Correlations between SmCPS1 Promoter Polymorphism and Tanshinone Contents in Salvia miltiorrhiza

Ying SUN, Xin CHEN*, Xiao-Yan GAN, Zhu-Yun YAN, Di-Xiu MU and Qing-Rong WANG

Chengdu University of Traditional Chinese Medicine, Chengdu, P.R. China

*Corresponding author: chenxin64@hotmail.com

Abstract


Copalyldiphosphate synthase 1 (SmCPS1) is the first committed enzyme in tanshinone biosynthesis. The promoter region plays an important role in the transcriptional regulation of genes. Mutations in the promoter region may affect gene expression, resulting in changes in the amount of metabolites. In this study, we investigated the SmCPS1 gene promoter region together with the 388 bp downstream from the translation start site and the content of tanshinones of 12 different genotypes of Salvia miltiorrhiza. The cis-elements of SmCPS1 promoter were predicted and analysed by the Plant Transcriptional Regulatory Map database. We found (1) a different correlation between the polymorphism in the promoter region and the contents of tanshinones; (2) functional polymorphic loci – four tandem repeat variations, three indels and five single nucleotide polymorphisms (SNPs) in five cis-elements, three SNPs in exons and two SNPs in introns; (3) the correlation coefficient was higher when only functional (informative) polymorphic loci were considered. These findings have laid the foundation for further exploring the interspecific variation of S. miltiorrhiza and its relationship with the contents of tanshinones.

Keywords: cis-elements; copalyldiphosphate synthase 1; danshen; polymorphism

Salvia miltiorrhiza Bunge (danshen) is a medicinal model plant widely distributed throughout China (Song et al. 2010). Its root is used for the treatment of cardiovascular and cerebrovascular diseases (Yan 2015). The main effective components of the root of S. miltiorrhiza are tanshinones, including dihydrotanshinone I, cryptotanshinone, 1,2-dihydrotanshinone, tanshinone IIA, tanshinone I and so on (Zeng et al. 2017).

Both sexual and vegetative propagation is involved in the cultivation of S. miltiorrhiza. But the crude field management of some origins of S. miltiorrhiza resulted in quality degradation and impaired medicinal value of the herb. Modern technologies, such as DNA markers based on Amplified Fragment Length Polymorphism (AFLP), Random Amplified Polymorphic DNA (RAPD), ISSR (Inter Simple Sequence Repeat) (Song et al. 2010; Zhang et al. 2013) and Sequence Related Amplified Polymorphism (SRAP) (Li et al. 2008; Wang et al. 2010) were developed to identify different genotypes of S. miltiorrhiza. However, none of them is associated with important pharmacological ingredients. A more reliable trait for the identification of good quality danshen is required.

Former studies revealed the metabolic pathway of tanshinones (Ma et al. 2012; Yang et al. 2013). The product of the 2-C-methyl-d-erythritol 4-phosphate (MEP) pathway, geranylgeranyl diphosphate (GGPP), is catalyzed by Copalyl diphosphate synthase 1 (SmCPS1) and Kaurene synthase like (SmKSL) together into miltiradiene (Cui et al. 2015), and then processed by CYP76AH1 to produce various tanshinone precursors (Guo et al. 2013). SmCPS1 and SmCPS2 have been confirmed to be involved in the tanshinone biosynthesis. But SmCPS1 rather than SmCPS2 plays the dominant role in tanshinone biosynthesis (Cui et al. 2015; Li et al. 2015; Liu et al. 2016).
Therefore, the regulation of SmCPS1 will directly impact the downstream biosynthesis of tanshinones. The promoter carries the basic information of the spatio-temporal expression pattern of the gene, which is the main regulatory region of the gene. The promoter is a region located upstream of the transcription start site and contains short sequences which are named cis-elements. Ranging from 5 to 30 bp in length (Stewart et al. 2012) and variable both in sequence and in location, cis-elements are the binding sites for corresponding transcriptional factors (TFs) and can be species-specific (Wittkopp & Kalay 2011). The combination of the cis-elements determines the expression pattern of a gene.

By changing the binding capacity between the promoter elements and corresponding trans-factors, or the distance between co-functional elements, mutations from either substitution or insertion/deletion (indel) in the promoter can alter the expression pattern of a gene and subsequently lead to a phenotypic change (Pham et al. 2017). Therefore, it is tempting to speculate that the variation of tanshinone contents of different S. miltiorrhiza genotypes is related to the promoter variation. In this study, we investigated the SmCPS1 gene promoter region together with the 388bp downstream from the translation start site and the content of tanshinones of 12 different genotypes of S. miltiorrhiza to find the correlation between promoter variations and the content of tanshinones.

**MATERIAL AND METHODS**

**Material.** Samples were collected from the major origins of S. miltiorrhiza in China, these areas spanned the geographical range and have different ecological environments (Figure 1, Table 1). In order to eliminate the effect of the environment on the content of tanshinones, all samples were cultivated for 18 months in the Medicinal Botanical Garden of Chengdu University of Traditional Chinese Medicine before experiment.

**Detection of tanshinone content.** High performance liquid chromatography (HPLC) analysis was carried out on an Agilent 1200 Series HPLC instrument with Agilent ZORBAX Eclipse Plus C18 column (Analytical 4.6 × 250 mm, 5 µm; Agilent Technologies, USA), and Diode Array Detector (G315C 1260; Agilent Technologies). A gram of powder made from the fresh root of S. miltiorrhiza was used to extract the tanshinones with 10 ml methanol under ultrasonic conditions for 1 h. After cooling to room
temperature, methanol was added to replenish the lost weight. Then the mixture was filtered through 0.45 µl microporous membrane. A volume of 20 µl of the liquid was used for detection. All peaks were detected at the wavelength of 270 nm. The column temperature was maintained at room temperature. The flow rate was 1.0 ml/min. Acetonitrile-water (V/V) was used as the mobile phase. The program was set with a linear gradient T from 0 to 30 min: T = 0 min, acetonitrile/water (55:45, V/V); T = 30 min, acetonitrile/water (0:100, V/V).

**SmCPS1 promoter amplification.** DNA was extracted using the Plant Genomic DNA Kit (DP305, Tiangen Biotech Co., Beijing). PCR was adopted to amplify the promoter region together with the downstream coding areas of SmCPS1. The detailed information about the primer pairs and the PCR reactions is in File S1 in Electronic Supplementary Material (ESM). The sequences were aligned and truncated to the same ends for further analysis. Their GenBank accessions are in Table 1.

**Promoter analysis.** The Binding Site Prediction Tool of Plant Transcriptional Regulatory Map database (http://planttfdb.cbi.pku.edu.cn/prediction.php) (Jin et al. 2017) was used to identify the cis-elements in the species-specific datasets that it provides. For overlapped predictions, a conservative point of view that different TFs can compete for the same binding site was taken; so different TFs were retained if they shared a common binding locus. But for the same TF, if it had overlapped binding sites, only the result with minimum P-value was retained. The aligned consensus sequence was used for prediction. Then the local sequences that excluded gaps were used for the second prediction to eliminate the prediction bias resulting from gaps.

The cis-element locus on which no alternative bases existed according to the position weight matrix of a cis-element was defined as a key binding site. Polymorphic loci in the key binding sites, exons and introns such as tandem repeat variations (TRV), single nucleotide polymorphisms (SNP) and indels were defined as functional polymorphic loci.

**Statistical analysis.** Sequence alignment was conducted using clustw by MEGA 7 (Kumar et al. 2016) with default settings. The genetic distance matrix was generated by GenAlex (Peakall & Smouse 2012). Chemical component content distance matrices were generated using Euclidean distance. The Mantel test was conducted by GenAlex with a setting of 999 permutations.

**RESULTS**

**Detection of tanshinone content.** The contents of five tanshinones are shown in Figure 2 and Table S1 in ESM.

**Promoter analysis and functional polymorphic loci identification.** In the promoter upstream of untranslated region (5’ UTR, +1), 33 regulatory regions resulted from 81 TF and 117 TF binding sites (Table S2 in ESM).

A few mutation sites were found to be functional polymorphic loci (Table 2, Table S3 in ESM). These mutations included four TRVs, three indels and five SNPs in five cis-elements, three SNPs in exons and two SNPs in introns.

**Correlations between the content of tanshinones and SmCPS1 polymorphism.** The Mantel test suggested a significant (P < 0.05) relationship between genetic distance matrix (Table S4 in ESM)
and content distance matrix of the chemical components (Table S5 in ESM) except for tanshinone I and dihydrotanshinone (P > 0.05, data not shown) (Table 3) when either all polymorphic sites or only functional polymorphic loci were considered. But the correlation coefficient was higher when only functional polymorphic loci were used. Cryptotanshinone showed the highest values of correlation coefficients; R = 0.440 when all polymorphic loci were counted and R = 0.473 when only functional polymorphic loci were counted. 1,2-dihydrotanshinone showed the lowest values of correlation coefficients (R = 0.285 for all polymorphic loci and R = 0.297 for functional polymorphic loci).

**DISCUSSION**

SmCPS1 is a key gene in the metabolism pathway of tanshinones. Any change of its expression will directly influence the content of tanshinones. Since a promoter contains the basic information about the expression pattern of the gene, we speculated that the variation of tanshinone contents of different *Salvia miltiorrhiza* genotypes is related to the promoter

**Table 2. Characterization of functional polymorphic loci of SmCPS1 and their variation pattern**

<table>
<thead>
<tr>
<th>Location</th>
<th>Regions</th>
<th>Variation pattern</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>–1685 to –1650</td>
<td>MIKC_MADS/C2H2</td>
<td>#</td>
<td>TRV</td>
</tr>
<tr>
<td>–1458</td>
<td>ARR-B</td>
<td>A/G</td>
<td>SNP</td>
</tr>
<tr>
<td>–1306</td>
<td>CPP</td>
<td>A/T</td>
<td>SNP</td>
</tr>
<tr>
<td>–1092 to –1091</td>
<td>MYB</td>
<td>T/C, A/C</td>
<td>SNP</td>
</tr>
<tr>
<td>–1070 to –1065</td>
<td>AP2/MIKC-MAC</td>
<td>(A)6/–</td>
<td>TRV</td>
</tr>
<tr>
<td>–1015 to –1003</td>
<td>TCP</td>
<td>AGTATGACCACAC/–</td>
<td>Indel</td>
</tr>
<tr>
<td>–687</td>
<td>Dof</td>
<td>T/–</td>
<td>Indel</td>
</tr>
<tr>
<td>–553</td>
<td>TCP</td>
<td>C/T</td>
<td>SNP</td>
</tr>
<tr>
<td>–541</td>
<td>MIKC_MADS</td>
<td>G/A</td>
<td>SNP</td>
</tr>
<tr>
<td>–269 to –267</td>
<td>C2H2</td>
<td>TAC/–</td>
<td>Indel</td>
</tr>
<tr>
<td>–204 to –142</td>
<td>(ATT)n</td>
<td>n = 0, 9, 12, 21</td>
<td>TRV</td>
</tr>
<tr>
<td>–22 to –21</td>
<td>TATA-box</td>
<td>AT/–</td>
<td>TRV</td>
</tr>
<tr>
<td>168</td>
<td>exon</td>
<td>C/T</td>
<td>SNP</td>
</tr>
<tr>
<td>179</td>
<td>intron</td>
<td>C/T</td>
<td>SNP</td>
</tr>
<tr>
<td>287</td>
<td>exon</td>
<td>G/C</td>
<td>SNP</td>
</tr>
<tr>
<td>328</td>
<td>exon</td>
<td>T/A</td>
<td>SNP</td>
</tr>
<tr>
<td>362</td>
<td>intron</td>
<td>T/C</td>
<td>SNP</td>
</tr>
</tbody>
</table>

*S* untranslated region (+1); # see Table S3 in ESM for details; SNP – single nucleotide polymorphism; TRV – tandem repeat variation; Indel – insertion/deletion variation

**Table 3. The Mantel test between genetic distance matrix and Euclidean distance matrix of chemical component content in *Salvia miltiorrhiza***

<table>
<thead>
<tr>
<th>Chemical component</th>
<th>Genetic distance matrix based on all polymorphic loci</th>
<th>Genetic distance matrix based on functional polymorphic loci</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R²</td>
</tr>
<tr>
<td>1,2-dihydrotanshinone</td>
<td>0.028</td>
<td>0.285</td>
</tr>
<tr>
<td>Tanshinone IIA</td>
<td>0.034</td>
<td>0.345</td>
</tr>
<tr>
<td>Cryptotanshinone</td>
<td>0.042</td>
<td>0.440</td>
</tr>
<tr>
<td>Tanshinones*</td>
<td>0.018</td>
<td>0.368</td>
</tr>
<tr>
<td>Total tanshinones*</td>
<td>0.018</td>
<td>0.374</td>
</tr>
</tbody>
</table>

*Mantel correlation coefficient; *significant level (P < 0.05); *tanshinones is the combination of all 5 tanshinones that were investigated; *total tanshinones is the sum of contents of all 5 tanshinones
polymorphism of SmCPS1. Our results that the polymorphic sites in SmCPS1 promoter are in correlation with the variation of tanshinone contents except for tanshinone I and dihydrotanshinone supported the hypothesis.

Functional polymorphic loci are polymorphic loci located in the regulatory regions or coding regions of a gene and may change the expression pattern or function of the gene (Kage et al. 2016). Changes in exons, cis-elements, or introns of a gene may change the function or expression pattern of the gene (Le Hir et al. 2003; Nithianantharajah & Hannan 2007; Wittkop & Kalay 2011). Hence the polymorphic loci such as SNPs, indels and TRVs were considered as functional polymorphic loci. According to the position weight matrix of cis-element, some sites contain alternative bases. Since no information about the effect of these alternative bases on the function of a cis-element was available, a key binding site was restricted on a cis-element locus where no alternative bases exist in this study. Such polymorphic loci in the key binding sites of cis-elements were considered as functional polymorphic loci.

Both genetic distance matrices generated based on all polymorphic loci and functional polymorphic loci showed the same significant relationships with the tanshinone contents. But the correlation coefficient was slightly higher when only functional polymorphic loci were considered. Considering the fact that a former study based on the whole genome polymorphism failed to find significant associations between DNA markers and the content of tanshinones (Zhang et al. 2013), it is obvious that all polymorphic loci included in the calculation concealed a correlation between tanshinones and truly informative (functional) polymorphic loci. Hence, it is necessary to filter out informative polymorphic loci from non-informative ones when studying DNA markers.

Different correlations between SmCPS1 functional polymorphism and tanshinone contents of S. miltiorrhiza may be the result of different distance between these chemical components and SmCPS1 in the downstream pathway (Yang et al. 2013). If there is a stronger correlation suggesting a closer distance or a weaker correlation suggesting a longer distance between a chemical component and SmCPS1, our data suggest that the distances between the different tanshinones and SmCPS1 are as follows: cryptotanshinone < tanshinone IIA < 1,2-dihydrorotanshinone < tanshinone I and dihydrotanshinone. This supports part of the hypothesis from Yang et al. (2013) about the relationship: cryptotanshinone < tanshinone IIA < tanshinone I. Though dihydrotanshinone and 1,2-dihydrorotanshinone were not included in this hypothesis, our results provided a probable place in the downstream pathway of tanshinone biosynthesis (see above).

This study is the first to find out correlations between the tanshinone content and polymorphism in SmCPS1. It exhibited the potential of SmCPS1 sequence polymorphism to be developed as a new kind of DNA marker and will be of great value in the breeding and cultivation of S. miltiorrhiza. In conclusion, exploring the relationship between sequence polymorphism and phenotype provides us a new thread in establishing a DNA marker.

References


Received for publication July 10, 2017
Accepted after corrections March 5, 2018
Published online June 13, 2018