Medicinal Plant Biology

ARTICLE

Open Access

https://doi.org/10.48130/mpb-0026-0021

Medicinal Plant Biology **2026**, 5: e026

Identification and functional characterization of betacyanin biosynthetic genes in the Tibetan medicinal plant *Mirabilis himalaica*

Jiang Liu[1#], Ting Sun[2#], Zhe Wang[3], Yajun Wang[2], Zhihua Liao[2], Fangyuan Zhang[2*] and Xiaozhong Lan[1*]

[1] *Key Laboratory of Xizang Medicine Resources Conservation and Utilization of Xizang Autonomous Region, Xizang Agriculture and Animal Husbandry University, Linzhi 860000, China*

[2] *Southwest Bio-resources R&D Key Laboratory of Sichuan Province, School of Life Sciences, Southwest University, Chongqing 400715, China*

[3] *College of Chemistry, Beijing University of Chemical Technology, Beijing 100029, China*

[#] Authors contributed equally: Jiang Liu, Ting Sun

[*] Correspondence: fyzhang@swu.edu.cn (Zhang F); lanxiaozhong@163.com (Lan X)

Abstract

Mirabilis himalaica (Edgew.) Heimerl (Nyctaginaceae) is a Tibetan medicinal plant of the Caryophyllales that accumulates red-violet betacyanins, yet its betacyanin biosynthetic genes and the 3,4-dihydroxyphenylalanine (DOPA) 4,5-dioxygenase (DODA) paralog responsible for betalamic acid formation betalamic acid formation have not been characterised. From a six-tissue *de novo* transcriptome, we identified and cloned six candidate genes (*MhADH*, *MhCYP76AD1*, *MhDODA1*, *MhDODA2*, *MhSDODA*, and *MhcDOPA5GT*) and analysed their phylogeny and tissue expression. *MhSDODA* denotes the candidate that groups with the functionally characterised *Mirabilis jalapa* MjDODA1. We reconstituted the pathway by transient co-expression in *Nicotiana benthamiana* and overexpressed *MhSDODA* in *M. himalaica*. Under the transient-expression conditions tested, *MhCYP76AD1* + *MhSDODA* was the smallest combination of introduced genes producing visible pigmentation (188 μg betanin equivalents·g^{-1} fresh weight determined by high-performance liquid chromatography). Betacyanin content rose to 351 μg·g^{-1} when *MhcDOPA5GT* was added and to 544 μg·g^{-1} when *MhADH* was added as well. Among the three DODA paralogs, only *MhSDODA* co-expressed with *MhCYP76AD1* produced strong pigmentation, whereas *MhDODA1* and *MhDODA2* yielded little to no pigment under the same conditions. Native overexpression of *MhSDODA* increased betacyanin content approximately threefold across three independent lines. These results provide the first functional characterisation of betacyanin biosynthetic genes in *M. himalaica* and identify *MhSDODA* as the major functional DODA paralog among those tested.

Keywords: *Mirabilis himalaica*, Betacyanin, DOPA 4,5-dioxygenase, Transient reconstitution, Tibetan medicinal plant

Citation: Liu J, Sun T, Wang Z, Wang Y, Liao Z, et al. 2026. Identification and functional characterization of betacyanin biosynthetic genes in the Tibetan medicinal plant *Mirabilis himalaica*. *Medicinal Plant Biology* 5: e026 https://doi.org/10.48130/mpb-0026-0021

Introduction

Mirabilis himalaica (Edgew.) Heimerl is a perennial herb of the family Nyctaginaceae distributed in the dry, hot river valleys of the Qinghai–Tibet Plateau, where it is used as a traditional Tibetan medicinal plant[1]. Its native habitat imposes high irradiance, strong ultraviolet (UV) radiation, and water limitation, conditions under which protective specialized metabolites are typically favoured. Previous work on *M. himalaica* has characterised its root rotenoids and UV-B-responsive specialised metabolism[2,3] and cloned the flavonoid-pathway entry genes *chalcone synthase* (*MhCHS*) and *chalcone isomerase* (*MhCHI*), demonstrating their function by *Agrobacterium rhizogenes*-mediated hairy root transformation[4]; more recently, a tissue-culture-independent (cut–dip–budding) transformation system has been developed for the species[5]. Its conspicuous red-violet betacyanin pigments, however, have not been characterised at the level of biosynthetic genes.

Pigmented species of the Caryophyllales generally accumulate betalains rather than anthocyanins, and, in nature, the two pigment classes are mutually exclusive[6,7]. Betalains comprise red-violet betacyanins and yellow betaxanthins. Betacyanins contribute to photoprotection, antioxidant capacity, and tolerance of abiotic stress[8,9]. They are therefore plausible contributors to environmental adaptation in high-altitude plants. Arogenate dehydrogenase (ADH) supplies the tyrosine from which betalain biosynthesis proceeds: the cytochrome P450 CYP76AD1 is bifunctional, hydroxylating tyrosine to L-DOPA and further oxidising L-3,4-dihydroxyphenylalanine (L-DOPA) to cyclo-DOPA[10,11]; 3,4-dihydroxyphenylalanine (DOPA) 4,5-dioxygenase (DODA) cleaves L-DOPA to yield 4,5-seco-DOPA, which spontaneously forms betalamic acid, the chromophore common to all betalains[12]; and cyclo-DOPA 5-O-glucosyltransferase (*cDOPA5GT*) glucosylates cyclo-DOPA en route to the stable betacyanin betanin[13,14].

A recurrent feature of this pathway is the functional divergence of DODA paralogs. Within a species, DODA duplicates typically comprise a lineage that enters the betalain pathway and one or more paralogs with little or no betalain activity[15], and high DODA activity has arisen repeatedly and convergently across the Caryophyllales rather than being predictable from its phylogenetic position[16]. Residue-based predictions of DODA's activity remain contested[17]. Determining which DODA paralog carries the pathway flux in a given species therefore requires direct functional testing rather than sequence inspection alone. Transient co-expression in *Nicotiana benthamiana* has become a standard route to reconstitute and dissect the betacyanin pathway, including as the visual RUBY reporter[18,19].

Although the betalain pathway is well described in *Beta vulgaris*, *Mirabilis jalapa*[20], and *Portulaca* species, the betacyanin

© The Author(s)

https://doi.org/10.48130/mpb-0026-0021

biosynthetic genes of *M. himalaica*, their tissue distribution, and the DODA paralog that is functionally dominant remain unknown. Here, we (i) generated a six-tissue *de novo* transcriptome and used it to identify and clone six candidate genes, (ii) analysed their phylogeny and their tissue-specific and light-responsive expression, and (iii) reconstituted the pathway by transient expression in *N. benthamiana* and overexpressed *MhSDODA* in the native species to identify an effective core combination in the transient expression system and compare the functional contributions of the three DODA candidates.

Materials and methods

Plant material and growth conditions

M. himalaica seeds were collected in Linzhi, Xizang, China (94°21′ E, 29°36′ N; ~3,150 m bove sea level) and identified by Prof. Xiaozhong Lan (Xizang Agriculture and Animal Husbandry University). Seeds were germinated in autoclaved humus/vermiculite/perlite (6:3:1, v/v/v), and seedlings were grown in a chamber at 25 ± 1 °C (day)/20 ± 1 °C (night), 16 h light/8 h dark, 160 μmol·m^{-2}·s^{-1}, and 40% ± 5% relative humidity. For the gene expression analysis, 10-day-old seedlings were divided into a control group (CK, growth under light, 160 μmol·m^{-2}·s^{-1}) and a high-light group (HL, 370 μmol·m^{-2}·s^{-1}). After 7 d of treatment, roots, stems and leaves were harvested separately from each group, with three biological replicates per tissue, and all other conditions were identical between the two groups. *N. benthamiana* was grown at 25 ± 1 °C, 16 h light/8 h dark, and 40% ± 5% relative humidity and used for infiltration at 6–8 weeks old.

De novo transcriptome assembly and candidate gene identification

A *de novo* transcriptome was assembled to mine betacyanin pathway genes in *M. himalaica*. Quality-filtered paired-end RNA sequencing (RNA-seq) reads from six vegetative tissues (taproot, secondary root, fibrous root, stem, old stem, and leaf) were further trimmed with fastp v1.3.6. This additional filtering retained 304,589,718 clean reads (152,294,859 read pairs; 44.6 Gb), with per-library Q30 proportions of 94.56%–96.11%. The trimmed reads were assembled with Trinity v2.15.2 (*in silico* normalisation to a maximum read coverage of 50; minimum k-mer coverage = 2). The final assembly contained 301,197 transcripts (N50: 1,553 bp; BUSCO v6.1.0, eudicots_odb10 completeness C: 91.7%). Candidate genes for each pathway step were identified by homology (BLAST) to characterised betalain biosynthesis enzymes of *B. vulgaris*, *M. jalapa*, and other Caryophyllales, and their full-length coding sequences (CDSs) were cloned (see section *Gene cloning*). The six candidates and their source transcripts are listed in Table 1.

Gene cloning

Full-length CDSs of the six candidates were amplified from mixed-tissue *M. himalaica* cDNA with PrimeSTAR Max DNA polymerase using gene-specific primers designed in Primer Premier 6.0 (Supplementary Table S1) and verified by Sanger sequencing. For functional assays, the coding sequences were assembled directly into the expression vector pEAQ-HT (see section *Transient expression in N. benthamiana*) by In-Fusion cloning, without an intermediate cloning vector. The coding regions were 1,203, 1,488, 822, 855, 816, and 1,503 bp for *MhADH*, *MhCYP76AD1*, *MhDODA1*, *MhDODA2*, *MhSDODA*, and *MhcDOPA5GT*, respectively. The six coding sequences were deposited in GenBank under accession numbers PZ746414 (*MhADH*), PZ746415 (*MhCYP76AD1*), PZ746416 (*MhDODA1*), PZ746417 (*MhDODA2*), PZ746418 (*MhSDODA*), and PZ746419 (*MhcDOPA5GT*).

Sequence and phylogenetic analysis

Deduced protein sequences of the six candidates and reference homologs from other Caryophyllales (Supplementary Table S2) were aligned with FAMSA. For each gene family (*ADH*, *CYP76AD*, *DODA*, and *cDOPA5GT*), a neighbor-joining tree was reconstructed from the p-distances with 1,000 bootstrap replicates and midpoint rooting (Biopython); bootstrap values ≥ 50% are shown. For an additional comparison of previously proposed activity-associated DODA residues, the 18 DODA protein sequences used in the phylogenetic analysis were aligned with MAFFT v7.526 using the automatic strategy. Alignment positions were mapped to ungapped BvDODA1 (HQ656027) and to the corresponding residue numbers in each *M. himalaica* paralog. This additional comparison was separate from the FAMSA alignment used for phylogenetic reconstruction.

RNA extraction and quantitative real-time polymerase chain reaction

Total RNA was extracted with TRIzol and reverse-transcribed (FastKing). Transcript levels were measured by quantitative real-time polymerase chain reaction (qRT-PCR) (FastFire SYBR Green premix; QuantStudio 5) with gene-specific primers (Supplementary Table S1), with *β-actin770* as the internal reference for *M. himalaica* and *NtACTIN* for *N. benthamiana*. For the tissue and light-response analysis, expression was calculated using the $2^{-\Delta\Delta C_T}$ method, where ΔCt = Ct(target gene) − Ct(internal reference gene), without normalization to a separate calibrator sample. Three biological replicates and three technical replicates were used for each measurement.

Transient expression in *N. benthamiana*

CDSs were inserted into pEAQ-HT and introduced into *Agrobacterium tumefaciens* GV3101 (pSoup19). Overnight cultures were resuspended to an optical density at 600 nm (OD_{600}) of 0.6 in an infiltration buffer (half-strength Murashige and Skoog medium, pH 5.7, 10 mM $MgCl_2$, 10 mM 2-[*N*-morpholino]ethanesulfonic acid, and

Table 1. The six candidate betacyanin pathway genes of *M. himalaica* and their *de novo* transcriptome transcripts.

Gene name	Functional annotation	Source transcript (Trinity)	CDS length (bp)
MhADH	Arogenate dehydrogenase	TRINITY_DN4957_c0_g1_i9	1,203
MhCYP76AD1	Cytochrome P450 CYP76AD1	TRINITY_DN163877_c0_g1_i1	1,488
MhDODA1	L-DOPA 4,5-dioxygenase 1	TRINITY_DN7768_c0_g1_i2	822
MhDODA2	L-DOPA 4,5-dioxygenase 2	TRINITY_DN1969_c0_g2_i1	855
MhSDODA	L-DOPA 4,5-dioxygenase paralog SDODA	TRINITY_DN6453_c5_g1_i2	816
MhcDOPA5GT	Cyclo-DOPA 5-*O*-glucosyltransferase	TRINITY_DN220813_c0_g1_i1	1,503

200 μM acetosyringone) and dark-adapted for 3 h. Cultures were infiltrated into *N. benthamiana* leaves individually or in binary, ternary, and quaternary combinations, mixed in equal volumes at a constant total OD_{600} of 0.6, with an empty vector (EV) strain as the infiltration control. The positive control was pYL1300H-CDGAeG, containing *BvADH*, *BvCYP76AD1S*, *BvDODA1S,* and *cDOPA5GT* together with *eGFP*, as described previously[5]. Visible pigmentation was recorded. Infiltrated regions were harvested at 2 d post-infiltration for pigment quantification, with the untransformed wild-type (WT) *N. benthamiana* leaf as the baseline.

Betacyanin extraction and high-performance liquid chromatography quantification

Betacyanins were extracted by grinding 1 g of fresh tissue (infiltrated *N. benthamiana* leaf or *M. himalaica* tissue) in liquid nitrogen, two rounds of extraction in 5 mL precooled double-distilled water (40 °C ultrasonication, 60 min; 8,000 rpm, 10 min), and combining the supernatants. Extracts were filtered (0.22 μm) and analysed by high-performance liquid chromatography (HPLC) (Shimadzu, photodiode array detector; Kromasil C18 column, 4.6 mm × 250 mm, 5 μm) with an isocratic mobile phase of methanol : 0.2% aqueous formic acid (15:85, v/v) at 0.8 mL·min^{-1}, 40 °C, 10 μL injection, and detection at 535 nm. Content was quantified against a betanin external standard (CAS 7659-95-2) from the calibration curve (y = 79,969x + 27,122, R^2 = 0.9865) and expressed as μg betanin equivalents·g^{-1} fresh weight (FW). These estimates do not establish the identity or absolute concentration of each individual betacyanin derivative. Three replicates were analysed per sample.

Overexpression of *MhSDODA* in *M. himalaica*

The *MhSDODA* coding sequence was cloned into pCAMBIA1300 (KpnI/BamHI) under the CaMV 35S promoter, together with a tdTomato fluorescent reporter, and introduced into *M. himalaica* using tissue-culture-independent cut–dip–budding transformation with *Agrobacterium rhizogenes* K599, as described previously[5]. Regenerated shoots were first screened for tdTomato fluorescence, and fluorescence-positive putative transformants were subsequently confirmed by genomic polymerase chain reaction (PCR). Thereafter, *MhSDODA* transcript levels were measured by qRT-PCR (as above), and leaf betacyanin content was quantified by HPLC (as above).

Statistical analysis

Data are presented as the mean ± standard deviation (SD) of three biological replicates. For each comparison against the relevant reference (HL versus CK within a tissue, each gene combination versus the control, and each overexpression line versus the WT), significance was assessed by Student's *t*-test and is indicated by asterisks (* $p < 0.05$, ** $p < 0.01$). Analyses were performed in SPSS 26.0, and figures were prepared in GraphPad Prism 8.0.

Results

De novo transcriptome mining yields six pathway candidates, including three DODA paralogs

From the *de novo* transcriptome, we identified and cloned six candidate genes spanning the betacyanin pathway: one arogenate dehydrogenase (*MhADH*), one CYP76AD-family P450 (*MhCYP76AD1*), three DODA-family genes (*MhDODA1*, *MhDODA2*, and *MhSDODA*), and one cyclo-DOPA 5-*O*-glucosyltransferase (*MhcDOPA5GT*) (Table 1; Fig. 1). In neighbor-joining trees, MhADH grouped with *Mirabilis jalapa* MjADH, MhCYP76AD1 with catalytically active CYP76AD proteins, and MhcDOPA5GT with characterised cyclo-DOPA glucosyltransferases including MjcDOPA5GT (Fig. 2a–c). Within the DODA family (Fig. 2d), MhSDODA grouped with *M. jalapa* MjDODA1 (bootstrap 100%), the functionally active DODA of that species, whereas MhDODA1 grouped with MjDODA2 (bootstrap 100%), the paralog reported to show no detectable betalain-producing activity under the tested conditions, and MhDODA2 grouped with MvDODA (bootstrap 99%)[15]. Because high DODA activity has arisen convergently and is not reliably predicted from clade position alone[16,17], these phylogenetic relationships were used to prioritise candidates for functional testing rather than as evidence of activity.

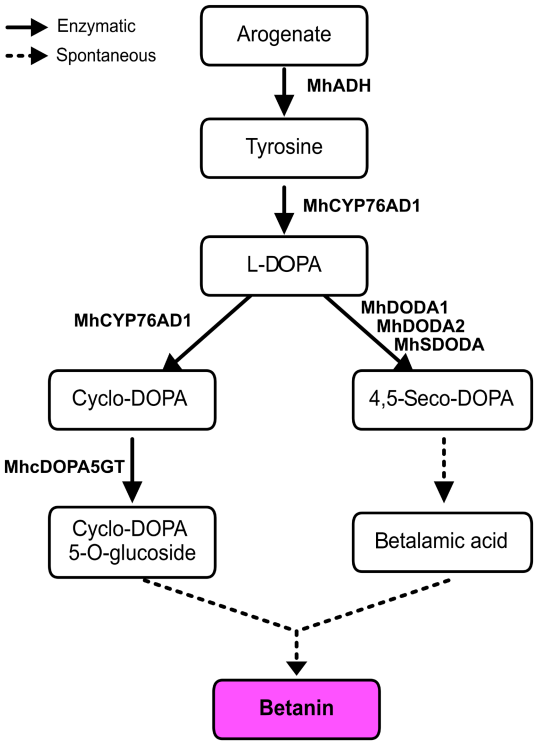

Fig. 1 Proposed betacyanin biosynthetic pathway in *M. himalaica* and the six candidate genes identified in this study (*MhADH*, *MhCYP76AD1*, *MhDODA1*, *MhDODA2*, *MhSDODA*, and *MhcDOPA5GT*), with the reactions each is proposed to catalyse.

High light induces pigmentation and broadly upregulates the six genes

High light (HL, 370 μmol·m^{-2}·s^{-1}, 7 d) induced a marked red-violet pigmentation of *M. himalaica* seedlings and leaves relative to CK (160 μmol·m^{-2}·s^{-1}), most visibly in leaves and shoots (Fig. 3a, b). Under CK, the six genes showed distinct tissue biases rather than a shared tissue-specific peak (Fig. 3c–h): *MhCYP76AD1* and *MhDODA2* were highest in roots, whereas *MhSDODA* and *MhcDOPA5GT* were lowest in the roots, *MhSDODA* also being highest in the leaves. HL significantly increased transcript levels of all six genes in the roots and stems; in leaves, it increased all but *MhCYP76AD1*, whose increase was not significant (Fig. 3c–h). The tissue distribution of *MhSDODA* transcripts qualitatively paralleled visible pigmentation, whereas the transcript abundance of *MhCYP76AD1* did not follow

https://doi.org/10.48130/mpb-0026-0021

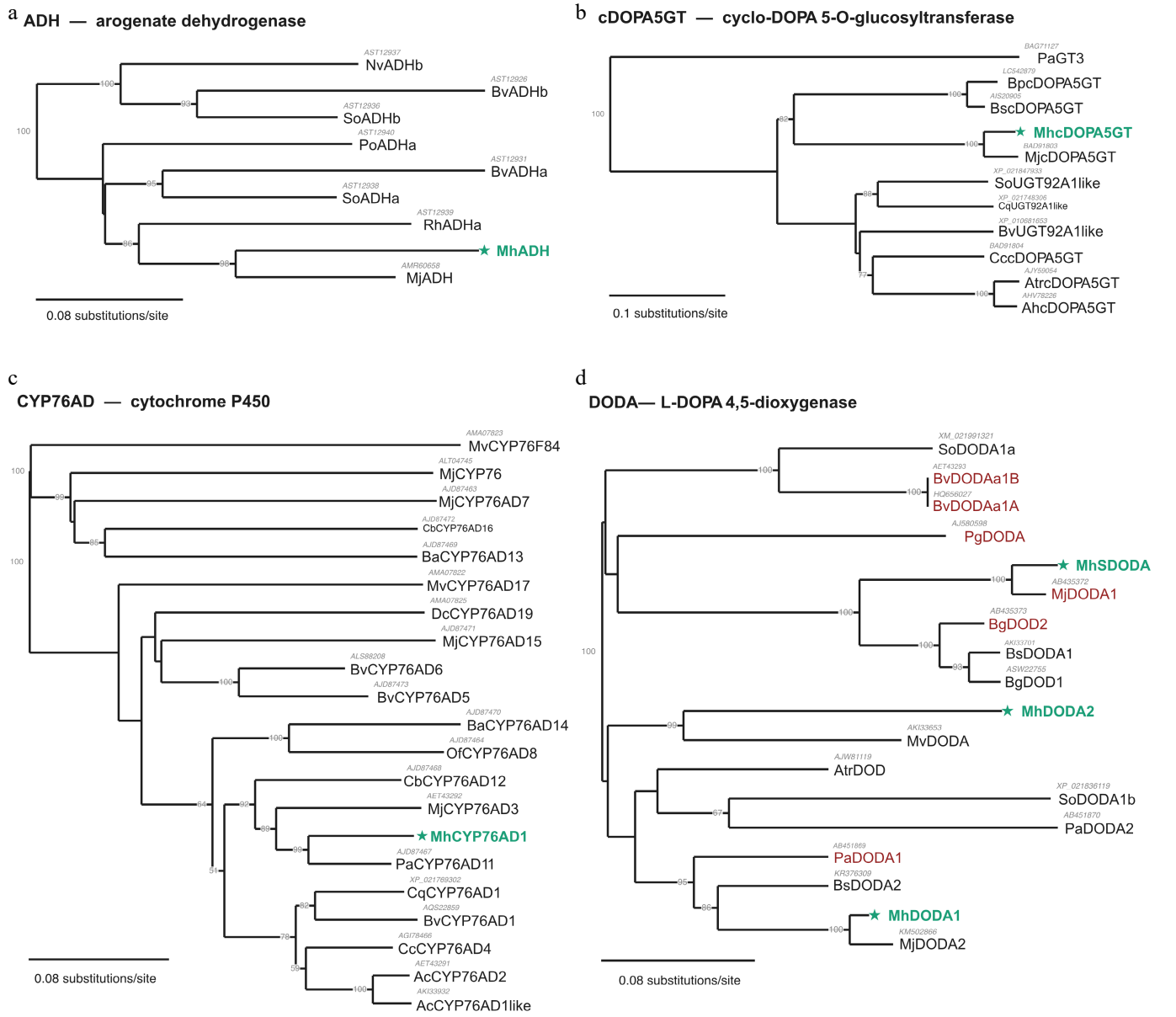

Fig. 2 Phylogenetic analysis of *M. himalaica* candidate proteins. Midpoint-rooted neighbor-joining trees of (a) ADH, (b) cDOPA5GT, (c) CYP76AD, and (d) DODA were constructed from FAMSA alignments using p-distances and 1,000 bootstrap replicates. Bootstrap values ≥ 50% are shown; scale bars indicate substitutions per site. Green labels and stars mark *M. himalaica* proteins. In (d), red labels indicate reference DODAs with demonstrated activity[12,15,21,22]. MjDODA2 and PaDODA2 showed no detectable activity[15,22]. Accessions are listed in Supplementary Table S2.

the same pattern. These observations describe an association between expression and pigmentation, rather than causal regulation of pigment accumulation by an individual gene.

Transient expression identifies an effective core combination and distinguishes DODA paralog phenotypes

To test the candidates functionally, we reconstituted the pathway in *N. benthamiana*. When each gene was infiltrated individually, none produced visible pigment beyond the empty-vector (EV) control (Supplementary Fig. S1a), so no single introduced gene was sufficient under these conditions. Among co-infiltrations of *MhCYP76AD1* with each DODA paralog on the same leaf, only *MhCYP76AD1* + *MhSDODA* produced clear pigmentation, whereas *MhCYP76AD1* + *MhDODA1* and *MhCYP76AD1* + *MhDODA2* yielded little to no visible pigment, similar to the EV control (Supplementary Fig. S1b). From the active *MhCYP76AD1* + *MhSDODA* module, adding *MhcDOPA5GT* and then *MhADH* progressively deepened the pigmentation (Fig. 4a, b). HPLC detected a pigment signal at 535 nm with a retention time matching the betanin standard (Fig. 4c), and betacyanin content increased from ~14 μg·g^{-1} FW in wild-type leaves to 188 μg·g^{-1} FW for *MhCYP76AD1* + *MhSDODA*, 351 μg·g^{-1} FW for the ternary combination (+ *MhcDOPA5GT*), and 544 μg·g^{-1} FW for the quaternary combination (+ *MhADH*) (Fig. 4d). These estimates are expressed as betanin equivalents. Under the conditions tested, *MhCYP76AD1* + *MhSDODA* was the smallest combination of introduced candidate genes that produced clear red-violet pigmentation. This result is specific to the *N. benthamiana* host background and does not establish that these two genes account for every biochemical step or are universally sufficient in other hosts. Among

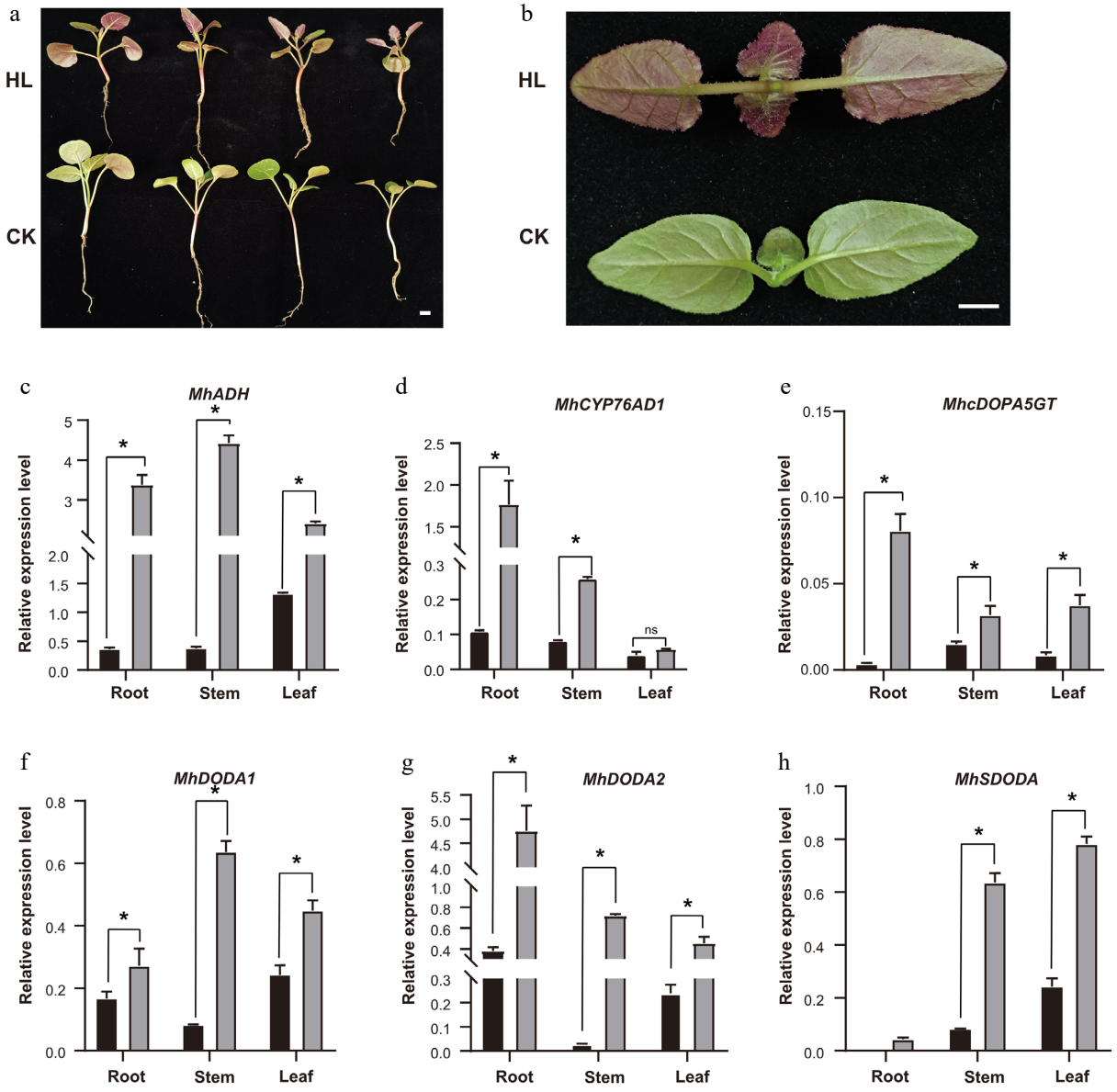

Fig. 3 High-light induction of betacyanin pigmentation and expression of the six candidate genes in *M. himalaica*. (a) Representative whole seedlings and (b) leaves grown under control light (CK, 160 μmol·m^{-2}·s^{-1}) or high light (HL, 370 μmol·m^{-2}·s^{-1}) for 7 d, showing HL-induced red-violet pigmentation (scale bars, 1 cm). Relative transcript levels under CK versus HL in roots, stems, and leaves for (c) *MhADH*, (d) *MhCYP76AD1*, (e) *MhcDOPA5GT*, (f) *MhDODA1*, (g) *MhDODA2*, and (h) *MhSDODA* (determined by qRT-PCR). Expression is shown as $2^{-\Delta\Delta CT}$, where ΔCt = Ct(target gene) − Ct(*β-actin770*). Values were normalised to the internal reference gene without a separate calibrator sample. Data are the mean ± SD (n = 3); * $p < 0.05$; ns, not significant (Student's *t*-test, HL vs CK).

the three DODA paralogs, *MhSDODA* produced the strongest pigmentation in this transient assay.

Native overexpression of *MhSDODA* increases betacyanin content

To test *MhSDODA*'s function in the native species, we generated *MhSDODA*-overexpressing lines of *M. himalaica* by the previously established cut–dip–budding transformation[5]. Following fluorescence prescreening, genomic PCR identified three independent transgenic lines (1#–3#; Fig. 5a). *MhSDODA* transcript levels in these lines were 50- to 115-fold higher than in the WT (Fig. 5b). Leaf betacyanin content, quantified by HPLC against the betanin standard (Fig. 5c), rose from ~14 μg·g^{-1} FW in the WT to 37–50 μg·g^{-1} in the three lines, an approximately threefold increase (all $p < 0.05$; Fig. 5d). Overexpression of *MhSDODA* is therefore sufficient to raise leaf betacyanin accumulation in *M. himalaica*.

Discussion

This study provides the first functional characterisation of betacyanin biosynthetic genes in *M. himalaica*. By combining *de novo* transcriptome mining, full-length cDNA cloning, heterologous reconstitution, and native overexpression, we defined a set of six pathway genes and identified *MhCYP76AD1* together with *MhSDODA* as an effective core combination in the *N. benthamiana* transient expression system, with *MhSDODA* as the major functional DODA paralog.

https://doi.org/10.48130/mpb-0026-0021

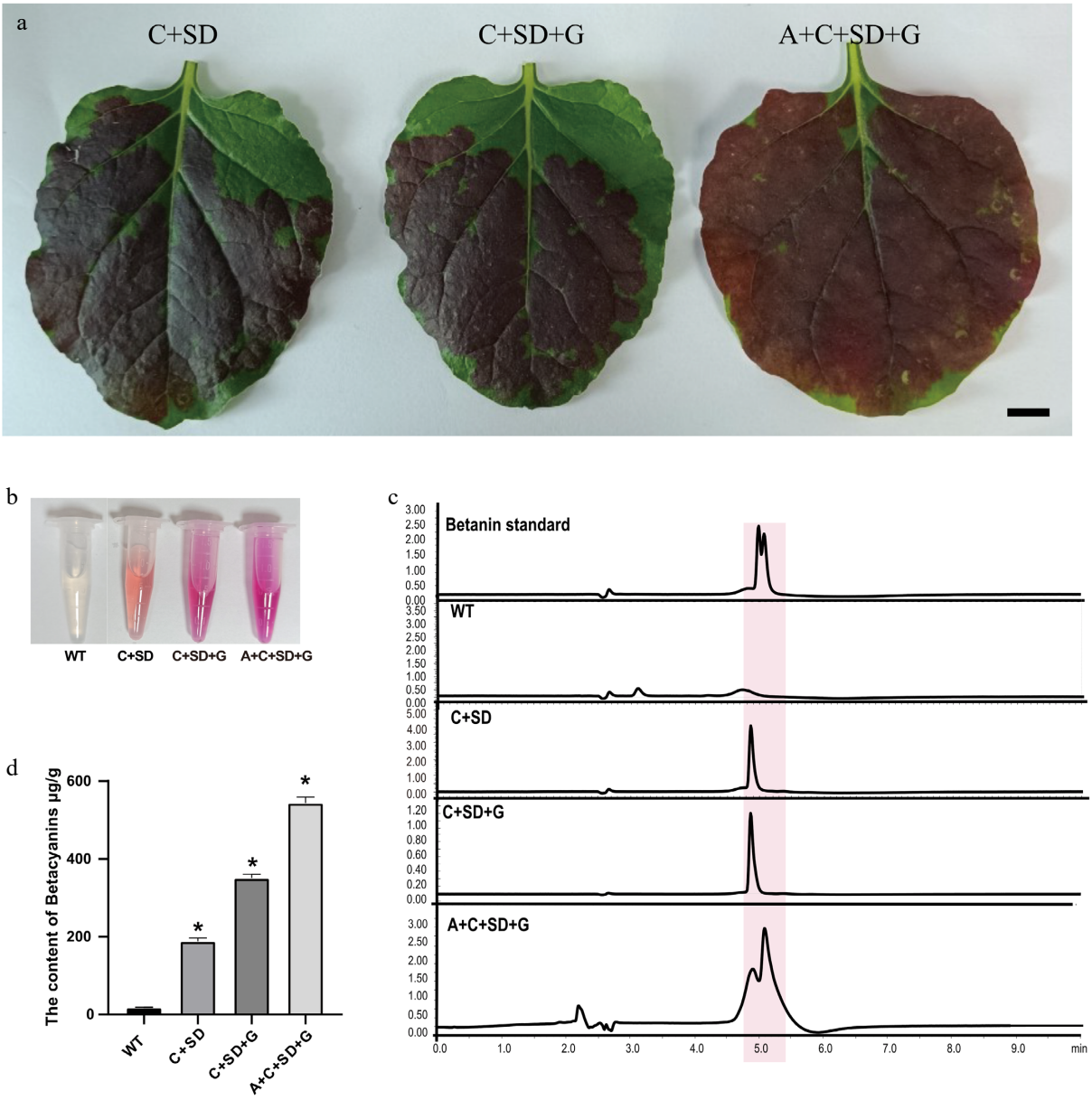

Fig. 4 Reconstitution of the betacyanin pathway in *N. benthamiana*. (a) Representative infiltrated leaves (scale bar, 1 cm) and (b) leaf extracts for *MhCYP76AD1* + *MhSDODA* (C + SD), the ternary combination + *MhcDOPA5GT* (C + SD + G), and the quaternary combination + *MhADH* (A + C + SD + G); WT, untransformed leaf. (c) HPLC chromatograms (535 nm) of a betanin standard, WT, and the three combinations; the shaded band marks the retention window corresponding to the betanin standard. (d) Betacyanin content determined by HPLC (μg betanin equivalents·g^{-1} FW; mean ± SD, $n = 3$; * $p < 0.05$ vs WT). C + SD, *MhCYP76AD1* + *MhSDODA*; C + SD + G, *MhCYP76AD1* + *MhSDODA* + *MhcDOPA5GT*; A + C + SD + G, *MhADH* + *MhCYP76AD1* + *MhSDODA* + *MhcDOPA5GT*. A, *MhADH*; C, *MhCYP76AD1*; SD, *MhSDODA*; G, *MhcDOPA5GT*.

A comparison with *M. jalapa* indicates that *M. himalaica* retains homologs of the same principal pathway components, including ADH, CYP76AD, DODA and cDOPA5GT. The paired phylogenetic relationships of MhSDODA with MjDODA1 and MhDODA1 with MjDODA2, together with the contrasting pigmentation phenotypes, are consistent with a shared ancestral functional divergence. However, the present analysis cannot determine its evolutionary timing or establish an independent origin in *M. himalaica*. Recovery of the additional candidate *MhDODA2* should not be interpreted as a species-specific gene gain without broader genomic sampling[15,20].

The comparison of the three DODA paralogs is the central functional result. Although MhDODA1, MhDODA2, and MhSDODA all belong to the DODA family, MhSDODA was markedly more effective than MhDODA1 or MhDODA2 at reconstituting betacyanin pigmentation when co-expressed with MhCYP76AD1. This outcome mirrors the phylogeny, in which MhSDODA groups with betalain-active *M. jalapa* MjDODA1 and MhDODA1 groups with MjDODA2, which showed no detectable betalain-producing activity in the reported assays[15]. It also fits the broader pattern: duplication of DODA is followed by functional divergence, and high activity is not reliably predicted from the clade position alone[16,17]. Our results reinforce the view that the functionally dominant DODA of a species must be identified experimentally, and they add *M. himalaica*, alongside its congener *M. jalapa*, to the species in which one DODA paralog is much more effective than the others[15]. Because this ranking was established in a heterologous transient assay and by

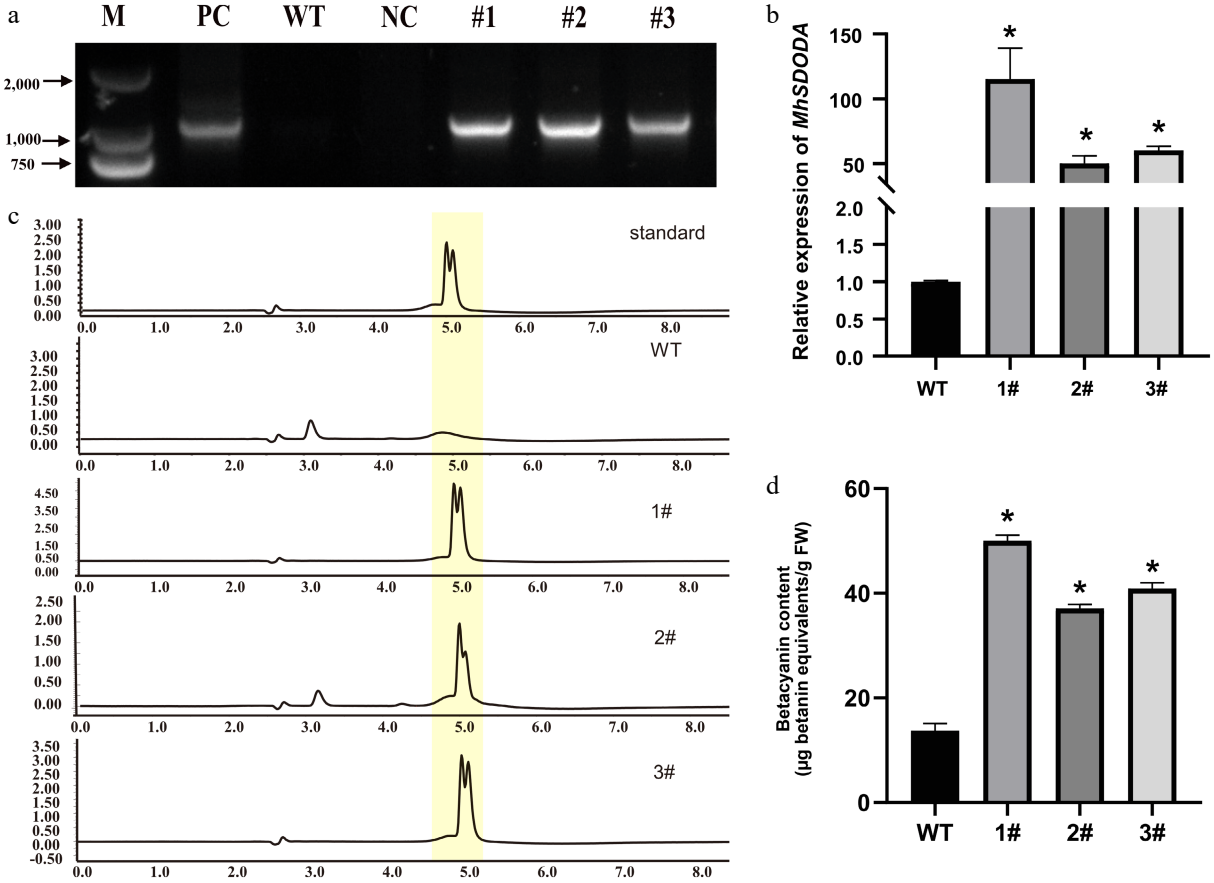

Fig. 5 Overexpression of *MhSDODA* in *M. himalaica*. (a) Genomic PCR confirmation of three independent overexpression lines (1#–3#). M, DNA size marker; PC, positive control (*MhSDODA* overexpression plasmid); WT, wild-type; NC, no-template control. (b) *MhSDODA* relative expression (qRT-PCR) in WT and lines 1#–3#. (c) HPLC chromatograms (535 nm) of a betanin standard, WT and the three lines; the shaded band indicates retention window corresponding to the betanin standard. (d) Betacyanin content (HPLC, μg betanin equivalents·g^{-1} FW) in WT and lines 1#–3#. Data are the mean ± SD (n = 3); * p < 0.05 (Student's t-test vs WT).

native overexpression rather than by loss of function, *MhSDODA* is best described as the major functional DODA paralog of the reconstituted pathway; establishing that it carries most of the endogenous flux will require knockdown or knockout of the paralogs in *M. himalaica*.

Comparison of the DODA candidate sequences identified differences at previously proposed activity-associated positions. At positions corresponding to BvDODA1 residues 81–83, 159, 185–186, and 210, MhSDODA has DDY, N145, DD and L196, respectively, matching MjDODA1. MhDODA1 has YGF, K157, NN and T208, whereas MhDODA2 has YGF, N166, RA and T218. These are separated alignment positions rather than a continuous peptide motif. These positions are candidates for functional tests, but sequence similarity is not proof of catalytic activity. The sufficiency of the substitutions proposed in earlier work[15] has been challenged by subsequent experiments[17]. Differences in transcript abundance, protein accumulation, or catalytic properties could also contribute to the transient expression phenotypes; no site-directed mutagenesis or purified enzyme assays of the Mh proteins were performed here.

The increase observed after adding *MhcDOPA5GT* is consistent with its proposed glucosyltransferase function, but the present assays do not distinguish increased product formation from altered product persistence. Similarly, the further increase associated with *MhADH* is compatible with an effect on tyrosine supply, but neither tyrosine abundance nor pathway flux was measured. These mechanisms remain interpretations rather than direct measurements in this experiment. Pigment accumulation also depends on the endogenous metabolism of the host; the result therefore does not define a universally minimal biosynthetic pathway.

The tissue expression data provide a descriptive comparison with visible pigmentation. *MhSDODA*'s transcript abundance was lowest in the roots, which showed little pigmentation, and was highest in the leaves, whereas *MhCYP76AD1* was highest in the roots. Without tissue-resolved metabolite measurements or loss-of-function experiments, this pattern does not demonstrate that *MhSDODA* limits pathway flux across tissues. The HL treatment increased transcripts' abundance broadly, although the increase in leaf *MhCYP76AD1* expression was not statistically significant.

The light-responsive pigmentation may be relevant to the exposed, dry, hot valley habitat of *M. himalaica*. In *Disphyma australe*, experimental induction of betacyanins reduced photoinhibition and accumulation of hydrogen peroxide under combined light and salinity stress[23]. By analogy, betacyanin accumulation could help moderate photo-oxidative stress in *M. himalaica*. This remains a hypothesis: our experiment tested increased visible light intensity, not UV exposure or drought, and did not measure photosystem performance, reactive oxygen species, or fitness. Pigment-dependent protection requires physiological measurements together with manipulation of betacyanin biosynthesis.

https://doi.org/10.48130/mpb-0026-0021

Several limitations should be stated. The product assignments are provisional: matching the retention time and detection at 535 nm relative to a betanin standard (Figs 4c and 5c) do not provide unambiguous structural identification, exclude co-eluting compounds, or identify individual betacyanin derivatives. Mass spectrometry analysis, preferably liquid chromatography–tandem mass spectrometry (LC–MS/MS) with appropriate reference standards, is required to confirm these assignments. Quantities are therefore reported as betanin equivalents under the analytical conditions used. The *de novo* assembly used one library each from six vegetative tissues and did not include flowers or seeds, so it supports the recovery of candidate sequences but not an exhaustive inventory of gene copies, isoforms, or tissue-restricted transcripts. The candidate genes are supported by transcriptomic, qRT-PCR, heterologous, and overexpression evidence rather than by loss of function in the native species. Future work should include product confirmation via mass spectrometry, *in vitro* comparison of the DODA paralogs, and loss-of-function analysis of *MhSDODA* in *M. himalaica*.

Conclusions

We cloned six genes of the betacyanin pathway of *M. himalaica* (*MhADH*, *MhCYP76AD1*, three DODA paralogs, and *MhcDOPA5GT*) and showed that *MhCYP76AD1* and *MhSDODA* form an effective core combination for visible pigmentation in the *N. benthamiana* transient expression system tested here. Among the three DODA paralogs, *MhSDODA* is by far the most effective, whereas *MhDODA1* and *MhDODA2* yielded little or no pigmentation under the same conditions, and native overexpression of *MhSDODA* increases betacyanin accumulation. This work provides the first functional description of the betacyanin biosynthetic genes of this Tibetan medicinal plant and a basis for further study of pigment regulation and metabolic engineering in the species.

Author contributions

The authors confirm their contributions to the paper as follows: study conception and design: Zhang F, Lan X; revising the paper: Sun T, Liao Z; gene cloning, plasmid construct, and genetic transformation: Sun T; transient expression and genetic transformation: Wang Z; molecular biology experiments and *N. benthamiana* transformation: Liu J; metabolite detection: Liu J, Wang Y; data analysis: Zhang F; manuscript writing: Zhang F, Lan X. All authors reviewed the results and approved the final version of the manuscript.

Data availability

Nucleotide sequences: GenBank PZ746414–PZ746419 (*MhADH*, *MhCYP76AD1*, *MhDODA1*, *MhDODA2*, *MhSDODA*, and *MhcDOPA5GT*). The quality-filtered paired-end RNA-seq reads from the six libraries used in the *de novo* assembly have been submitted to the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA1524643 (BioSample accessions SAMN62982685–SAMN62982690).

Acknowledgments

ZFYC Plan for Science and Technology Innovation Team (ZFYC202401001TD), the Key R&D and Transformation Special Program of Xizang Autonomous Region (XZ202601ZY0254), and the Fourth National Survey of Traditional Chinese Medicine Resources, Chinese or Tibetan Medicinal Resources Investigation in Xizang Autonomous Region (National Administration of Traditional Chinese Medicine 20200501–542301).

Conflicts of interest

The authors declare no competing interests.

Supplementary information accompanies this paper online at: https://doi.org/10.48130/mpb-0026-0021.

Dates

Received 7 August 2026; Revised 8 September 2026; Accepted 12 October 2026; Published online 30 October 2026

References

[1] Linghu L, Fan H, Hu Y, Zou Y, Yang P, et al. 2014. Mirabijalone E: a novel rotenoid from *Mirabilis himalaica* inhibited A549 cell growth *in vitro* and *in vivo*. *Journal of Ethnopharmacology* 155(1):326–333

[2] Gu L, Zhang ZY, Quan H, Li MJ, Zhao FY, et al. 2018. Integrated analysis of transcriptomic and metabolomic data reveals critical metabolic pathways involved in rotenoid biosynthesis in the medicinal plant *Mirabilis himalaica*. *Molecular Genetics and Genomics* 293(3):635–647

[3] Gu L, Zheng W, Li M, Quan H, Wang J, et al. 2018. Integrated analysis of transcriptomic and proteomics data reveals the induction effects of rotenoid biosynthesis of *Mirabilis himalaica* caused by UV-B radiation. *International Journal of Molecular Sciences* 19(11):3324

[4] Lan X, Quan H, Xia X, Yin W, Zheng W. 2016. Molecular cloning and transgenic characterization of the genes encoding chalcone synthase and chalcone isomerase from the Tibetan herbal plant *Mirabilis himalaica*. *Biotechnology and Applied Biochemistry* 63(3):419–426

[5] Sun T, Han X, Jiang Y, Li Q, Xu Y, et al. 2025. A method of genetic transformation of Xizang medical plants without tissue culture. *Medicinal Plant Biology* 4:e002

[6] Brockington SF, Yang Y, Gandia-Herrero F, Covshoff S, Hibberd JM, et al. 2015. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytologist* 207(4):1170–1180

[7] Timoneda A, Feng T, Sheehan H, Walker-Hale N, Pucker B, et al. 2019. The evolution of betalain biosynthesis in Caryophyllales. *New Phytologist* 224(1):71–85

[8] Jain G, Gould KS. 2015. Are betalain pigments the functional homologues of anthocyanins in plants? *Environmental and Experimental Botany* 119:48–53

[9] Polturak G, Aharoni A. 2018. "La vie en rose": biosynthesis, sources, and applications of betalain pigments. *Molecular Plant* 11(1):7–22

[10] Hatlestad GJ, Sunnadeniya RM, Akhavan NA, Gonzalez A, Goldman IL, et al. 2012. The beet R locus encodes a new cytochrome P450 required for red betalain production. *Nature Genetics* 44(7):816–820

[11] Sunnadeniya R, Bean A, Brown M, Akhavan N, Hatlestad G, et al. 2016. Tyrosine hydroxylation in betalain pigment biosynthesis is performed by cytochrome P450 enzymes in beets (*Beta vulgaris*). *PLoS One* 11(2):e0149417

[12] Christinet L, Burdet FX, Zaiko M, Hinz U, Zrÿd JP. 2004. Characterization and functional identification of a novel plant 4,5-extradiol dioxygenase involved in betalain pigment biosynthesis in *Portulaca grandiflora*. *Plant Physiology* 134(1):265–274

[13] Sasaki N, Adachi T, Koda T, Ozeki Y. 2004. Detection of UDP-glucose: cyclo-DOPA 5-O-glucosyltransferase activity in four o'clocks (*Mirabilis jalapa* L.). *FEBS Letters* 568(1–3):159–162

[14] Sasaki N, Wada K, Koda T, Kasahara K, Adachi T, et al. 2005. Isolation and characterization of cDNAs encoding an enzyme with glucosyltransferase activity for cyclo-DOPA from four o'clocks and feather cockscombs. *Plant & Cell Physiology* 46(4):666–670

[15] Bean A, Sunnadeniya R, Akhavan N, Campbell A, Brown M, et al. 2018. Gain-of-function mutations in beet *DODA2* identify key residues for betalain pigment evolution. *New Phytologist* 219(1):287−296

[16] Sheehan H, Feng T, Walker-Hale N, Lopez-Nieves S, Pucker B, et al. 2020. Evolution of L-DOPA 4,5-dioxygenase activity allows for recurrent specialisation to betalain pigmentation in Caryophyllales. *New Phytologist* 227(3):914−929

[17] Guerrero-Rubio MA, Walker-Hale N, Guo R, Sheehan H, Timoneda A, et al. 2023. Are seven amino acid substitutions sufficient to explain the evolution of high L-DOPA 4,5-dioxygenase activity leading to betalain pigmentation? Revisiting the gain-of-function mutants of Bean et al. (2018). *New Phytologist* 239(6):2265−2276

[18] Polturak G, Breitel D, Grossman N, Sarrion-Perdigones A, Weithorn E, et al. 2016. Elucidation of the first committed step in betalain biosynthesis enables the heterologous engineering of betalain pigments in plants. *New Phytologist* 210(1):269−283

[19] He Y, Zhang T, Sun H, Zhan H, Zhao Y. 2020. A reporter for noninvasively monitoring gene expression and plant transformation. *Horticulture Research* 7(1):152

[20] Polturak G, Heinig U, Grossman N, Battat M, Leshkowitz D, et al. 2018. Transcriptome and metabolic profiling provides insights into betalain biosynthesis and evolution in *Mirabilis jalapa*. *Molecular Plant* 11(1):189−204

[21] Sasaki N, Abe Y, Goda Y, Adachi T, Kasahara K, et al. 2009. Detection of DOPA 4,5-dioxygenase (DOD) activity using recombinant protein prepared from *Escherichia coli* cells harboring cDNA encoding DOD from *Mirabilis jalapa*. *Plant & Cell Physiology* 50(5):1012−1016

[22] Takahashi K, Yoshida K, Sakuta M. 2015. Comparative analysis of two DOPA dioxygenases from *Phytolacca americana*. *Natural Product Communications* 10(5):713−716

[23] Jain G, Schwinn KE, Gould KS. 2015. Betalain induction by L-DOPA application confers photoprotection to saline-exposed leaves of *Disphyma australe*. *New Phytologist* 207(4):1075−1083

Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit https://creativecommons.org/licenses/by/4.0/.