Cloning and Expression Analysis of the Ml3GT1 Gene in Magnolia liliflora 'Hongyuanbao' Post-print

Authors: Zhuowei Wang, Dai Mengyi, Cheng Shaoyu, Wang Xiaode, Wang Yaling, Shen Yamei, Zhang Chao

Date: 2021-08-09T00:00:00+00:00

Abstract

UDP-flavonoid 3-O-glucosyltransferase (3GT) is one of the important catalytic enzymes in the anthocyanin biosynthesis pathway. To investigate its role in the anthocyanin biosynthesis pathway of Magnolia liliflora, this study used the Magnolia liliflora cultivar 'Hongyuanbao' as material, designed primers based on the 3GT sequence obtained from transcriptome sequencing, cloned the structural gene Ml3GT1 in the anthocyanin biosynthesis pathway using RT-PCR technology, and conducted bioinformatic and expression pattern analyses. The results showed: (1) The cDNA sequence of the Ml3GT1 gene was 1,863 bp in length, with the longest open reading frame (ORF) being 1,374 bp, encoding a peptide chain of 457 aa, with a relative molecular mass of 49.37 kDa and a theoretical isoelectric point (pI) of 6.04. (2) Amino acid sequence alignment revealed that it possesses the typical plant secondary product glycosyltransferase signal sequence (PSPG box). (3) Phylogenetic analysis results indicated that the Ml3GT1 protein clustered in the same clade with 3GT proteins from species such as freesia, petunia, and sweet potato. (4) qRT-PCR results showed that the expression of the Ml3GT1 gene exhibited spatiotemporal specificity, with the highest expression level in flowers, low expression in young and old leaves, and almost no expression in roots and stems; moreover, as the flower developed, the expression level of the Ml3GT1 gene showed a trend of first decreasing and then increasing, reaching its peak at the full-bloom stage. These results suggest that Ml3GT1 may be involved in the 3-O-glycosylation of flavonoids, and the findings of this study will lay a foundation for flower color breeding research in Magnolia species.

Full Text

Abstract

UDP-flavonoid 3-O-glucosyltransferase (3GT) is one of the important catalytic enzymes in the anthocyanin biosynthesis pathway. To study the function of 3GT in anthocyanin biosynthesis of *Magnolia liliflora*, *M. liliflora* 'Hongyuanbao' was employed as material. Primers were designed based on the 3GT sequence obtained from the transcriptome database of *M. liliflora* 'Hongyuanbao' , and the structural gene *Ml3GT1* in the anthocyanin biosynthesis pathway was cloned by RT-PCR (reverse transcription-PCR). Bioinformatics and expression pattern analyses were subsequently performed. The results were as follows: (1) The cDNA of *Ml3GT1* was 1,863 bp, and the open reading frame was 1,374 bp, encoding 457 amino acid residues. The relative molecular weight of Ml3GT1 was 49.37 kDa, and its isoelectric point was 6.04. (2) The deduced amino acid sequence of Ml3GT1 contains a conserved plant secondary product glycosyltransferase signature sequence (PSPG box). (3) Results of the phylogenetic analysis showed that Ml3GT1 was closely related to 3GT proteins from *Freesia hybrida*, *Petunia* × *hybrida*, and *Ipomoea batatas*. (4) Results of fluorescence quantitative PCR revealed that *Ml3GT1* has spatio-temporal specificity, with the highest expression level in flowers, lower expression levels in young and old leaves, and trace expression levels in roots and stems. With flower development, the expression level of *Ml3GT1* decreased first, then increased, and showed the highest expression level at the fully-opening stage. These results suggest that Ml3GT1 may be involved in flavonoid glycosylation. This study will lay a foundation for flower color breeding of *Magnolia* plants.

Keywords: *Magnolia liliflora*, flower color, glycosyltransferase, gene cloning, expression analysis

Introduction

Flower color is one of the most important quality traits in ornamental plants. Anthocyanins, which are glycosides composed of anthocyanidins and sugars, represent a crucial class of pigments that determine flower color (戴思兰和洪艳，2016). To date, research on anthocyanin biosynthesis pathways has been relatively comprehensive. As shown in Figure 1 [Figure 1: see original paper], three main branches of the anthocyanin biosynthetic pathway produce cyanidin, pelargonidin, and delphinidin. Cyanidin can be methylated to form peonidin, while petunidin and malvidin are generated through varying degrees of methylation of delphinidin, collectively constituting the six major anthocyanidins found in nature (朱丽娟等，2012). Glycosylation of anthocyanidins to form stable anthocyanins plays a critical role in color development (招雪晴等，2017).

In the process of flower color formation in plants, UDP-flavonoid 3-O-glucosyltransferase (3GT) typically acts at the final step of the anthocyanin biosynthesis pathway, catalyzing the transfer of sugar molecules from UDP-

glucose to the C3 hydroxyl position of anthocyanidins, thereby converting anthocyanidins into the first stable anthocyanins in the pathway (Sui et al., 2011). Anthocyanidins are unstable under natural conditions and readily form anthocyanins through glycosidic bonds with sugars (Nakatsuka et al., 2008). The main monosaccharides comprising anthocyanins include glucose, rhamnose, galactose, xylose, and arabinose (庄维兵等，2018). From a chemical perspective, sugar conjugation enhances the stability and water solubility of anthocyanins. Anthocyanins most commonly undergo O-glycosylation at the C3 position, followed by glycosylation at the C5 position. Glycosylation facilitates the transfer of anthocyanins from their production site in the cytoplasm to the vacuole (Nakatsuka et al., 2008), causing a slight shift in color toward red (Tanaka et al., 2008). Thus, glycosylation plays a key role in flower color formation.

The first enzyme discovered to catalyze glycosylation of the anthocyanin C3 hydroxyl group was maize (*Zea mays*) Bronze-1 (X13500). Due to reduced anthocyanin accumulation, Bronze mutants exhibit pale-colored kernels (Dooner & Nelson, 1977; Larson & Coe, 1977). To date, 3GT genes have been cloned from various ornamental plants, including petunia (*Petunia hybrida*) (Jonsson et al., 1984), three-flowered gentian (*Gentiana triflora*) (Tanaka et al., 1996), Dutch iris (*Iris hollandica*) (Yoshihara et al., 2005), freesia (*Freesia hybrida*) (Sui et al., 2011), *Paeonia delavayi* (王毅等，2017), and grape hyacinth (*Muscari armeniacum*) (杜灵娟等，2017). In most plants, 3GT expression is positively correlated with anthocyanin accumulation, as seen with *Fh3GT* and *MdUFGT*; however, in some plants such as *Ze3GT*, this correlation is not observed (Ban et al., 2009; Sui et al., 2011; Hu et al., 2016; Qian et al., 2021). *PeUFGT3* plays an important role in red color formation in *Phalaenopsis equestris*; suppressing *PeUFGT3* expression leads to significant reduction in anthocyanin content (Chen et al., 2011). Shu et al. used VIGS technology to silence the *PsUF3GT* gene in tree peony (*Paeonia suffruticosa*), resulting in varying degrees of reduction in both 3G-type and 3G5G-type glycosides (舒庆艳等，2018). These findings demonstrate that 3GT genes play important roles in flower color formation.

Magnolia liliflora 'Hongyuanbao' is a perennial deciduous shrub belonging to the family Magnoliaceae and genus *Magnolia*, and is a cultivated variety of *M. liliflora* with deeper flower color than the wild type. Our previous research showed that both *M. liliflora* and *M. liliflora* 'Hongyuanbao' tepals contain four types of anthocyanins modified with rutinoside at the C3 position: cyanidin 3-O-rutinoside-5-O-glucoside (Cy3Ru5G), cyanidin 3-O-rutinoside (Cy3Ru), peonidin 3-O-rutinoside-5-O-glucoside (Pn3Ru5G), and peonidin 3-O-rutinoside (Pn3Ru) (Wang et al., 2019). The C3 rutinoside modification of anthocyanidins occurs in two steps: first, O-glucosylation at the C3 position catalyzed by 3GT, after which rhamnosyltransferase (3-O-rhamnosyltransferase, 3RT) can transfer O-rhamnoside to connect with O-glucoside, forming a rutinoside group (Yamazaki et al., 2002). We therefore hypothesize that the 3GT gene plays an important role in the synthesis of these four anthocyanins during flower color formation in *M. liliflora* 'Hongyuanbao' . To further elucidate its function,

this study screened and cloned the 3GT gene based on transcriptome data and performed expression analysis.

Materials and Methods

1.1 Plant Materials

Ten-year-old seedlings of *Magnolia liliflora* 'Hongyuanbao' were planted at the Pingshan Experimental Base of Zhejiang A&F University (119°42 54.67 E, 30°15 50.09 N). Healthy, disease-free plants were selected, and outer petals were collected at five developmental stages in April 2019: bud stage (S1), color-showing stage (S2), initial-opening stage (S3), half-opening stage (S4), and full-bloom stage (S5) (Figure 2 [Figure 2: see original paper]). Samples were wrapped in tinfoil, flash-frozen in liquid nitrogen, and stored at -80 °C for RNA extraction and fluorescence quantitative analysis. Since *M. liliflora* 'Hongyuanbao' flowers and leaves emerge simultaneously, root, stem, old leaf, and young leaf tissues were also collected.

1.2 Total RNA Extraction and cDNA Synthesis

Total RNA was extracted from petals at stages S1–S5 and from root, stem, old leaf, and young leaf tissues using the UltraClean Polysaccharide and Phenol Plant RNA Purification Kit (DNA-free) (NHUC002S) from Novogene. Reverse transcription was performed using PrimeScript™ RT Master Mix (Perfect Real Time) (TaKaRa code: RR036A). The resulting cDNA was stored at -20 °C for subsequent use.

1.3 Cloning of the *Ml3GT1* Gene

Using the transcriptome database of *M. liliflora* 'Hongyuanbao' petals constructed by our research group (unpublished results), we searched gene annotations against the Nr (non-redundant protein database) database using "anthocyanidin/flavonoid 3-O-glucosyltransferase" as the query term. The obtained sequence CL3388.Contig1 was further analyzed by blastx alignment using NCBI's BLAST function (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Primers were designed using Prime Prime 5.0 software flanking the open reading frame (ORF) (Table 1) and synthesized by Hangzhou Youkang Biotechnology Co., Ltd. Equal amounts of cDNA from the five flower developmental stages (S1–S5) were mixed, diluted 10-fold, and used as template for RT-PCR amplification. The 20 L PCR reaction mixture contained: 1 L each of forward and reverse primers (10 mol·L^{-1}), 1 L cDNA template, 10 L Premix Taq, and 7 L ddH_2O. The reaction program consisted of initial denaturation at 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 59.6 °C for 30 s, and 72 °C for 2 min, with a final extension at 72 °C for 10 min and cooling at 10 °C for 5 min. After PCR, target fragments were separated by 1% (w/v) agarose gel electrophoresis. Gel extraction was performed using the TaKaRa MiniBEST Agarose Gel DNA Extraction Kit Ver.4.0 (TaKaRa code: No.9762) according to the manufacturer's instructions.

The purified product was ligated into the pMD™18-T Vector (TaKaRa code: No.6011). The ligation product was transformed into *Escherichia coli* DH5α Competent Cells (TaKaRa code: No.9057). Positive clones were selected by blue-white screening and confirmed by colony PCR. Clones with correct bands were sent to Hangzhou Youkang Biotechnology Co., Ltd. for DNA sequencing.

1.4 Bioinformatics Analysis of *Ml3GT1*

Vector sequences were removed using SnapGene 4.1.8 software to obtain the target gene sequence. The nucleotide sequence was translated into protein sequence using DANMAN 7.0 software. The coding region was analyzed using NCBI' s ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/). Conserved domains were analyzed using the CDD protein conserved domain database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Physicochemical properties of the Ml3GT1 protein sequence, including molecular weight, isoelectric point, instability coefficient, aliphatic index, and hydrophobicity, were predicted using the ProtParam online tool (https://web.expasy.org/protparam/). Hydrophobicity/hydrophilicity was analyzed using the protscale online server (https://web.expasy.org/protscale/). Secondary structure was predicted using SOPMA online software (https://npsa-prabi.ibcp.fr/cgi-bin/secpred_{sopma}.pl). Tertiary structure was predicted using the SWISS-MODEL online server (swissmodel.expasy.org). Multiple sequence alignment was performed using DNAMAN 7.0, and a phylogenetic tree was constructed using MEGA 6.0 software with the neighbor-joining (NJ) method and 1,000 bootstrap replicates to obtain branch reliability.

1.5 Expression Analysis of *Ml3GT1* at Different Flowering Stages and in Different Tissues

Relative gene expression was quantified using a LightCycler 480 II (Roche) real-time PCR system. The reaction mixture contained: 2 L template, 0.8 L each of forward and reverse primers, 10 L BCG Qpcr Master Mix (2×), and 6.4 L ddH_2O. qRT-PCR was performed using a two-step method: initial denaturation at 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s, then 95 °C for 5 s, 60 °C for 1 min, and 95 °C for 15 s. The cDNA was diluted in 5-fold gradient series. *MbTEF* was used as the reference gene (王宁杭等，2019), with three biological replicates for each sample. Relative expression levels were calculated using the $2^{-}\Delta$Ct method. Data analysis and graphing were performed using SigmaPlot 14.0 software.

Results

2.1 Cloning and Sequence Analysis of *Ml3GT1*

RT-PCR amplification of *M. liliflora* 'Hongyuanbao' yielded a product of 1,863 bp (Figure 3 [Figure 3: see original paper]). Sequence alignment with the transcriptome sequence CL3388.Contig1 revealed 99.84% nucleotide similarity.

Analysis using NCBI' s ORF finder identified a longest open reading frame of 1,374 bp encoding 457 amino acids. The deduced amino acid sequence showed 100% similarity with the transcriptome sequence. The gene was designated *Ml3GT1* and deposited in GenBank under accession number MW454862.

Conserved domain prediction using NCBI' s CDD database revealed that Ml3GT1 possesses a uridine diphosphate-glycosyltransferase domain (PLN02670) and a UDP-glucuronosyl/glucosyltransferase conserved domain (UDPGT), indicating that Ml3GT1 belongs to the glycosyltransferase superfamily. The presence of a GT1_{Gtf}-like domain suggests it belongs to family 1 glycosyltransferases (GT1s), the largest family in the plant kingdom. Based on three-dimensional folding patterns, glycosyltransferases can be broadly classified into GT-A, GT-B, and GT-C families (Coutinho et al., 2003). Prediction results indicated that Ml3GT1 is a GT-B glycosyltransferase.

BLASTp analysis showed that the Ml3GT1 amino acid sequence shares highest similarity (60.35%) with 3GT from *Cinnamomum micranthum*, and high similarity (49.00%-54.27%) with 3GTs from *Nelumbo nucifera*, *Morella rubra*, *Phoenix dactylifera*, *Vitis labrusca*, and *Ricinus communis*. Multiple sequence alignment using DNAMAN 7.0 software revealed that Ml3GT1 shares 46.20%, 44.20%, 41.60%, 38.40%, 47.20%, and 41.80% similarity with 3GT proteins from *Arabidopsis*, apple, maize, petunia, grape, and three-flowered gentian, respectively (Figure 4 [Figure 4: see original paper]). Similar to other 3GT proteins, Ml3GT1 contains a typical conserved sequence of 44 amino acids at its C-terminus, known as the plant secondary product glycosyltransferase signal sequence (PSPG box), indicating that Ml3GT1 possesses characteristics of plant glycosyltransferases and may be involved in glycosylation modification of secondary metabolites. The PSPG motif of Ml3GT1 is WAPQTMVLGHVALGAFVTHCGWNSVMESITAGVPMICRPFFGDQ. Previous studies have shown that sugar donor specificity is partially determined by the last amino acid residue in the PSPG box: Q indicates glucose as the sugar donor, while H indicates galactose (Kubo et al., 2004). The last residue in the Ml3GT1 PSPG motif is Q, suggesting that Ml3GT1 may be a UDP-flavonoid glycosyltransferase using UDP-glucose as the sugar donor.

2.2 Bioinformatics Analysis of *Ml3GT1*

Physicochemical property analysis using ProtParam revealed that Ml3GT1 encodes 457 amino acids with a molecular formula of $C_{2216}H_{3457}N_{601}O_{638}S_{20}$. The relative molecular weight is 49.37 kDa, and the theoretical isoelectric point (pI) is 6.04, indicating it is an acidic protein. The numbers of negatively charged (Asp + Glu) and positively charged (Arg + Lys) residues are 50 and 45, respectively. Alanine is the most abundant amino acid (10.9%), while tyrosine is the least abundant (0.9%). The protein instability coefficient is 40.71, with values greater than 40 indicating an unstable protein. The grand average of hydropathicity is 0.090, with positive values indicating hydrophobicity; thus, Ml3GT1 is a hydrophobic protein.

Secondary structure prediction using SOPMA indicated that Ml3GT1 consists mainly of α-helices (39.39%) and random coils (38.29%), with extended strands (16.85%) and β-turns (5.47%) (Figure 5 [Figure 5: see original paper]). The Ml3GT1 sequence was submitted to SWISS-MODEL for tertiary structure prediction. The 3D model showed a QMEAN score of -0.88 (Figure 6 [Figure 6: see original paper]), where values closer to 0 (range: -4 to 0) indicate better match and higher reliability between the target and template proteins. The QMEAN score suggests good matching between Ml3GT1 and the template protein UDP-glucose:flavonoid 3-O-glucosyltransferase (2c1x.1.A), with high similarity (52.67%), indicating that Ml3GT1 may be a flavonoid 3-O-glucosyltransferase. The GMQE reliability score was 0.83 (range: 0–1), with values closer to 1 indicating better model quality, suggesting that the model built using this protein template is of good quality. Ml3GT1 possesses the typical Rossmann fold domain of the GT-B family, confirming it as a GT-B family member and consistent with CDD prediction results.

2.3 Phylogenetic Analysis of Ml3GT1 Amino Acid Sequence

A phylogenetic tree was constructed using Ml3GT1 and reported flavonoid pathway glycosyltransferases from other plants. As shown in Figure 7 [Figure 7: see original paper], glycosyltransferases (GTs) are divided into 5GT, 7GT, and 3GT evolutionary branches. Ml3GT1 clustered with *Freesia hybrida* Fh3GT1 in the 3GT major branch, which also includes grape Vv3GT, strawberry Fa3GT, petunia Ph3GT, and sweet potato Ib3GT. Ml3GT1 showed distant relationships with other flavonoid glycosyltransferases (5GTs and 7GTs), suggesting it may function similarly to Fh3GT1 as a member of the 3GT glycosyltransferase family involved in flavonoid 3-O glycosylation (Sun et al., 2016).

2.4 Expression Analysis of *Ml3GT1* at Different Flowering Stages and in Different Tissues

Using cDNA from different flowering stages (S1–S5) and from old leaves, young leaves, roots, and stems of *M. liliflora* 'Hongyuanbao' as templates, we investigated the expression pattern of *Ml3GT1* (Figure 8 [Figure 8: see original paper]). The results showed that *Ml3GT1* expression decreased initially then increased during flower opening, reaching its peak at the full-bloom stage (S5). Tissue-specific expression analysis revealed that *Ml3GT1* expression was highest in flowers, with low expression in old and young leaves, and almost no expression in roots and stems. Expression levels at all flower developmental stages were higher than in other tissues.

Discussion and Conclusion

This study cloned the *Ml3GT1* gene from *M. liliflora* 'Hongyuanbao' . Bioinformatics analysis revealed that the Ml3GT1 sequence contains the plant secondary product glycosyltransferase signal sequence (PSPG box),

suggesting it may be involved in glycosylation modification of secondary metabolites. Based on their catalytic sites, glycosyltransferases can be classified into 3-O-glycosyltransferases (3GTs), 5-O-glycosyltransferases (5GTs), and 7-O-glycosyltransferases (7GTs) (王应丽等，2014). Phylogenetic clustering showed that Ml3GT1 has a close relationship with *Freesia hybrida* Fh3GT1, further indicating that it may function similarly to Fh3GT1 in flavonoid 3-O glycosylation modification (Sun et al., 2016).

Previous studies have demonstrated that 3GT exhibits organ-specific expression patterns (王毅等，2017). In *Paeonia delavayi* (王毅等，2017) and grape hyacinth (杜灵娟等，2017), 3GT genes are highly expressed in tissues with abundant anthocyanin accumulation, but show extremely low or no expression in tissues with little or no anthocyanins. Consistent with these findings, *Ml3GT1* was highly expressed in flowers but showed minimal or no expression in other vegetative tissues, suggesting its involvement in anthocyanin biosynthesis.

3GT participates in anthocyanin biosynthesis in freesia and grape hyacinth petals, and its expression is often positively correlated with anthocyanin accumulation (Sui et al., 2011; 梁沛雯等，2019). However, the expression pattern of *Ml3GT1* differs from that in freesia and grape hyacinth, showing a trend of initial decrease followed by increase during flower development. This discrepancy may be related to different anthocyanin glycoside types in different plant petals. Grape hyacinth petals contain only glucoside-type anthocyanins (梁沛雯等，2019), so anthocyanin accumulation is closely correlated with 3GT expression. In *M. liliflora* 'Hongyuanbao' petals, all four anthocyanins undergo rutinoside modification at the C3 position, with Pn3Ru being the most abundant (approximately 71% of total anthocyanins) (Wang et al., 2019). Since C3 rutinoside modification of anthocyanidins requires sequential catalysis by 3GT and 3RT (Yamazaki et al., 2002), the synthesis of rutinoside-type anthocyanins in *M. liliflora* 'Hongyuanbao' is regulated not only by *3GT* but also by *3RT* expression. These results indicate that *Ml3GT1* is involved in regulating anthocyanin biosynthesis in *M. liliflora* 'Hongyuanbao' , but is not the rate-limiting enzyme in this pathway.

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