Cloning and Bioinformatics Analysis of the Psammosilene tunicoides Transcription Factor ptMYC2 Postprint

Authors: Meng Wenjun, Zhang Aili, Jiang Lexiao, Zhu Wenjie, Duan Li, Qian Zigang

Date: 2018-12-19T00:00:00+00:00

Abstract

Psammosilene tunicoides is an important constituent of various Chinese patent medicines including "Yunnan Baiyao", with triterpenoid saponins as its active components. MYC-like transcription factors play a crucial role in regulating the accumulation of plant triterpenoid secondary metabolites. This study, based on transcriptome sequencing data of Psammosilene tunicoides, cloned two full-length genes of the PtMYC2 transcription factor; bioinformatics software was utilized to perform preliminary predictive analysis on the homology, physicochemical properties, hydrophobicity, transmembrane structure, subcellular localization, domains, and target gene binding sites of these two transcription factors. Experimental results indicated that the proteins encoded by the two transcription factors are hydrophilic, lack transmembrane regions, are non-secretory proteins, and contain no signal peptides; the subcellular localization of the two transcription factors is in the nucleus; domain analysis revealed that both genes contain the bHLH family domain; it was predicted that promoters of genes such as HMGR, FPS, SF, and FPS in the triterpenoid saponin synthesis pathway of Psammosilene tunicoides may harbor E-box specific binding sites for MYC2. This experiment will establish a foundation for further investigation of the regulatory mechanism of the PtMYC2 gene in the triterpenoid saponin synthetic metabolic pathway of Psammosilene tunicoides.

Full Text

Preamble

DOI:10.11931/guihaia.gxzw201810005

Cloning and Bioinformatics Analysis of the Psammosilene tunicoides Transcription Factor ptMYC2

MENG Wenjun, ZHANG Aili, JIANG Lexiao, ZHU Wenjie, DUAN Li, QIAN Zigang*

(Yunnan University of Traditional Chinese Medicine/Engineering Center for Superior Germplasm Breeding of Chinese Medicinal Materials, Kunming 650500, China)

CLC number: Q943 Document code: A

Abstract: *Psammosilene tunicoides* is an important component of various Chinese patent medicines such as "Yunnan Baiyao," with triterpenoid saponins as its active ingredients. MYC transcription factors play crucial roles in regulating plant secondary metabolic accumulation of triterpenoids. Based on transcriptome sequencing data of *P. tunicoides*, this study cloned two full-length genes of the ptMYC2 transcription factor and conducted preliminary predictive analysis of homology, physicochemical properties, hydrophobicity, transmembrane structure, subcellular localization, domains, and target gene binding sites using bioinformatics software. The experimental results showed that the proteins encoded by the two transcription factors are hydrophilic, lack transmembrane regions, are non-secretory, and contain no signal peptides. Both transcription factors localize to the nucleus. Domain analysis revealed that both genes contain the bHLH family domain. We predicted that promoters of genes involved in the triterpenoid saponin synthesis pathway of *P. tunicoides*, including HMGR, FPS, SE, and -AS, may contain E-box binding sites for MYC2. This study provides a foundation for further investigation of the regulatory mechanisms of ptMYC2 genes in the triterpenoid saponin metabolic pathway.

Keywords: *Psammosilene tunicoides*, transcription factor, MYC2, triterpenoid saponin, biosynthetic pathway, bioinformatics analysis

Funding: This work was supported by the Yunnan Provincial Department of Science and Technology-Yunnan University of Traditional Chinese Medicine Applied Foundation Joint Special Project (2017FF117-001; 2017FF117-030) and the Yunnan Applied Basic Youth Project (2015FD036).

Author Contributions: MENG Wenjun (1992-), female, from Qujing, Yunnan, Master' s degree candidate, engaged in research on development and utilization of traditional Chinese medicine resources, (E-mail) 1909242258@qq.com. *Corresponding author: QIAN Zigang, Ph.D., Professor, engaged in research on development and molecular biology of traditional Chinese medicine resources, (E-mail) qianzig@aliyun.com.

Introduction

Psammosilene tunicoides is a native endangered medicinal plant in Yunnan Province and an important component of various famous Chinese patent medicines such as "Yunnan Baiyao." Its active components are oleanane-type triterpenoid saponins, which exhibit significant pharmacological activities including analgesic and anti-inflammatory effects [?, ?]. Due to these prominent pharmacological properties, the medicinal resources of this plant have become severely depleted, making it imperative to elucidate the regulatory mechanisms of triterpenoid saponin biosynthesis to enable directed accumulation of these valuable compounds.

Jasmonic acid (JA) and its methyl ester derivative (MeJA) are important plant hormones that serve as key signal transduction molecules in plant responses to biotic and abiotic stresses [?, ?]. These compounds regulate the synthesis of secondary metabolites such as flavonoids and terpenoids, thereby enhancing plant stress resistance. The MYC2 transcription factor is a crucial component of the JA signaling pathway, regulating downstream gene expression by binding to target gene promoters [?, ?]. Under normal conditions, nuclear JAZ (Jasmonate-ZIM Domain) proteins and co-repressors such as TOPLESS (TPL) or TPL-related proteins (TPLs) suppress JA signaling by preventing transcription factors from binding to target gene promoters [?, ?]. Conversely, when plants are stressed, JA conjugates with isoleucine to form JA-Ile, which promotes the formation of a conjugate complex between JAZ proteins and Skp1/Cullin1/F-box protein COI1 (SCF COI1). This complex is degraded by the 26S proteasome, releasing transcription factors to regulate stress-responsive gene expression and promote secondary metabolite accumulation [?, ?].

Previous studies have reported on MYC2 transcription factors regulating secondary metabolite synthesis in various plants. For instance, nicotine synthesis in tobacco requires NtMYC2, which forms nuclear complexes with the repressor NtJAZ1 and regulates multiple jasmonate-inducible steps in nicotine biosynthesis [?, ?]. Hong et al. (2012) demonstrated that MYC2 activates the expression of sesquiterpene synthases TPS21 and TPS11 by directly binding to their promoter regions, thereby regulating terpene synthase gene expression and volatile sesquiterpene synthesis. In *Catharanthus roseus*, CrMYC2 acts as an early MeJA-responsive factor that regulates the expression of ORCA genes, which in turn control the expression of enzymes involved in indole alkaloid (TIA) biosynthesis [?, ?]. However, no studies on *P. tunicoides* transcription factors have been reported to date. While our research group has previously cloned multiple key enzyme genes in the oleanane-type triterpenoid saponin biosynthetic pathway of *P. tunicoides*, the transcriptional regulatory mechanisms of these genes remain unclear. Therefore, this study cloned the ptMYC2 gene from *P. tunicoides* and performed bioinformatics analysis to predict potential target genes that may interact with ptMYC2, laying a foundation for elucidating the transcriptional regulatory mechanisms of triterpenoid saponin metabolism in this important medicinal plant.

Materials and Methods

1.1 Plant Materials

Psammosilene tunicoides samples were collected from Lijiang City, Yunnan Province and identified as *Psammosilene tunicoides* W. C. Wu et C. Y. Wu (Caryophyllaceae) by Professor Qian Zigang of Yunnan University of Traditional Chinese Medicine.

1.2 Major Instruments

High-speed refrigerated centrifuge (Eppendorf), electrophoresis apparatus (BIO-RAD), DYC-33A mini electrophoresis tank (BIO-RAD), gel imaging system (BIO-RAD), PCR thermal cycler (BIO-RAD), pipettes (Eppendorf), and general consumables purchased from Kunming Dingguo Biotechnology Co., Ltd.

1.3 Major Reagents

TaKaRa Minibest Universal RNA Extraction Kit (Lot: AK1602), TaKaRa PrimeScript™ 1st Strand cDNA Synthesis Kit (Lot: AK4501), TaKaRa MiniBEST Agarose Gel DNA Extraction Kit Ver 4.0 (Lot: AK1901), TaKaRa MiniBEST Plasmid Purification Kit Ver 4.0, DL2000 DNA Marker (Lot: A2101A), and other reagents.

1.4 Experimental Methods

1.4.1 Primer Design Based on two ptMYC2 gene sequences identified in the *P. tunicoides* transcriptome (see Appendix), specific primer pairs were designed for amplification (Table 1).

1.4.2 Cloning of *P. tunicoides* Transcription Factor ptMYC2 Total RNA was extracted from *P. tunicoides* roots using the RNA extraction kit. First-strand cDNA was synthesized from total RNA using Takara' s PrimeScript™ 1st Strand cDNA Synthesis Kit. Using this cDNA as template, full-length amplification of two ptMYC2 genes was performed with the designed primers. The PCR reaction systems are shown in Table 2 and Table 3 .

After optimization, the final PCR conditions were: ptMYC2-1: 98°C for 1 min, followed by 32 cycles of 98°C for 10 s, 48°C for 15 s, and 72°C for 45 s, with a final extension at 72°C for 10 min and termination at 4°C. ptMYC2-2: 98°C for 1 min, followed by 32 cycles of 98°C for 10 s, 55°C for 15 s, and 72°C for 45 s, with a final extension at 72°C for 10 min and termination at 4°C. PCR products were analyzed by 1.0% agarose gel electrophoresis, purified using the gel extraction kit (TaKaRa), and sequenced.

1.4.3 T-Vector Construction of ptMYC2 Purified PCR products were ligated into the pMD™18-T vector in a 10 L reaction system. The ligation mixture was transformed into competent *E. coli* cells, plated on LB solid medium containing ampicillin, and incubated at 37°C. Single colonies were selected for overnight culture, plasmids were extracted using the plasmid purification kit, and positive clones were verified by PCR and sequencing.

1.4.4 Bioinformatics Analysis of ptMYC2 Comprehensive bioinformatics analysis was performed on the two ptMYC2 genes. Sequence homology was analyzed using BLAST on the NCBI website (https://www.ncbi.nlm.nih.gov/). Open reading frames (ORFs) were identified using ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/). Physicochemical properties were predicted using ProtParam (https://web.expasy.org/protparam/). Hydrophobicity analysis was conducted using ProtScale (https://web.expasy.org/protscale/). Transmembrane regions were predicted using TMHMM (http://www.cbs.dtu.dk/services/TMHMM/). Signal peptide analysis was performed using SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP-3.0/). Subcellular localization was predicted using TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/) and Cell-PLoc 2.0 (http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/). Conserved domains were analyzed using SMART (http://smart.embl-heidelberg.de/). Secondary structure was predicted using CFSSP (http://www.biogem.org/tool/chou-fasman/index.php). Tertiary structure was modeled using SWISS-MODEL (https://swissmodel.expasy.org/interactive). Phylogenetic trees were constructed using MEGA 5.0 software. To identify potential target genes, the highest similarity species for each gene were identified through BLAST searches, and 2000 bp upstream regions were downloaded as putative promoter sequences. Binding sites for ptMYC2 were predicted using the JASPAR 2016 server.

Results

2.1 Cloning of *P. tunicoides* Transcription Factor ptMYC2

Total RNA extracted from *P. tunicoides* roots showed clear bands on 1.0% agarose gel electrophoresis, indicating high quality. After reverse transcription to cDNA and PCR amplification, products were analyzed by 1.0% agarose gel electrophoresis and verified by sequencing. The full-length sequences were cloned into the T-vector, and sequencing results confirmed that the amplified sequences matched the transcriptome data.

2.2 Bioinformatics Analysis of ptMYC2

2.2.1 Homology Analysis BLAST analysis revealed that both ptMYC2 transcription factors share homology with MYC2 transcription factors from other plant species, confirming the successful cloning of ptMYC2-1 and ptMYC2-2 from *P. tunicoides*.

2.2.2 Physicochemical Properties of Encoded Proteins Analysis of ptMYC2-1 using the ProtParam online tool revealed an open reading frame of 1,713 bp encoding 570 amino acids. The predicted molecular formula was C H N O S , with a molecular weight of 63,746.8 Da and a theoretical isoelectric point of 6.06. Serine (Ser) comprised 11.4% of the amino acid composition. The instability index was 47.66, indicating an unstable protein, with an aliphatic index of 76.91.

Similar analysis of ptMYC2-2 showed an ORF of 1,902 bp encoding 633 amino acids. The predicted molecular formula was C H N O S , with a molecular weight of 69,330.42 Da and a theoretical isoelectric point of 5.33. Serine was the most abundant amino acid at 10.6%. The instability index was 52.4, also indicating an unstable protein, with an aliphatic index of 70.88.

2.2.3 Hydrophobicity Analysis ProtScale analysis of the amino acid sequences revealed a typical hydrophilic region around position 379 in ptMYC2-1 and around position 89 in ptMYC2-2 (Figure 1 [Figure 1: see original paper] and Figure 2 [Figure 2: see original paper]).

2.2.4 Transmembrane Region Analysis TMHMM2.0 prediction indicated that neither protein contains transmembrane regions, suggesting they are not membrane proteins.

2.2.5 Signal Peptide and Subcellular Localization Prediction SignalP 3.0 analysis predicted no signal peptides for either ptMYC2-1 or ptMYC2-2. Both neural network and hidden Markov models confirmed that these proteins are non-secretory and lack signal peptides. TargetP 1.1 Server analysis indicated localization to other organelles, while Cell-PLoc 2.0 specifically localized both factors to the nucleus.

2.2.6 Domain Analysis SMART software predicted a highly conserved bHLH domain from positions 39-218 in ptMYC2-1 and from positions 28-208 in ptMYC2-2 (Figure 3 [Figure 3: see original paper] and Figure 4 [Figure 4: see original paper]), confirming these as typical bHLH family members.

2.2.7 Secondary and Tertiary Structure Prediction CFSSP secondary structure prediction revealed that ptMYC2-1 contains 70% -helix, 57.2% -sheet, and 14.0% turns, while ptMYC2-2 contains 65.9% -helix, 51.3% -sheet, and 14.4% turns. Both proteins exhibit mixed secondary structures. Using *Arabidopsis* AtMYC2 protein (Accession: NM-102998.3) as a reference template, SWISS-MODEL predicted the three-dimensional structures (Figure 5 [Figure 5: see original paper], Figure 6 [Figure 6: see original paper], and Figure 7 [Figure 7: see original paper]). The bHLH regions of *P. tunicoides* ptMYC2 proteins showed structural similarity to *Arabidopsis* AtMYC2. The helix-loop-helix structures in both similar and different bHLH transcription factors can interact to form homodimers or heterodimers that bind to different regions of target gene promoters, thereby exerting transcriptional regulatory functions.

2.2.8 Phylogenetic Tree Construction Phylogenetic analysis using MEGA 5.0 Neighbor-Joining method showed that both ptMYC2-1 and ptMYC2-2 are closely related to quinoa (*Chenopodium quinoa*) MYC2 (Figure 8 [Figure 8: see original paper] and Figure 9 [Figure 9: see original paper]). Since quinoa belongs to the Amaranthaceae family while *P. tunicoides* belongs to Caryophyllaceae, both representing relatively primitive taxonomic groups with close phylogenetic relationships, the cloned ptMYC2 proteins cluster with quinoa cqMYC2.

2.2.9 Transcription Factor-Target Gene Binding Site Analysis MYC transcription factors regulate target genes by binding to E-box motifs in their promoters. Since the *P. tunicoides* genome has not been sequenced, we performed BLAST searches for key enzyme genes in the triterpenoid saponin biosynthetic pathway: HMGR, SE, FPS, and -AS. HMGR showed highest similarity to *Populus euphratica*, SE1 to *Beta vulgaris* subsp., while SE2, FPS, and -AS showed highest similarity to *Chenopodium quinoa*. The 2000 bp upstream regions of these reference genes were downloaded as putative promoter sequences. Binding site prediction results are shown in Table 4 .

Discussion

Psammosilene tunicoides is a rare and endangered medicinal herb in southwestern China and a crucial component of famous Chinese patent medicines. Its active ingredients are pentacyclic triterpenoid saponins. Due to significant pharmacological activities, wild populations have been severely depleted, leading to its classification as an endangered species by IUCN. MYC transcription factors are core components of the jasmonate hormone response pathway. Recent studies in tobacco [?, ?], *Arabidopsis* [?, ?], *Catharanthus roseus* [?, ?], and *Taxus cuspidata* [?, ?] have demonstrated that MYC2 transcription factors regulate secondary metabolite synthesis. Among the various MYC transcription factors, MYC2 is one of the most extensively studied. This research investigated ptMYC2 transcription factors from *P. tunicoides* to identify regulators of its active compound biosynthesis, providing a foundation for directed accumulation of these valuable metabolites.

This study successfully cloned two ptMYC2 transcription factors with complete ORFs. ptMYC2-1 and ptMYC2-2 contain ORFs of 1,713 bp and 1,902 bp, encoding 570 and 633 amino acid residues with molecular weights of 63.75 kDa and 69.33 kDa, respectively. SMART analysis confirmed that both proteins contain a highly conserved bHLH domain, characteristic of bHLH family members. Tertiary structure prediction revealed that both ptMYC2 proteins share structural similarity with *Arabidopsis* AtMYC2, particularly in the conserved helix-loop-helix domain that enables binding to target gene promoters and transcriptional regulation [?, ?].

BLASTX homology comparison showed that the two MYC2-encoded proteins share over 50% identity with tobacco (*Nicotiana tabacum*) NtMYC2, which regulates nicotine biosynthesis [**?**, **?**], suggesting these ptMYC2 transcription factors may also be involved in regulating secondary metabolism in *P. tunicoides*.

MYC2 transcription factors regulate gene expression primarily through DNA binding. These specific binding regions are typically located in target gene promoters [**?**, **?**]. Therefore, predicting MYC2-promoter interactions is essential for understanding transcriptional regulation. bHLH proteins are classified into groups A, B, C, D, E, and F based on their DNA binding patterns. Group B bHLH proteins bind to E-box motifs with the consensus sequence 5 -CACGTG-3 , including the MYC family [**?**, **?**]. Using this E-box motif, we selected key enzyme genes from the triterpenoid saponin biosynthetic pathway as potential targets. BLAST searches identified the most similar species for each gene, and 2000 bp upstream regions were analyzed as putative promoters. The prediction revealed potential E-box binding sites in the promoters of HMGR, SE, FPS, and -AS genes, suggesting ptMYC2 may interact with and regulate these genes. Shen et al. [**?**, **?**] demonstrated that *Artemisia annua* AaMYC2 binds to promoters of CYP71AV1 and DBR2 genes in the artemisinin biosynthetic pathway. Similarly, our binding site predictions provide valuable candidate genes for future functional studies.

In summary, this study cloned two ptMYC2 transcription factors and performed comprehensive bioinformatics analysis, providing a scientific basis for future functional validation experiments such as yeast two-hybrid assays.

References

CHINI A, FONSECA S, FERNANDEZ G, et al, 2007. The JAZ family of repressors is the missing link in jasmonate signalling [J]. Nature, 448(7154): 666-671.

DEVOTO A, TURNER JG, 2010. Jasmonate-regulated Arabidopsis stress signaling network [J]. Physiol Plant, 123(2): 161-172.

HEDHILI S, MATTEI MVD, COUDERT Y, et al, 2010. Three non-autonomous signals collaborate for nuclear targeting of CrMYC2, a Catharanthus roseus, bHLH transcription factor [J]. BMC Res Notes, 3(1): 301-309.

HONG GJ, XUE XY, MAO YB, et al, 2012. Arabidopsis MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression [J]. Plant Cell, 24(6): 2635-2648.

LEDENT V, VERVOORT M, 2001. The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis [J]. Genome Res, 11(5): 754-770.

LENKA SK, NIMS NE, VONGPASEUTH K, et al, 2015. Jasmonate-responsive expression of paclitaxel biosynthesis genes in Taxus cuspidata cultured cells is negatively regulated by the bHLH transcription factors TcJAMYC1, TcJAMYC2, and TcJAMYC4 [J]. Front Plant Sci, 6: 115.

NIMS E, VONGPASEUTH K, ROBERTS SC, et al, 2009. WITHDRAWN: TcJAMYC: A bHLH transcription factor that activates paclitaxel biosynthetic pathway genes in yew [J]. J Biol Chem, 290(33): 1-19.

SHEN Q, LU X, YAN T, et al, 2016. The jasmonate-responsive AaMYC2 transcription factor positively regulates artemisinin biosynthesis in Artemisia annua [J]. New Phytol, 210(4): 1269-1281.

SHEN Q, LU X, ZHANG L, et al, 2012. Advances in research on the function of MYC2 transcription factors in plants[J]. J Shanghai Jiaotong Univ(Agric Sci Ed), 30(6): 51-57. [沈乾，陆续，张凌，等，2012. 植物中 MYC2 转录因子功能研究进展 [J]. 上海交通大学学报：农业科学版，30(6)：51-57.]

STEVENS JD, ROALSON EH, SKINNER MK, 2010. Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: genomic approach to cellular differentiation[J]. Differentiation, 76(9): 1006-1022.

THINES B, KATSIR L, MELOTTO M, et al, 2007. JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling [J]. Nature, 448(7154): 661-665.

ZHANG H, HEDHILI S, MONTIEEL G, et al, 2011. The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in Catharanthus roseus [J]. Plant J, 67(1): 61-71.

ZHANG HB, BOKOWIEC MT, RUSHTON PJ, et al, 2012. Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple jasmonate-inducible steps in nicotine biosynthesis [J]. Mol Plant, 5(1): 73-84.

ZHANG J, CAO W, TIAN J, et al, 2012. Evaluation of novel saponins from Psammosilene tunicoides and their analogs as immunomodulators [J]. Int Immunopharmacol, 14(1): 21-26.

Note: Figure translations are in progress. See original paper for figures.