second round of candidate pairs screening. At last, we defined two genes as a co-expressed and co-located pair if they were spaced by < 100 kb and co-expressed.

Construction of the diterpenoid pathway related lncRNA-mRNA-TF networks and finding hub genes

Only genes with |PCC| ≥≥ 0.8 and p ≤ 0.05 were considered to be potential target genes of lncRNA, forming lncRNA-mRNA-TF co-expression pairs. The lncRNA-mRNA-TF co-expression network related to the diterpenoid biosynthesis of *S. miltiorrhiza* was constructed using Cytoscape 3.2 (Shannon et al., 2003) to visualize the interaction network.

Precursors of miRNA and miRNA targets prediction in *S. miltiorrhiza* lncRNAs

~~Differential expression analysis of lncRNAs~~

~~To evaluate the gene expression relationship between the two periods, we used the count matrix and R package (4.1.2) Deseq2 (Love, Huber & Anders, 2014) to filter out the transcripts with the lower expression levels and detected differentially expressed lncRNAs. Transcripts with |log2 FoldChange| > 1 and padj ≤ 0.05 were identified as differentially expressed lncRNAs.~~

To explore whether the candidate lncRNAs may act as the precursors of miRNAs, the 203 known *S. miltiorrhiza* miRNAs in miRBase (Kozomara, Birgaoanu & Griffiths-Jones, 2019) (Release 22.1, http://www.mirbase.org/) and PmiREN2.0 (Guo et al., 2022) (https://pmiren.com/) were aligned to the sequences of the candidate lncRNAs. And predicted the secondary structure of lncRNAs by RNAfold (Gruber et al., 2008) (http://rna.tbi.univie.ac.at/). LncRNAs with classical stem-loop hairpins were regarded as the putative precursors of miRNA (Zhou et al., 2017). A lncRNA referred to as *IPS1* (*INDUCED BY PHOSPHATE STARVATION1*) acts as a TM of miR-399 in *Arabidopsis thaliana* (Franco-Zorrilla et al., 2007). miRNA targets and lncRNAs have highly similar miRNA-binding sites, thereby the miRNA can be sequestered by the lncRNA. This mechanism describing the inhibition of miRNA activity is defined as “target mimicry”, also known as “target mimics” in plants (Paschoal et al., 2017). Three kinds of prediction software were used to determine the miRNAs targeted to candidate lncRNAs. The first was TAPIR (Bonnet et al., 2010) (http://bioinformatics.psb.ugent.be/webtools/tapir), The server offers the possibility to search for plant miRNA targets using a fast and precise algorithm (score ≤ 4, free energy ratio ≥ 0.7). The second was psRobot (Wu et al., 2012b) (Version 1.2) with default parameters, a widely used online miRNA target prediction tool. The third was psRNATarget (2017 release) with default settings, psRNATarget was developed to identify plant sRNA targets by (i) analyzing complementary matching between the sRNA sequence and target mRNA sequence using a predefined scoring schema and (ii) by evaluating target site accessibility (Dai, Zhuang & Zhao, 2018), targets with an E value less than 5.0 were retained as potential miRNA targets.

Subcellular localization of lncRNAs

~~RNA localization is the main mechanism to control cell function, and RNA subcellular localization can help us to better understand RNA function (Lécuyer et al., 2007). Compared with mRNA, since the final product of lncRNA is an RNA, the function of lncRNA mainly depends on RNA physical interaction (Carlevaro-Fita & Johnson, 2019), and the local concentration and subcellular localization of lncRNAs determine their molecular interaction network (Carlevaro-Fita & Johnson, 2019). Therefore, subcellular localization plays an important role in studying the function and mechanism of lncRNAs (Chen, 2016).The subcellular localization of lncRNAs was predicted by lncLocator (Cao et al., 2018).~~

The subcellular localization of lncRNAs was predicted by lncLocator (Cao et al., 2018). LncLocator is an ensemble classifier-based predictor, which adopts both *k-mer* features and high-level abstraction features generated by unsupervised deep models and constructs four classifiers by feeding these two types of features to support vector machine and random forest, respectively. The current lncLocator can predict five subcellular localizations of lncRNAs, including cytoplasm, nucleus, cytosol, ribosome, and exosome.

Tissue-specific expression, RNA extraction, and qRT-PCR analysis ~~of lncRNAs in *S. miltiorrhiza*~~

~~Most lncRNAs have significant tissue-specific expression, and their expression may be limited to specific cell types or induced by specific signals (Liu et al., 2012). So we examine the organizational differences of the candidate lncRNAs. The previously obtained candidate lncRNAs were selected, of which 11 lncRNAs with good~~The expression patterns of lncRNAs in the co-located and co-expressed pairs among different tissues were detected by qRT-PCR (quantitative real-time Polymerase Chain Reaction). Of which 11 lncRNAs with observable expression were analyzed in one-year-old, greenhouse-grown *S. miltiorrhiza* plants. Total RNA was extracted from fresh tissue samples of four parts of *S. miltiorrhiza*: roots, stems, leaves, and flowers. RNA isolation, quantification, and qualification total RNA were isolated from the roots treated using the RNeasy plant kit (No.PH-01013-B, Foregene, Chengdu, China). RNA degradation and contamination were monitored using 1% RNase-free agarose gel electrophoresis, and the RNA purity was analyzed using a NanoPhotometer^{TM}-N60 ultra-micro spectrophotometer. The reverse transcription reaction used RT Easy^{TM} II (With gDNase) kit (Version Number: 1.0-1904, Foregene, Chengdu, China) following the instruction manual. PCR was performed in a 20 μL reaction volume containing primers, Real-Time PCR EasyTM-SYBR Green I kit (Cat.No.QP-01011/01012/01013/01014, Foregene, Chengdu, China), and diluted cDNA templates on the LineGene K Plus Real-Time PCR Detection System. Experiments were performed with three biological repetitions. The primer pairs for qRT-PCR were designed using the Primer-BLAST (Ye et al., 2012) and were shown in Table S1. *β-actin* (Yang et al., ~~and flowers, and the expression were detected by quantitative real-time Polymerase Chain Reaction~~

~~(qRT-PCR) after reverse transcription. *β-actin* (Yang et al.,~~ 2010) gene was selected as a reference since it showed stable expression in the *S. miltiorrhiza* tissues analyzed compared to others. The reaction program was as follows: 3 min at 95°C, 10 s at 95°C, 30 s at 60°C, and 40 cycles. The temperature was then gradually increased to produce melting curves for amplification specificity verification. The mean value of three replicates was normalized using *β-actin* as the ~~The arithmetic formula 2(-deltadeltaCT) method was used to achieve results for relative quantification in Excel. Three biological replicates were carried out for each experiment. Primer sequences were designed using the Primer-BLAST (Ye et al., 2012), and were reported in *Table S1*~~endogenous control. The 2(-delta-delta CT) method was used for calculating the relative expression levels of genes.

~~Plant materials, treatment condition, RNA extraction and~~Candidate co-expression co-located lncRNA-mRNA/TF pairs treated with MeJA

In order to further explore whether candidate co-expression co-located lncRNA-mRNA/TF pairs have a response to the diterpenoid biosynthetic pathway, using the qRT-PCR ~~analysis~~ method to analyze expression patterns of observable lncRNA-mRNA/TF pairs. The *S. miltiorrhiza* stools were cultured in a greenhouse under 20°C to 25°C. The plant materials were propagated from the medicinal plant garden of Chengdu University of Traditional Chinese Medicine. The plants were authenticated by Professor Xin Chen of the Chengdu University of Traditional Chinese Medicine using the morphological identification approach of the Flora of China. The plants were sprayed with 200 μM MeJA (Sigma-Aldrich Corp., St. Louis, MO) solution as mentioned in the previous report (Luo et al., 2014). After being treated with MeJA solution for 0, 6, 12, 24, and 48 hours, ~~respectively,~~ the MeJA-induced roots and the mock-treatment were collected and rinsed with water. The roots were then dried gently and quickly with absorbent paper. The cleaned roots were immediately frozen in liquid nitrogen and stored at ~~-~~-80°C until RNA isolation.

~~RNA isolation, quantification, and qualification total RNA~~ qRT-PCR was ~~isolated from~~ achieved according to the ~~roots~~process used in the tissue-specific expression analysis. Plant materials treated ~~using the RNeasy plant kit (No.PH-01013-B, FOREGENE). RNA degradation and contamination was monitored using 1% agarose gels, and the RNA purity was analyzed using NanoPhotometer TM-N60.The reverse transcription reaction used RT-EasyTM II (With gDNase) kit (Version Number: 1.0-1904, FOREGENE) following the instruction manual. The primer pairs for qRT-PCR were designed using the Primer-BLAST and were shown in Table S1. The qRT-PCR was carried out using Real-Time PCR-EasyTM-SYBR-Green I kit (Cat.No.QP-01011/01012/01013/01014, FOREGENE)~~ with MeJA dissolving media were used as a control (0 h), and three biological ~~repetitions. The reaction program was as follows: 3 min at 95°C, 10 s at 95°C, 30 s at 60°C, 40 cycles.~~replications were carried out. ~~The temperature was then gradually increased to produce melting curves for amplification specificity verification. The mean value of three replicates was normalized using *β-actin* as the internal control.~~ The 2(-~~deltadeltaCT~~delta-

Formatado: Fonte: Times, 12 pt, Não Negrito

delta CT) method was used for calculating the relative expression levels of genes. ANOVA was calculated using SPSS (Version 23.0, IBM, USA), and $p < 0.05$ ~~was~~and $p < 0.01$ were considered statistically significant. The primers were designed by Primer-BLAST. Primer sequences were reported in Table S1.

Module detection in the analysis of time-series gene expression

~~Expression~~The qRT-PCR results of ~~candidate~~the co-located and co-expressed lncRNA-mRNA/TF ~~genes in *S. miltiorrhiza*~~pairs induced by MeJA ~~as time-series gene expression data. The time-series gene expression matrix was constructed using this data.~~were used to calculate PCC ~~was calculated~~ by using the psych ~~(Revelle, 2022)~~ package ~~in R.~~. The co-expression networks were visualized with the program Cytoscape 3.2. The hub genes were identified by using the cytoHubba (Chin et al., 2014) plug-in of Cytoscape. To increase the sensitivity and specificity, we proposed Maximal Clique Centrality (MCC) to discover featured nodes. Subsequently, we constructed a network module of these hub genes.

Results

~~lncRNA~~LncRNA identification and characterization analysis

Candidate lncRNAs screening was performed in the RNA library we constructed from de novo assembled whole unigenes files (a total of 70,357 unigenes). Based on the above identification process of lncRNA, 40,010 transcripts were annotated in the RNA library, while 30,347 transcripts were not annotated. We ~~total~~ obtained a total of 21,742 transcripts with length ~~></~~> 200 nt and ORF < 100 aa. Then, 5 unigenes with annotation information matched in Pfam. A total of 21,468 ~~nocoding~~noncoding unigenes were obtained ~~by intersecting~~from the intersection of PLEK and CPC2 ~~predictions, see Figure~~prediction results (Fig. 1~~.~~). The RNA family members were identified in ~~the~~ Rfam database and a total of 91 predictive information was obtained (Table ~~*S1*~~S2, Table ~~*S2*~~S3). Filtered out ~~the low~~ expression transcripts with FPKM < 0.05, and a total of 6,651 candidate lncRNAs were ~~obtained.~~selected for further analysis (Table S4).

~~And we characterized the genomic features of *S. miltiorrhiza*~~ These 6,651 candidate lncRNAs ~~by comparing them with protein-coding mRNAs assembled in this study. The length of lncRNA in *S. miltiorrhiza*~~ ranged from 201 nt to 4,340 nt in length, of which most (approximately 68.95%) were 200 to 400 nt (average length = 391.50 nt), ~~lower than the mRNA value (average length =1514.61 nt). On average, the length of lncRNAs was also found to be significantly~~ which were shorter than mRNAs of *S. miltiorrhiza* (average length = 1,514.61 nt). The ~~in other species (Hao et al., 2015; Liu et al., 2018; Shen et al., 2018). In the distribution of~~ GC content~~, the GC content in lncRNA~~ of lncRNAs was mainly concentrated at 31.28~37.28% (accounting for 46.32%), while ~~that in mRNA was~~mRNAs were mainly concentrated at 39.28~41.28% (accounting for 36.00~~%) (Figure 2). We compared the lncRNA~~%). The mean GC content of lncRNAs was approximately 36.86%, a little lower than that of mRNA sequences ~~to obtain 134 (accounting for~~

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Não Itálico

Formatado: Fonte: Não Itálico

Formatado: Fonte: Times, 12 pt

~~2.01%) lncRNA located within 100 kb of the adjacent protein-coding gene, indicating~~(approximately 43.42%) (Fig. 2, Table S5).

Conservation analysis showed that ~~most lncRNAs were transcribed independently of the adjacent protein-coding gene.~~

~~At the sequence-level, we estimated conservation~~21.02% and 15.40% of *S. miltiorrhiza* lncRNAs were conserved in ~~other related species (~~*Salvia splendens* and ~~*Sesamum indicum*) and another distantly related specie (*Arabidopsis thaliana*) based on both sequence similarity. The conservation of *S. miltiorrhiza* lncRNAs in distantly related species is less than that in related species. We got the results for four coverage levels, the cutoff threshold was set to coverage of > 20% of matched regions. Our results indicate that 20.85% and 0.51% of *S. miltiorrhiza* lncRNAs are conserved compared with *Salvia splendens, Sesamum indicum*~~*hispanica* of the same genus, respectively. ~~But, compared with *Arabidopsis thaliana*~~However, in the same family of the different genera of *Mentha longifolia*, *Scutellaria baicalensis*, *Pogostemon cablin* and *Sesamum indicum*, conservation is 6.54%, 0.96%, 0.83% and 0.51%, respectively. Among plants of different families, only 0.~~015% conservative.~~05%, 0.02% and 0.02% of *S. miltiorrhiza* lncRNAs conserved in *Nicotiana tabacum*, *Arabidopsis thaliana* and *Brassica napus*, respectively. There was no match at all in *Selaginella moellendorffii*. The results of the conservative analysis are presented in Table ~~*S3*~~S6.

~~LncRNAs might be used as precursors or target mimics of *S. miltiorrhiza* miRNA~~

~~In this study, we identified lncRNA Smlnc0015916 as a precursor of *S. miltiorrhiza* miR12112 in the miRBase database, as can seen from the *Fig. S1A, Fig. S1B*. Twelve lncRNAs (Smlnc0015916, Smlnc0023222, Smlnc0063827, Smlnc0016149, Smlnc0053518, Smlnc0064035, Smlnc0027179, Smlnc0035109, Smlnc0002673, Smlnc0066457, Smlnc0063004, Smlnc0012094) were potential targets for *S. miltiorrhiza* miRNA in the miRBase database, as can be seen from the *Fig .S1C*. It has been reported that miR12112 regulates the post-transcription of *SmPPO* (Li et al., 2017) in *S. miltiorrhiza*, and *SmPPO* positively regulates the synthesis of Salvianolic acid B.~~

LncRNAs related to the diterpenoid biosynthetic pathway

In transcriptome annotation, we obtained 57 diterpenoid biosynthetic pathway genes, including 46 diterpene pathway genes: GPPS (Van Schie et al., 2007), FPPS, GGPPS, CPS, *ent*-CPS, KS, *ent*-KS, CYP76AH1 (Guo et al., 2013; Ma et al., 2016), CYP76AK2, CYP76AK3 (Li et al., 2021), CYP76AK5 (PU Xiangdong & WANG Lizhi, 2017), *ent*-kaurene oxidase (KO) (Hedden & Thomas, 2012), *ent*-kaurenoic acid oxidase (KAO) (Helliwell et al., 2001), GA 2-oxidase (GA2ox), GA 3-oxidase (GA3ox), GA 20-oxidase (GA20ox) (Hedden & Thomas, 2012; Du et al., 2015), and 11 transcription factors (TFs): bHLH148, ERF6, GRAS1, MYB36 and WRKY2

Formatado: Fonte: Não Itálico

Formatado: Cor da fonte: Automática, Padrão: Transparente

Formatado: Fonte: Times, 12 pt, Não Negrito

(Li & Lu, 2014; Zhang et al., 2015; Li et al., 2015, 2019; Ji et al., 2016), belonging to five TF families: bHLH, ERF, GRAS, MYB, and WRKY. The list of the above genes is shown in Table S7.

~~At the positional level, using the genome of Xu et al. as reference and BLAST for comparison, a total of 64 pairs of lncRNA-mRNA/TF located adjacent to each other were compared with E value < 1e-5 and distance within 100 kb (including ± 3 kb) as the threshold.~~

In the PCC matrix, we obtain 56,892 correlation gene pairs with |PCC| ~~≥~~≥ 0.4 ~~as the PCC threshold~~ (p ~~≤~~≤ 0.05), of which~~,~~ 47,946 pairs were positively correlated and 5,775 pairs were negatively correlated. ~~According to the correlation (greater than 0.4) between previous genes,~~ At the positional level, using the genome of Xu *et al*. (2016). as a reference. A total of 64 pairs of lncRNA-mRNA/TF were located adjacent to each other within 100 kb (including ± 3 kb), of which 9 pairs were both co-located and co-expressed. According to the genomic *gff* file, 5 of ~~which~~these were lincRNAs (long intergenic noncoding RNAs), 2 ~~of which~~ were incRNAs (intronic noncoding RNAs), and other spanning introns and exons. Using the reference genome of *S. miltiorrhiza* in the NCBI database as ~~the~~a reference, ~~and performing BLAST comparison with the same threshold,~~ a total of 135 pairs of lncRNA-mRNA/~~TFs~~TF located adjacent to each other were obtained, ~~and a total~~ of which 16 pairs were co-located and co-expressed. ~~A total of~~ We finally got 23 lncRNA-mRNA/TF pairs ~~were~~ co-located, with |PCC| greater than 0.4 (Table ~~*S4*~~S8).

LncRNA-mRNA-TF Networks and Hub Genes

The gene co-expression network was constructed with |PCC| ~~≥~~≥ 0.8 as the threshold~~,~~ and the visualization was shown in ~~(~~Fig. S2~~,~~ and Table ~~*S5*).~~S9. We obtain 1,496 correlation gene pairs with |PCC| ~~≥~~≥ 0.8 ~~as the PCC threshold~~ (p ~~≤~~≤ 0.05). We obtained 24 mRNAs with a degree > 10, which were considered hub genes. In the 23 pairs of lncRNA-mRNA/TF, 8 mRNAs were present in these hub genes. Only two mRNAs *Sm0012648* (GA2ox~~),~~) and *Sm0037093* (kaurene synthase 2, KS2) with their pair existed in the network, showing high expression correlation in their pairs, the correlations were 0.85 and 0.80 respectively.

LncRNAs ~~differentially expressed during tanshinones accumulation period in~~ might be used as precursors or target mimics of *S. miltiorrhiza* miRNAs

~~In the analysis of differential expression in lncRNAs, a total of 18 were obtained, of which, 10 lncRNAs were up-regulated and 8 were down-regulated at different times, with the rest showing a stable expression trend. Unfortunately, there were no differentially expressed genes in lncRNAs of 23 lncRNA-mRNA/TF pairs. LncRNA may only function for a short period of time, rather than for a long period of time (*Table S6*).~~

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Times, 12 pt, Não Negrito

Formatado: Fonte: Itálico

Formatado: Fonte: Não Itálico

Formatado: Fonte: 12 pt, Não Negrito

Formatado: Fonte: Não Itálico

Formatado: Fonte: Não Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

In this study, we identified 17 lncRNAs probable as 41 precursors of *S. miltiorrhiza* miRNA (Table S7). In order to improve the prediction accuracy, the intersection of three kinds of software prediction results was selected. Total identified 14 lncRNAs as potential targets of 66 miRNAs, these miRNAs can be divided into 18 families according to PmiREN2.0 (Table S10).

Subcellular localization predictions indicated that most *S. miltiorrhiza* lncRNAs were found in the cytoplasm and nucleus

Subcellular localization of lncRNA is closely related to its function. At present, there are also many studies ~~indicate~~indicating that lncRNA contains specific RNA motifs with nuclear localization (Zhang et al., 2014). Meanwhile, in the subcellular localization of lncLocator (Cao et al., 2018), a total of 3,381 ~~cytoplasm~~cytoplasms, 3,316 nucleus, 83 ~~exosome~~exosomes, 51 ~~cytosol~~cytosols, and 20 ~~ribosome~~ribosomes were obtained from 6,651 candidate lncRNAs. Among our candidate lncRNA-mRNA/TF regulatory pairs, 13 were found in the nucleus, 9 in the cytoplasm, and 1 in the exosome. (Table ~~S7~~. S11).

Formatado: Fonte: Não Itálico

Validation of tissue expression patterns of candidate lncRNAs in *S. miltiorrhiza*

By analyzing the expression patterns and genome location of lncRNAs during the accumulation of tanshinones, we screened out the observable lncRNAs: Smlnc0000154, Smlnc0008477, Smlnc0008662, Smlnc0012647, Smlnc0018769, Smlnc0019429, Smlnc0032870, Smlnc0042160, Smlnc0052170, Smlnc0063419, ~~Smlnc0070114 which may be involved in the biosynthesis of *S. miltiorrhiza* diterpenes. According to the above, lncRNAs showed obvious differential expression in tissues, which may be related to its regulatory function. Therefore, to further verify their expression in different tissues of *S. miltiorrhiza*, the tissue expression patterns were verified by qRT-PCR. Total RNA was extracted from fresh tissues of root, stem, leaf and flower from four parts of *S. miltiorrhiza*, and the expression level was detected by RT-PCR (Reverse Transcription-Polymerase Chain Reaction) after reverse transcription.~~and Smlnc0070114, which may be involved in the biosynthesis of *S. miltiorrhiza* diterpenes. The results showed that the candidate lncRNAs were differentially expressed in different tissues, among which Smlnc0012647, Smlnc0032870, Smlnc0042160, Smlnc0063419, and Slnc0070114 were highly expressed in root tissue. Smlnc0000154, Smlnc0008477, Smlnc0019429, and Smlnc0052170 were more stem-specific. Smlnc0018769 ~~were~~was expressed mainly in flowers and stems (~~Figure~~Fig. 3). According to the above, lncRNAs showed obvious differential expression in different tissues, which may be related to their regulatory function.

Time-series expression analysis of candidate lncRNA-mRNA/TF pairs involved in diterpenoid biosynthesis of *S. miltiorrhiza* induced by MeJA

~~In order to further explore whether candidate lncRNA-mRNA/TF pairs have response to diterpene pathway, using the qRT-PCR method to analyse expression patterns of 23 lncRNA-mRNA/TF pairs, materials from roots of *S. miltiorrhiza* treated with MeJA solution for 6, 12, 24, and 48 h. Plant materials treated with MeJA dissolving media were used as a control (0 h), and three biological replications were carried out. The observable~~The results obtained from the preliminary analysis of MeJA-induced gene time-series expression were presented in ~~Figure~~Fig. 4.

According to the general trend, MeJA treatment significantly changed the expression of ~~most~~19 genes in *S. miltiorrhiza* (Fig. S3). The graph ~~shows~~showed that at the 6 h time point, the expression of most genes ~~has~~showed a ~~downtrend at 6 h time point,~~downward trend, and only the Smlnc0018769-*Sm0037093* pair and Smlnc0008662-*Sm0063385* pair ~~have~~had an upward trend ~~at this time point.~~. Smlnc0008477-*Sm0056000* pair~~,~~ and Smlnc0012647-Smlnc0032870 pair had the same expression trend. Smlnc0042160-*Sm0028870* pair had the same expression trends at 6 h, 12~~,~~ h, and 24 h ~~point of~~ time points. Smlnc0018769-*Sm0037093*, Smlnc0008662-*Sm0063385*, and Smlnc0052170-*Sm0067296* pairs had opposite expression trends at 48 h time point ~~of time~~, Smlnc0019429-*Sm0026208* pair ~~at the point of time in both 24 and 48 h~~ had ~~the~~ opposite trends~~.~~ ~~Our CYP76AH1 Sm0063385 was consistent with the response trends during 0~24 h found in previous studies (Li et al., 2021), however, we reached a peak at 48~~ at 24 h and 48 h time point.

~~Through the plug-in cytoHubba, we have calculated the top 10 hub genes (*Table S8*, Figure 5D).~~ From the data in ~~Figure~~Fig. 5A, we could be found that Smlnc0063419, *Sm0026208*, and *Sm0067296* had similar expression ~~pattern~~patterns, which form the lncRNA-mRNA-TF module. *Sm0026208* and *Sm0067296* were annotated with TF WRKY2 and mRNA GA3ox2, respectively. ~~WRKY might respond to GA_3 stress by regulating the development of phloem fibers in jute (*Corchorus capsularis*) (Zhang et al., 2020).~~ Smlnc0012647, Smlnc0032870, and *Sm0009433* had similar expression ~~pattern~~patterns, as shown in ~~the Figure~~Fig. 5B, *Sm0009433* ~~were~~was annotated with TF MYB36. Smlnc0042160 and *Sm0037093* had similar expression ~~pattern~~patterns, as shown in ~~the Figure~~Fig. 5C, *Sm0037093* ~~were~~was annotated with mRNA KS2. Through the plug-in cytoHubba, we calculated the top 10 hub genes (Table S12, Fig. 5D).

Discussion

In the characteristic analysis of *S. miltiorrhiza* lncRNAs, our study found that the length of lncRNAs was shorter than protein-coding genes, which was also found in other species (Hao et al., 2015; Liu et al., 2018; Shen et al., 2018). The mean GC content was a little lower than mRNAs also found in *Populus tomentosa* (Zhou et al., 2017).

Our sequence conservation analysis showed that species of the same genus were more conserved compared with plants of different families. And a previous study also showed that the majority of lncRNAs had high sequence conservation at the intra-species and sub-species levels (Deng et al.,

Formatado: 正文1
Formatado: Fonte: Não Itálico
Formatado: Fonte: Arial
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico
Formatado: Fonte: Itálico

2018). The conservative analysis of the candidate lncRNAs revealed that the primary sequence of lncRNAs varies widely, ~~In order to better explore the role of lncRNAs in diterpenoid biosynthesis of *S. miltiorrhiza*. We obtained abundant lncRNAs from RNA-seq data, and found that they have many different characteristics from mRNAs. Part of the lncRNAs may act as the precursor of miRNA or act on the target gene. By analyzing the expression profiles of lncRNAs, it has found that lncRNAs act extensively on mRNA/TF, and the same mRNA was regulated by multiple lncRNAs. On the basis of genomic location and expression patterns, we screened out some candidate lncRNA-mRNA-TF pairs, and found that the expressions of lncRNAs in different tissues were different. In addition, some lncRNAs and mRNAs/TFs in the lncRNA-mRNA-TF pairs had particular response to MeJA stimulation, forming network modules of lncRNA-mRNA and/or TF.~~

~~Like mRNA, most of the lncRNAs in eukaryotes are transcribed by Pol II, and these are considered stable classical lncRNA, which is similar to mRNA, has a 5' cap structure, polyA tail, and promoter sequence of a part of lncRNAs. The splicing site is also similar to mRNA and has some conservation (Rinn & Chang, 2020).~~ Polymerase C-terminal domain modifications [e.g., threonine 4 phosphorylation (CTD-T4P)] are ~~enriched at lncRNA~~on the promoters~~, resulting in less~~ of lncRNAs, which leads to the decrease of the polymerase ~~pausing~~pause and ~~earlier~~the advance of the termination ~~throughout~~in the whole lncRNA ~~gene bodies. therefore the~~ genome. The transcription rate of ~~lncRNA~~lncRNAs is very fast, which means that ~~it~~they can quickly act on the regulatory target and respond to the signal~~, therefore~~. Therefore, transcription ~~precision is low~~accuracy and sequence conservation ~~is~~were low (Rinn & Chang, 2020). ~~In addition,~~lncRNA function was maintained across large evolutionary distances even when the lncRNA sequence substantially diverged (Ulitsky, 2016). RNA secondary structures are the units of lncRNA words, and disparate sequences form similar structure–function relationships to transmit symbolic language like hieroglyphics, thus forming the molecular grammar of lncRNAs (Rinn & Chang, 2020).

In genomic location analysis, we obtained 23 (accounting for 1.64%, 23 lncRNAs) co-located pairs observed in our co-expression ~~of lncRNA is more~~network. This indicates that most lncRNAs are not co-expressed with their nearby coding genes and are transcribed independently (Liao et al., 2011).

Although advances have been made in the miRNA and miRNA target prediction fields, the precision of miRNA target prediction needs to be improved (Akgül et al., 2022). To reduce false positives, we used three kinds of prediction software to predict miRNA targets. Although many miRNA databases and prediction software are published for plants, few of them are available (de Amorim, Pedro & Paschoal, 2022). This reduced our chance to find miRNAs associated with the lncRNA-mRNA/TF pairs in the diterpenoid biosynthetic pathway. Based on the relationship between miRNAs and lncRNAs, we predict that 14 lncRNAs were potential targets or target

Formatado: Fonte: Times, 12 pt

mimics of 66 miRNAs. However, all the results are predicted preliminarily based on bioinformatic analyses and need to be further validated.

Our study indicates that lncRNAs may have different subcellular localization and tissue-specific expression. Both of them are closely related to function. At present, there were many studies indicate that lncRNAs contain specific RNA motifs with nuclear localization (Zhang et al., 2014), many lncRNAs have significant tissue-specific expression, and their expression may be limited to specific cell types or induced by specific signals ~~specific~~ (Liu et al., 2012~~; Zhang et al., 2014)~~). And ~~and~~ it is more specific in different growth conditions (Yu et al., 2019) and developmental stages (Sanchita, Trivedi & Asif, 2020) ~~and growth conditions (Yu et al., 2019). These characteristics determine that it has a special "functional language" which is different from that encoding protein RNA and protein.~~

. These characteristics determine that it has a special "functional language" which is different from that encoding protein RNA and protein.

~~The current study indicated that characteristics of the lncRNA found that lncRNAs were generally shorter than protein-coding genes. These results match those observed in earlier studies (Deng et al., 2018). A conservative analysis of the candidate lncRNAs revealed that the primary sequence of lncRNA vary widely, despite all that, many lncRNAs exhibit syntenic conservation (Rinn & Chang, 2020), the transcription location of lncRNA are highly conserved among related species (Deng et al., 2018). In other study, only a few cucumber lincRNAs were conservative among in distantly related species (Hao et al., 2015). This also accords with our observations. Although the sequence is poorly conserved, RNA secondary structures are the units of lncRNA words, disparate sequences form similar structure–function relationships to transmit symbolic language like hieroglyphics, thus forming the molecular grammar of lncRNA (Rinn & Chang, 2020).~~

~~We obtained a large-scale regulatory network in co-expression network analysis, in which only two genes, Sm0012648 and Sm0037093 were highly expressed relative to their corresponding co-located genes. This shows that GRN studies often results in a static view of gene expression, which can make it difficult to disentangle the regulatory pathway structure response to a stimulus. Time-based analysis may uncover the temporal transcriptional logic for plant response system, and provide more accurate predictions for targeted breeding(Greenham & McClung, 2018). Through studying the time-series expression of some candidate regulatory pairs stimulated by MeJA, we found that some lncRNA was down-regulated in the early stage and the corresponding mRNA was up-regulated in the later stage between the regulatory pairs, plants response to signal had time delay. We speculated that lncRNA had a negative regulatory effect on it, but if we only observed the expression levels at two time points, the down-pattern of lncRNA-intermediate-expression was ignored, we will conclude that lncRNA has a positive~~

Formatado: Fonte: 12 pt, Cor da fonte: Automática, Não Sobrescrito/ Subscrito

Formatado: Fonte: Times, 12 pt

~~regulatory effect on the gene. All this is attributed to the complexity of biological signaling networks.~~

The results of our study showed that ~~the hub mRNAs~~just two pairs were co-expression and co-location in the high correlation co-expression network ~~above~~with |PCC| ≥ 0.8~~, its associated high correlation lncRNAs were not the lncRNAs in the co-expression and co-location pairs, and just two pairs were co-expression and co-location. However, the correlation between co-expression and co-location pairs was above 0.4, indicating~~. Indicating that there were many higher-level "acquaintances" between the regulatory pairs, who need to participate in the gene expression work more frequently. Therefore, the ~~similarity of gene expression~~ PCC value could only provide the threshold for defining whether the relationship pair was valid. ~~Rather than predicting the credibility of a particular function, we~~We need ~~to combine~~the combined interpretation of co-expression analysis with other genomic data to enhance the generation of biologically relevant information~~.~~ (Rao & Dixon, 2019). Another possible explanation ~~for this~~ is that the strength of required gene ~~coexpression~~co-expression may depend on the stability or toxicity of the metabolites, strong coexpression should only be required for unstable monomers (Obayashi & Kinoshita, 2009).

Building gene regulatory networks from transcriptomic studies often result in a static view of gene expression, which can make it difficult to disentangle the regulatory pathway structure response to a stimulus, and time-based analysis may uncover the temporal transcriptional logic for plant response systems, and provide more accurate predictions for targeted breeding (Greenham & McClung, 2018). Through studying the time-series expression of some candidate regulatory pairs stimulated by MeJA, we observed a more detailed landscape different from the gene expression at two-time points in RNA-seq data. And we found that some lncRNA expression was down-regulated in the early stage and the expression of the corresponding mRNAs was up-regulated in the later stage between the regulatory pairs, plants' response to the signal had a time delay. For example, in response to vernalization, *COOLAIR* is transiently induced by prolonged cold, reaching a maximum expression level after 2 weeks (Swiezewski et al., 2009). Meanwhile, we have also found that some genes show different expression patterns within a period, both up-regulated and down-regulated (e.g. *Sm0056000*, *Sm0063385*, Smlnc0008477, Smlnc0008662, Smlnc0012647, Smlnc0032870), which may be related to gene regulatory networks are inherently complex, with multiple feedback and feedforward loops (Wils & Kaufmann, 2017).

Based on the relationship in expression and location of lncRNA~~-~~, mRNA/TF genes, we discussed the gene expression of these candidate lncRNA-mRNA/TF pairs under MeJA ~~treatment~~induction. The increase ~~of~~in the content of bioactive compounds after ~~the treatment of~~ MeJA induction was consistent with the up-regulated expression of key genes (e.g. CPS and KS) under ~~the~~ MeJA induction, ~~The~~Our results ~~of our research suggest~~showed that the expression of

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Times, 12 pt

Sm0056000 (CPS ~~(Sm0056000~~) and ~~KS (~~*Sm0037093* (KS) were ~~also with the~~ up-regulated ~~expression~~under MeJA induction, which was also observed in another study (Luo et al., 2014)~~, however, they~~. However, these two genes started to raise the trend at 12 h and 24 h in our study, respectively. The response of TFs to MeJA was also observed in other ~~studies~~study (Luo et al., 2014), it was consistent with the WRKY TF we studied. And the response trends during 0~24 h of *Sm0063385* (CYP76AH1) were found in previous studies (Li et al., 2021), however, we reached a peak at 48 h. In Smlnc0063419-*Sm0026208*-*Sm0067296* module (*Sm0026208*: TF WRKY2, *Sm0067296*: GA3ox2), WRKY might respond to GA3 stress by regulating the development of phloem fibers in jute (*Corchorus capsularis*) in other study (Zhang et al., 2020), there may be a similar phenomenon in *S. miltiorrhiza*. Combined with the expression data of time~~ series, we can observe the gene expression in detail~~ series, which provides an exploratory method for the role of lncRNA and establishing the relationship between lncRNA and mRNA/TF at the gene expression level.

RNA regulatory sequences developed an RNA regulatory system ~~which~~that bypassed the complexity limits of regulatory networks operating with protein alone (Mattick, 2004). The true basis of the evolution and development programming of complex organisms needed to be re-examined for the RNA regulatory system to set up a new regulatory framework (Mattick, 2004).

Conclusions

To sum up, we identified 6,651 candidate lncRNAs from RNA-seq in *S. miltiorrhiza* tanshinones accumulation period. ~~In total, 18 lncRNAs were differentially expressed.~~We all obtained 23 co-located and co-expressed lncRNA-mRNA/TF pairs related to diterpenoid biosynthesis. In the co-expression network, two hub genes *Sm0012648*~~,~~ and *Sm0037093* were among the co-expressed and co-localized~~. Among the 23 co-located and co-expressed lncRNA-mRNA/TF pairs,~~ pairs. And 13 genes that showed a significant response to MeJA stimulation, ~~they~~ showed ~~the same~~similar or opposite expression ~~trend~~trends of the corresponding ~~mRNA/TF~~pair in different ~~time~~ periods. This study set out to investigate the lncRNAs that may be involved in the diterpenoid biosynthesis pathway of *S. miltiorrhiza* and their ~~regulatory~~related mRNA/~~TFs~~TF. As well as three lncRNA-mRNA and/or TF network modules were constructed. These findings indicate that lncRNA is complex in regulating mRNA/TF in general. One limitation of this study is that we have not studied the specific mechanisms of lncRNA that regulates mRNA or TF. Further research is required to establish the therapeutic efficiency of lncRNA regulatory mechanisms.

~~Acknowledgements~~

~~Lin Wang and Xin Chen conceived the ideas. Lin Wang designed the study. Lin Wang, Fang Liu, Peijin Zou, Rui Liu analyzed the data. Lin Wang interpreted the data and wrote the main manuscript text.~~

Formatado: Fonte: Itálico

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Times, 12 pt

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Arial, Negrito

References

Akgül B, Stadler PF, Hawkins LJ, Hadj-Moussa H, Storey KB, Ergin K, Çetinkaya R, Paschoal AR, Nachtigall PG, Tutar Y, Yousef M, Allmer J. 2022. 44 Current Challenges in miRNomics. In: 423–438. DOI: 10.1007/978-1-0716-1170-8_19.

de Amorim TS, Pedro DLF, Paschoal AR. 2022. MicroRNA Databases and Tools. In: 131–166. DOI: 10.1007/978-1-0716-1170-8_7.

Aoki K, Ogata Y, Shibata D. 2007. Approaches for Extracting Practical Information from Gene Co-expression Networks in Plant Biology. *Plant and Cell Physiology* 48:381–390. DOI: 10.1093/pcp/pcm013.

Bonnet E, He Y, Billiau K, van de Peer Y. 2010. TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics* 26:1566–1568. DOI: 10.1093/bioinformatics/btq233.

Brockdorff N. 1998. The role of Xist in X-inactivation. *Current Opinion in Genetics & Development* 8:328–333. DOI: 10.1016/S0959-437X(98)80090-7.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. DOI: 10.1186/1471-2105-10-421.

Cao Z, Pan X, Yang Y, Huang Y, Shen H-B. 2018. The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics* 34:2185–2194. DOI: 10.1093/bioinformatics/bty085.

Carlevaro-Fita J, Johnson R. 2019. Global Positioning System: Understanding Long Noncoding RNAs through Subcellular Localization. *Molecular Cell* 73:869–883. DOI: 10.1016/j.molcel.2019.02.008.

Chang Y, Wang M, Li J, Lu S. 2019. Transcriptomic analysis reveals potential genes involved in tanshinone biosynthesis in *Salvia miltiorrhiza*. *Scientific Reports* 9:14929. DOI: 10.1038/s41598-019-51535-9.

Chen L-L. 2016. Linking Long Noncoding RNA Localization and Function. *Trends in Biochemical Sciences* 41:761–772. DOI: 10.1016/j.tibs.2016.07.003.

Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* 8:S11. DOI: 10.1186/1752-0509-8-S4-S11.

Dai X, Zhuang Z, Zhao PX. 2018. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Research* 46:W49–W54. DOI: 10.1093/nar/gky316.

Deng P, Liu S, Nie X, Weining S, Wu L. 2018. Conservation analysis of long non-coding RNAs in plants. *Science China Life Sciences* 61:190–198. DOI: 10.1007/s11427-017-9174-9.

Du Q, Li C, Li D, Lu S. 2015. Genome-wide analysis, molecular cloning and expression profiling reveal tissue-specifically expressed, feedback-regulated, stress-responsive and alternatively spliced novel genes involved in gibberellin metabolism in *Salvia miltiorrhiza*. *BMC Genomics* 16:1087. DOI: 10.1186/s12864-015-2315-5.

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES. 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539:452–455. DOI: 10.1038/nature20149.

Formatado: Fonte: 11 pt

Formatado: Fonte: Itálico

Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA, Paz-Ares J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genetics* 39:1033–1037. DOI: 10.1038/ng2079.

Gao W, Hillwig ML, Huang L, Cui G, Wang X, Kong J, Yang B, Peters RJ. 2009. A Functional Genomics Approach to Tanshinone Biosynthesis Provides Stereochemical Insights. *Organic Letters* 11:5170–5173. DOI: 10.1021/ol902051v.

Gong H-Y, Zeng Y, Chen X-Y. 2014. Diterpene Synthases and Their Responsible Cyclic Natural Products. *Natural Products and Bioprospecting* 4:59–72. DOI: 10.1007/s13659-014-0012-8.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29:644–652. DOI: 10.1038/nbt.1883.

Greenham K, McClung CR. 2018. Time to build on good design: Resolving the temporal dynamics of gene regulatory networks. *Proceedings of the National Academy of Sciences* 115:6325–6327. DOI: 10.1073/pnas.1807707115.

Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. 2008. The Vienna RNA Websuite. *Nucleic Acids Research* 36:W70–W74. DOI: 10.1093/nar/gkn188.

Guil S, Esteller M. 2012. Cis-acting noncoding RNAs: friends and foes. *Nature Structural & Molecular Biology* 19:1068–1075. DOI: 10.1038/nsmb.2428.

Gundlach H, Müller MJ, Kutchan TM, Zenk MH. 1992. Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures. *Proceedings of the National Academy of Sciences* 89:2389–2393. DOI: 10.1073/pnas.89.6.2389.

Guo Z, Kuang Z, Zhao Y, Deng Y, He H, Wan M, Tao Y, Wang D, Wei J, Li L, Yang X. 2022. PmiREN2.0: from data annotation to functional exploration of plant microRNAs. *Nucleic Acids Research* 50:D1475–D1482. DOI: 10.1093/nar/gkab811.

Guo J, Zhou YJ, Hillwig ML, Shen Y, Yang L, Wang Y, Zhang X, Liu W, Peters RJ, Chen X, Zhao ZK, Huang L. 2013. CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proceedings of the National Academy of Sciences* 110:12108–12113. DOI: 10.1073/pnas.1218061110.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458:223–227. DOI: 10.1038/nature07672.

Hao Z, Fan C, Cheng T, Su Y, Wei Q, Li G. 2015. Genome-Wide Identification, Characterization and Evolutionary Analysis of Long Intergenic Noncoding RNAs in *Cucumber*. *PLOS ONE* 10:e0121800. DOI: 10.1371/journal.pone.0121800.

Hedden P, Thomas SG. 2012. Gibberellin biosynthesis and its regulation. *Biochemical Journal* 444:11–25. DOI: 10.1042/BJ20120245.

Helliwell CA, Chandler PM, Poole A, Dennis ES, Peacock WJ. 2001. The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis

Formatado: Fonte: Itálico

pathway. *Proceedings of the National Academy of Sciences* 98:2065–2070. DOI: 10.1073/pnas.98.4.2065.

Huang L, Dong H, Zhou D, Li M, Liu Y, Zhang F, Feng Y, Yu D, Lin S, Cao J. 2018. Systematic identification of long non-coding RNAs during pollen development and fertilization in *Brassica rapa*. *The Plant Journal* 96:203–222. DOI: 10.1111/tpj.14016.

Ji AJ, Luo HM, Xu ZC, Zhang X, Zhu YJ, Liao BS, Yao H, Song JY, Chen SL. 2016. Genome‐Wide Identification of the AP2/ERF Gene Family Involved in Active Constituent Biosynthesis in *Salvia miltiorrhiza*. *The Plant Genome* 9. DOI: 10.3835/plantgenome2015.08.0077.

Jiang M, Chen H, Liu J, Du Q, Lu S, Liu C. 2021. Genome-wide identification and functional characterization of natural antisense transcripts in *Salvia miltiorrhiza*. *Scientific Reports* 11:4769. DOI: 10.1038/s41598-021-83520-6.

Kai G, Liao P, Zhang T, Zhou W, Wang J, Xu H, Liu Y, Zhang L. 2010. Characterization, expression profiling, and functional identification of a gene encoding geranylgeranyl diphosphate synthase from *Salvia miltiorrhiza*. *Biotechnology and Bioprocess Engineering* 15:236–245. DOI: 10.1007/s12257-009-0123-y.

Kai G, Xu H, Zhou C, Liao P, Xiao J, Luo X, You L, Zhang L. 2011. Metabolic engineering tanshinone biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *Metabolic Engineering* 13:319–327. DOI: 10.1016/j.ymben.2011.02.003.

Kang Y-J, Yang D-C, Kong L, Hou M, Meng Y-Q, Wei L, Gao G. 2017. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research* 45:W12–W16. DOI: 10.1093/nar/gkx428.

Kopp F, Mendell JT. 2018. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 172:393–407. DOI: 10.1016/j.cell.2018.01.011.

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Research* 47:D155–D162. DOI: 10.1093/nar/gky1141.

Lécuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, Hughes TR, Tomancak P, Krause HM. 2007. Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function. *Cell* 131:174–187. DOI: 10.1016/j.cell.2007.08.003.

Li W, Bai Z, Pei T, Yang D, Mao R, Zhang B, Liu C, Liang Z. 2019. SmGRAS1 and SmGRAS2 Regulate the Biosynthesis of Tanshinones and Phenolic Acids in *Salvia miltiorrhiza*. *Frontiers in Plant Science* 10. DOI: 10.3389/fpls.2019.01367.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. DOI: 10.1186/1471-2105-12-323.

Li Y, Kong D, Fu Y, Sussman MR, Wu H. 2020. The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant Physiology and Biochemistry* 148:80–89. DOI: 10.1016/j.plaphy.2020.01.006.

Li B, Li J, Chai Y, Huang Y, Li L, Wang D, Wang Z. 2021. Targeted mutagenesis of CYP76AK2 and CYP76AK3 in *Salvia miltiorrhiza* reveals their roles in tanshinones biosynthetic pathway. *International Journal of Biological Macromolecules* 189:455–463. DOI: 10.1016/j.ijbiomac.2021.08.112.

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

~~Li C, Li D, Li J, Shao F, Lu S. 2017. Characterization of the polyphenol oxidase gene family reveals a novel microRNA involved in posttranscriptional regulation of PPOs in Salvia miltiorrhiza. *Scientific Reports* 7:44622. DOI: 10.1038/srep44622.~~

Li C, Li D, Shao F, Lu S. 2015. Molecular cloning and expression analysis of *WRKY* transcription factor genes in *Salvia miltiorrhiza*. *BMC Genomics* 16:200. DOI: 10.1186/s12864-015-1411-x.

Li C, Lu S. 2014. Genome-wide characterization and comparative analysis of R2R3-MYB transcription factors shows the complexity of MYB-associated regulatory networks in *Salvia miltiorrhiza*. *BMC Genomics* 15:277. DOI: 10.1186/1471-2164-15-277.

Li D, Shao F, 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.

Li S, Yamada M, Han X, Ohler U, Benfey PN. 2016. High-Resolution Expression Map of the *Arabidopsis* Root Reveals Alternative Splicing and lincRNA Regulation. *Developmental Cell* 39:508–522. DOI: 10.1016/j.devcel.2016.10.012.

Li A, Zhang J, Zhou Z. 2014. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved *k*-mer scheme. *BMC Bioinformatics* 15:311. DOI: 10.1186/1471-2105-15-311.

Liang Z-S, Yang D-F, Liang X, Zhang Y-J, Liu Y, Liu F-H. 2012. Roles of reactive oxygen species in methyl jasmonate and nitric oxide-induced tanshinone production in *Salvia miltiorrhiza* hairy roots. *Plant Cell Reports* 31:873–883. DOI: 10.1007/s00299-011-1208-6.

Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, Zhao G, Luo H, Bu D, Zhao H, Skogerbø G, Wu Z, Zhao Y. 2011. Large-scale prediction of long non-coding RNA functions in a coding–non-coding gene co-expression network. *Nucleic Acids Research* 39:3864–3878. DOI: 10.1093/nar/gkq1348.

Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua N-H. 2012. Genome-Wide Analysis Uncovers Regulation of Long Intergenic Noncoding RNAs in *Arabidopsis*. *The Plant Cell* 24:4333–4345. DOI: 10.1105/tpc.112.102855.

Liu H, Lu Y, Wang J, Hu J, Wuyun T. 2018. Genome-wide screening of long non-coding RNAs involved in rubber biosynthesis in *Eucommia ulmoides*. *Journal of Integrative Plant Biology* 60:1070–1082. DOI: 10.1111/jipb.12693.

~~Love MI, Huber W, Anders~~Lu S. ~~2014. Moderated estimation of fold change~~2021. Biosynthesis and ~~dispersion~~Regulatory Mechanisms of Bioactive Compounds in *Salvia miltiorrhiza* , a Model System for ~~RNA-seq data with DESeq2.~~ ~~*Genome*~~Medicinal Plant Biology ~~15:550.~~. *Critical Reviews in Plant Sciences* 40:243–283. DOI: 10.~~1186/s13059-014-0550-8~~1080/07352689.2021.1935719.

Luo H, Zhu Y, Song J, Xu L, Sun C, Zhang X, Xu Y, He L, Sun W, Xu H, Wang B, Li X, Li C, Liu J, Chen S. 2014. Transcriptional data mining of *Salvia miltiorrhiza* in response to methyl jasmonate to examine the mechanism of bioactive compound biosynthesis and regulation. *Physiologia Plantarum* 152:241–255. DOI: 10.1111/ppl.12193.

Ma Y, Cui G, Chen T, Ma X, Wang R, Jin B, Yang J, Kang L, Tang J, Lai C, Wang Y, Zhao Y, Shen Y, Zeng W, Peters RJ, Qi X, Guo J, Huang L. 2021. Expansion within the CYP71D subfamily drives the heterocyclization of tanshinones synthesis in *Salvia miltiorrhiza*. *Nature Communications* 12:685. DOI: 10.1038/s41467-021-20959-1.

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Não Itálico

Formatado: Fonte: Itálico

Ma Y, Ma X-H, Meng F-Y, Zhan Z-L, Guo J, Huang L-Q. 2016. RNA interference targeting CYP76AH1 in hairy roots of *Salvia miltiorrhiza* reveals its key role in the biosynthetic pathway of tanshinones. *Biochemical and Biophysical Research Communications* 477:155–160. DOI: 10.1016/j.bbrc.2016.06.036.

Ma X-H, Ma Y, Tang J-F, He Y-L, Liu Y-C, Ma X-J, Shen Y, Cui G-H, Lin H-X, Rong Q-X, Guo J, Huang L-Q. 2015. The Biosynthetic Pathways of Tanshinones and Phenolic Acids in *Salvia miltiorrhiza*. *Molecules* 20:16235–16254. DOI: 10.3390/molecules200916235.

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*. *Journal of Experimental Botany* 63:2809–2823. DOI: 10.1093/jxb/err466.

Mattick JS. 2004. RNA regulation: a new genetics? *Nature Reviews Genetics* 5:316–323. DOI: 10.1038/nrg1321.

Mercer TR, Dinger ME, Mattick JS. 2009. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics* 10:155–159. DOI: 10.1038/nrg2521.

~~Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Research* 49:D412–D419. DOI: 10.1093/nar/gkaa913.~~

Nagano T, Fraser P. 2011. No-Nonsense Functions for Long Noncoding RNAs. *Cell* 145:178–181. DOI: 10.1016/j.cell.2011.03.014.

Obayashi T, Kinoshita K. 2009. Rank of Correlation Coefficient as a Comparable Measure for Biological Significance of Gene Coexpression. *DNA Research* 16:249–260. DOI: 10.1093/dnares/dsp016.

Palazzo AF, Koonin E v. 2020. Functional Long Non-coding RNAs Evolve from Junk Transcripts. *Cell* 183:1151–1161. DOI: 10.1016/j.cell.2020.09.047.

Paschoal AR, Lozada-Chávez I, Domingues DS, Stadler PF. 2017. ceRNAs in plants: computational approaches and associated challenges for target mimic research. *Briefings in Bioinformatics*. DOI: 10.1093/bib/bbx058.

Ponting CP, Oliver PL, Reik W. 2009. Evolution and Functions of Long Noncoding RNAs. *Cell* 136:629–641. DOI: 10.1016/j.cell.2009.02.006.

PU Xiangdong, WANG Lizhi. 2017. Cloning and Bioinformatic Analysisof CYP76AK5 Related to Tanshinone Biosynt hesis in *Salvia miltiorrhiza*. *Modern Chinese Medicine* 19:1106–1110. DOI: 10.13313/j.issn.1673-4890.2017.8.011.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Research* 40:D290–D301. DOI: 10.1093/nar/gkr1065.

Quan M, Du Q, Xiao L, Lu W, Wang L, Xie J, Song Y, Xu B, Zhang D. 2019. Genetic architecture underlying the lignin biosynthesis pathway involves noncoding RNAs and transcription factors for growth and wood properties in *Populus*. *Plant Biotechnology Journal* 17:302–315. DOI: 10.1111/pbi.12978.

Rao X, Dixon RA. 2019. Co-expression networks for plant biology: why and how. *Acta Biochimica et Biophysica Sinica* 51:981–988. DOI: 10.1093/abbs/gmz080.

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Revelle W. 2022. psych: Procedures for Psychological, Psychometric, and Personality Research.

Rinn JL, Chang HY. 2020. Long Noncoding RNAs: Molecular Modalities to Organismal Functions. *Annual Review of Biochemistry* 89:283–308. DOI: 10.1146/annurev-biochem-062917-012708.

Sanchita, Trivedi PK, Asif MH. 2020. Updates on plant long non-coding RNAs (lncRNAs): the regulatory components. *Plant Cell, Tissue and Organ Culture (PCTOC)* 140:259–269. DOI: 10.1007/s11240-019-01726-z.

van Schie CCN, Ament K, Schmidt A, Lange T, Haring MA, Schuurink RC. 2007. Geranyl diphosphate synthase is required for biosynthesis of gibberellins. *The Plant Journal* 52:752–762. DOI: 10.1111/j.1365-313X.2007.03273.x.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* 13:2498–2504. DOI: 10.1101/gr.1239303.

Shen E, Zhu X, Hua S, Chen H, Ye C, Zhou L, Liu Q, Zhu Q-H, Fan L, Chen X. 2018. Genome-wide identification of oil biosynthesis-related long non-coding RNAs in allopolyploid *Brassica napus*. *BMC Genomics* 19:745. DOI: 10.1186/s12864-018-5117-8.

Shimane M, Ueno Y, Morisaki K, Oogami S, Natsume M, Hayashi K, Nozaki H, Kawaide H. 2014. Molecular evolution of the substrate specificity of *ent*-kaurene synthases to adapt to gibberellin biosynthesis in land plants. *Biochemical Journal* 462:539–546. DOI: 10.1042/BJ20140134.

Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA, Young RA. 2015. Transcription factor trapping by RNA in gene regulatory elements. *Science* 350:978–981. DOI: 10.1126/science.aad3346.

Swiezewski S, Liu F, Magusin A, Dean C. 2009. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* 462:799–802. DOI: 10.1038/nature08618.

Team R. 2010. R : a language and environment for statistical computing. R Foundation for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. *Computing* 14:12–21.

Ulitsky I. 2016. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nature Reviews Genetics* 17:601–614. DOI: 10.1038/nrg.2016.85.

Wang H-L v., Chekanova JA. 2017. Long Noncoding RNAs in Plants. In: 133–154. DOI: 10.1007/978-981-10-5203-3_5.

Wang Y, Zhang H, Li Q, Jin J, Chen H, Zou Y, Huang X, Ding Y. 2021. Genome-Wide Identification of lncRNAs Involved in Fertility Transition in the Photo-Thermosensitive Genic Male Sterile Rice Line Wuxiang S. *Frontiers in Plant Science* 11. DOI: 10.3389/fpls.2020.580050.

Wasternack C. 2007. Jasmonates: An Update on Biosynthesis, Signal Transduction and Action in Plant Stress Response, Growth and Development. *Annals of Botany* 100:681–697. DOI: 10.1093/aob/mcm079.

Formatado: Fonte: Itálico

Wils CR, Kaufmann K. 2017. Gene-regulatory networks controlling inflorescence and flower development in Arabidopsis thaliana. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1860:95–105. DOI: 10.1016/j.bbagrm.2016.07.014.

Wu B, Li Y, Yan H, Ma Y, Luo H, Yuan L, Chen S, Lu S. ~~2012~~2012a. Comprehensive transcriptome analysis reveals novel genes involved in cardiac glycoside biosynthesis and mlncRNAs associated with secondary metabolism and stress response in *Digitalis purpurea*. *BMC Genomics* 13:15. DOI: 10.1186/1471-2164-13-15.

Wu H-J, ~~Wang Z-M~~Ma Y-K, Chen T, Wang M, Wang X-J. ~~2013. Widespread Long Noncoding RNAs as Endogenous Target Mimics for MicroRNAs in Plants . *Plant Physiology* 161:1875–1884.~~2012b. PsRobot: a web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Research* 40:W22–W28. DOI: 10.~~1104/pp.113.215962~~1093/nar/gks554.

Xiao Y, Zhang L, Gao S, Saechao S, Di P, Chen J, Chen W. 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 RJ, Weirather J, Luo H, Liao B, Zhang X, Zhu Y, Ji A, Zhang B, Hu S, Au KF, Song J, Chen S. 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. *The Plant Journal* 82:951–961. DOI: 10.1111/tpj.12865.

Xu H, Song J, Luo H, Zhang Y, Li Q, Zhu Y, Xu J, Li Y, Song C, Wang B, Sun W, Shen G, Zhang X, Qian J, Ji A, Xu Z, Luo X, He L, Li C, Sun C, Yan H, Cui G, Li X, Li X, Wei J, Liu J, Wang Y, Hayward A, Nelson D, Ning Z, Peters RJ, Qi X, Chen S. 2016. Analysis of the Genome Sequence of the Medicinal Plant Salvia miltiorrhiza. *Molecular Plant* 9:949–952. DOI: 10.1016/j.molp.2016.03.010.

Yamaguchi S. 2008. Gibberellin Metabolism and its Regulation. *Annual Review of Plant Biology* 59:225–251. DOI: 10.1146/annurev.arplant.59.032607.092804.

Yang Y, Hou S, Cui G, Chen S, Wei J, Huang L. 2010. Characterization of reference genes for quantitative real-time PCR analysis in various tissues of *Salvia miltiorrhiza*. *Molecular Biology Reports* 37:507–513. DOI: 10.1007/s11033-009-9703-3.

~~Yang T, Ma H, Zhang J, Wu T, Song T, Tian J, Yao Y. 2019. Systematic identification of long noncoding RNAs expressed during light - induced anthocyanin accumulation in apple fruit. *The Plant Journal* 100:572–590. DOI: 10.1111/tpj.14470.~~

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13:134. DOI: 10.1186/1471-2105-13-134.

Yin D-D, Li S-S, Shu Q-Y, Gu Z-Y, Wu Q, Feng C-Y, Xu W-Z, Wang L-S. 2018. Identification of microRNAs and long non-coding RNAs involved in fatty acid biosynthesis in tree peony seeds. *Gene* 666:72–82. DOI: 10.1016/j.gene.2018.05.011.

Yu T, Tzeng DTW, Li R, Chen J, Zhong S, Fu D, Zhu B, Luo Y, Zhu H. 2019. Genome-wide identification of long non-coding RNA targets of the tomato MADS box transcription factor RIN and function analysis. *Annals of Botany* 123:469–482. DOI: 10.1093/aob/mcy178.

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Zhang B, Gunawardane L, Niazi F, Jahanbani F, Chen X, Valadkhan S. 2014. A Novel RNA Motif Mediates the Strict Nuclear Localization of a Long Noncoding RNA. *Molecular and Cellular Biology* 34:2318–2329. DOI: 10.1128/MCB.01673-13.

Zhang X, Luo H, Xu Z, Zhu Y, Ji A, Song J, Chen S. 2015. Genome-wide characterisation and analysis of bHLH transcription factors related to tanshinone biosynthesis in *Salvia miltiorrhiza*. *Scientific Reports* 5:11244. DOI: 10.1038/srep11244.

Zhang L, Wan X, Xu Y, Niyitanga S, Qi J, Zhang L. 2020. De novo assembly of transcriptome and genome-wide identification reveal GA3 stress-responsive WRKY transcription factors involved in fiber formation in jute (*Corchorus capsularis*). *BMC Plant Biology* 20:403. DOI: 10.1186/s12870-020-02617-8.

Zhou D, Du Q, Chen J, Wang Q, Zhang D. 2017. Identification and allelic dissection uncover roles of lncRNAs in secondary growth of Populus tomentosa. *DNA Research* 24:473–486. DOI: 10.1093/dnares/dsx018.

Zhou J, Liu R, Shuai M, Yan Z-Y, Chen X. 2021. Comparative transcriptome analyses of different *Salvia miltiorrhiza* varieties during the accumulation of tanshinones. *PeerJ* 9:e12300. DOI: 10.7717/peerj.12300.

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico

Formatado: Fonte: Itálico