

Unstable anthocyanin pigmentation in *Streptocarpus* sect. *Saintpaulia* (African violet) is due to transcriptional selectivity of a single MYB gene

Daichi Kurata¹ , Tomohisa Tsuzaki¹, Fumi Tatsuzawa² , Kenta Shirasawa³ , Hideki Hirakawa³  and Munetaka Hosokawa^{1,4} 

¹Graduate School of Agriculture, Kindai University, 3327-204, Nakamachi, Nara, 631-8505, Japan; ²Faculty of Agriculture, Iwate University, 3-18-8, Ueda, Morioka, Iwate, 606-8550, Japan;

³Kazusa DNA Research Institute, 2-6-7, Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan; ⁴Agricultural Technology and Innovation Research Institute Kindai University (ATIRI), 3327-204, Hanamachi, Nara, 631-8505, Japan

Author for correspondence:

Munetaka Hosokawa

Email: mune@nara.kindai.ac.jp

Received: 24 January 2025

Accepted: 6 May 2025

New Phytologist (2025) 247: 1426–1444

doi: 10.1111/nph.70286

Key words: alternative promoter usage, anthocyanin, epigenetics, pattern formation, periclinal chimera.

Summary

- *Saintpaulia* (African violet) pigmentation is notoriously unstable and sometimes forms white stripes, particularly following passage through tissue culture. White-striped petals were thought to be due to periclinal chimeras, but we confirmed that white stripes result from epigenetic regulation rather than periclinal chimeras based on the flower color traits of plants obtained from tissue culture.
- Gene expression in several plant lines, anthocyanin quantification, bisulfite sequencing, and methylation analyses were used to demonstrate the presence of a single MYB gene responsible for pigment variation. We identified *SiMYB2* as the cause of variations in tissue color patterning, and that two RNAs were generated from *SiMYB2*.
- *SiMYB2-Long* was expressed in colored tissues, while *SiMYB2-Short* was expressed only in noncolored tissues. Functional analyses revealed that *SiMYB2-Long* is an anthocyanin biosynthesis activator and *SiMYB2-Short* is nonfunctional. Exon 3 of *SiMYB2* was generated by the insertion of a transposon-like sequence. A mutant lacking the element was obtained from cultivars with noncolored tissues. Anthocyanin content and *SiMYB2-Long* expression in the mutant were greatly increased compared to wild-type.
- Our results suggest that the white-striped petals of *Saintpaulia* are not formed by periclinal chimeras but through the transcriptional selectivity of epigenetically regulated *SiMYB2*.

Introduction

African violet (*Streptocarpus* sect. *Saintpaulia ionanthus* Wendl.) is an ornamental plant native to Kenya and Tanzania (Möller & Cronk, 1997). *Saintpaulia* accumulates anthocyanin in its petals, resulting in diverse color patterns such as stripe, picotee, splash, center eye, and spots. Notably, periclinal chimeras with bicolor striped petals are valuable due to the difficulties associated with their propagation (Eyerdom, 1981). *Saintpaulia* is typically propagated via adventitious shoots from tissue culture or leaf cuttings. However, in many cases adventitious shoots originate from multiple epidermal cells, leading to monochromatic petals rather than stripes when periclinal chimeras are propagated through adventitious buds (Naylor & Johnson, 1937; Broertjes & Keen, 1980; Ohki, 1994; Nabeshima *et al.*, 2017; Yang *et al.*, 2017). Thus, propagation of polychromatic periclinal chimera patterns through adventitious shoots is not possible.

Although some *Saintpaulia* cultivars have striped petals similar to those observed in periclinal chimeras, the plants regenerated

from them often exhibit varying color patterns. The petals of these cultivars exhibit a pinwheel-like pattern with a central-white region surrounded by a pigmented marginal region. Many white-striped cultivars have been bred and have become a popular *Saintpaulia* pigmentation pattern. These white-striped cultivars were thought to be periclinal chimeras (Lineberger & Druckenbrod, 1985; Kazemian *et al.*, 2019), but plants regenerated from these cultivars sometimes have white-striped petals, in addition to monochromatic petals, which makes it difficult to conclude that the white stripes are due to periclinal chimeras. Regenerated plants with white petals have unstable pigmentation and may also produce petals that are either white-striped or of a uniform color. Therefore, the mechanism of pattern formation in white-striped petals may be attributed to the instability of gene expression as a result of changes in epigenetic regulation (Yang *et al.*, 2017).

Ong-Abdullah *et al.* (2015) identified a gene (*EgDeficiens1*) that controls a trait that results in malformed fruit produced via tissue culture in African oil palm (*Elaeis guineensis*). In addition,

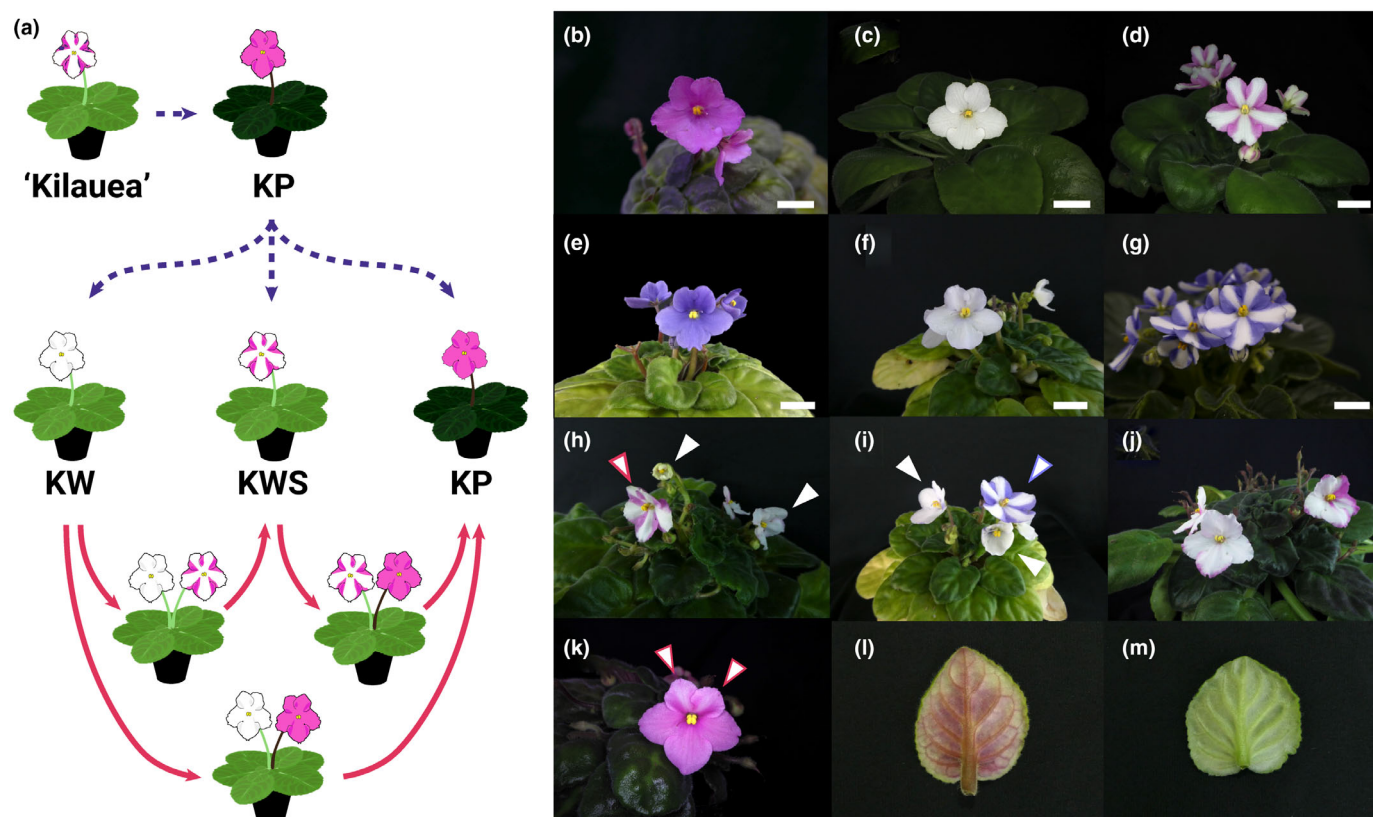


Fig. 1 (a) Variation in petals and leaf color of *Saintpaulia ionantha* cv 'Kilauea' that occurs via tissue culture (blue dashed arrow), and changes that occur during cultivation (red solid arrow). (b–g) Flowers of *Saintpaulia*. Bars, 1.5 cm. (b) 'Kilauea' Pink (KP). (c) 'Kilauea' White (KW). (d) 'Kilauea' White Stripe (KWS). (e) 'Stephanie' (ST), (f) 'Stephanie' White (SW), (g) 'Stephanie' White Stripe (SWS). (h, i) White and white striped flowers are present on a single plant simultaneously. A white arrowhead indicates a white flower and an arrowhead with colored margins indicates a white-striped flower. (j) Randomly pigmented KW petals. (k) Entirely pigmented KWS petals. (l) Leaf of KP. (m) Leaf of KW.

Tang *et al.* (2022) identified a gene (*CmMYB6*) responsible for floral color variation emerging from lateral buds in *Chrysanthemum* (*Chrysanthemum morifolium*). Because both studies successfully utilized cytosine methylation analysis to identify the epigenetic target, genes with different methylation levels among different traits are likely candidates as the causal genes. In this study, we have investigated the long-enigmatic phenomenon of pigment accumulation and pattern formation in *Saintpaulia*, based on the hypothesis that they result from epigenetic mutations. Using an integrated approach combining gene expression analysis and epigenomic profiling, we aimed to identify the gene or genes underlying this instability.

Materials and Methods

Plant materials and tissue culture

Saintpaulia ionantha Wendl. cv 'Kilauea' has white-striped petals with blue spots on a pink background at the margins. The plants with pink petals were obtained via tissue culture of this cultivar (Fig. 1a,b). Propagated plants with pink petals were labeled as 'Kilauea' Pink (KP). Leaves collected from KP plants were surface sterilized with 1% sodium hypochlorite supplemented with 10 μ l

Tween-20 for 15 min. The leaves were then rinsed with double-distilled water and cut into 1–2 cm² pieces using a razor. To obtain adventitious shoots, the pieces were cultured on full-strength MS medium (Murashige & Skoog, 1962) containing 2% sucrose (w/v), 0.3% gellan gum (w/v), 0.1 ppm 1-naphthalene acetic acid (NAA), and 0.5 ppm 6-benzyladenine (BA) for 4 wk. Regenerated plants with white petals were labeled 'Kilauea' White (KW) (Fig. 1c), and those with white-striped petals were labeled 'Kilauea' White Stripe (KWS) (Fig. 1d). The cultivar 'Stephanie' (ST) was used as a comparator line. Stephanie plants emerging from leaf cuttings with white petals ('Stephanie' White; SW) and with white-striped petals ('Stephanie' White Stripe; SWS) were labeled accordingly (Fig. 1e–g). Cultivars 'Akira', 'Tomoko', 'Georgia', 'Barbara', 'Manitoba', 'Taro', 'Saturn', 'Shimai' (SH), and mutants of SH (m-SH) were also included for anthocyanin assays and gene analyses (Supporting Information Fig. S1a–k). All plants were grown at 21 $\pm$ 2°C, 14 h : 10 h, light : dark, under fluorescent light (BioLux, 20 W, Hotalux, c. 50 μ mol cm⁻² s⁻¹). The plants were planted in 220 ml plastic pots filled with Metro-Mix 350J soil (HYPONeX, Osaka, Japan), and fertilizer (HYPONeX N-P-K = 6-10-5, HYPONeX) was applied once every 2 wk.

KW leaves were cultured on MS medium as above for 4 wk to obtain adventitious shoots. If KWS plants were chimeric with

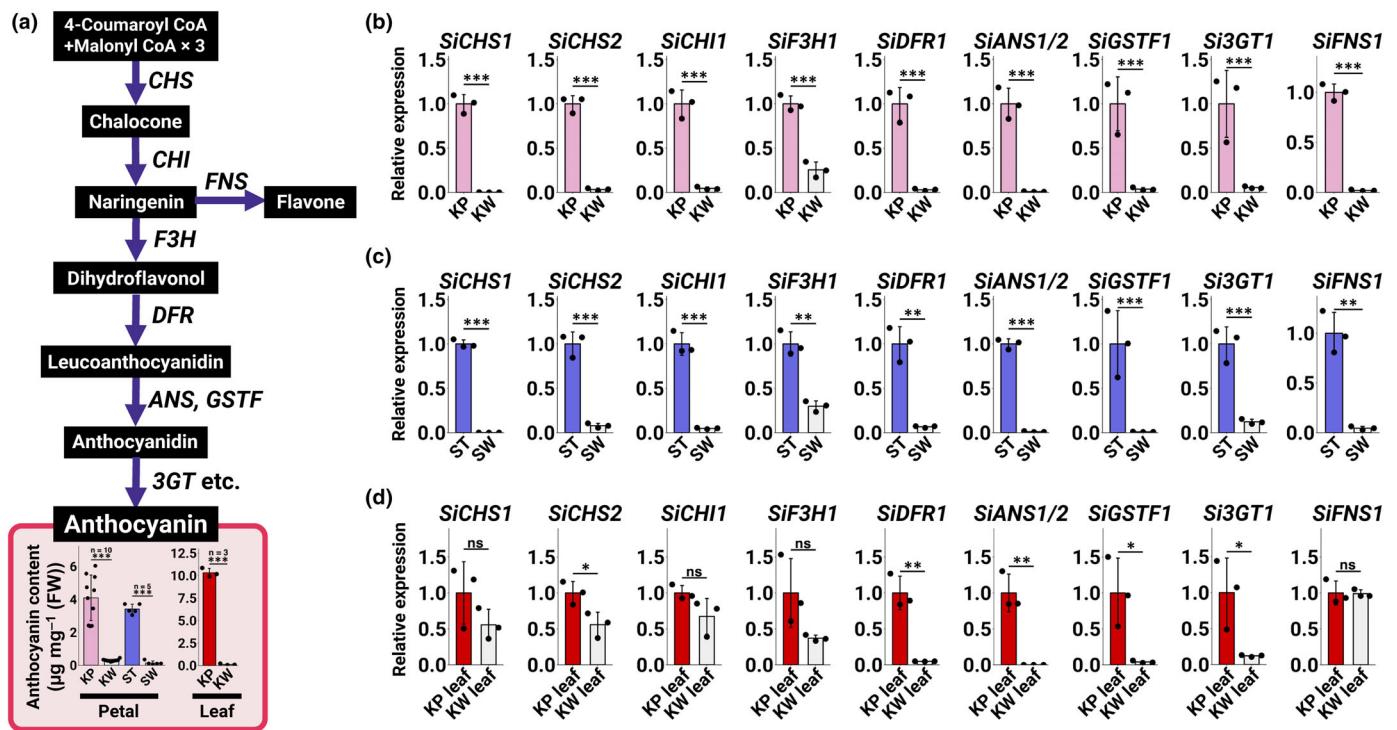


Fig. 2 (a) Anthocyanin biosynthesis pathway and anthocyanin content in *Saintpaulia ionanthes* petals and leaves. Asterisks indicate significant differences (Student's *t*-test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). ns indicates no significant difference. Error bars represent mean $\pm$ SD ($n = 10$ for KP and KW petals, $n = 5$ for ST and SW petals, and $n = 3$ for KP and KW leaves). (b–d) Quantitative reverse transcription PCR (qRT-PCR) analysis of anthocyanin biosynthesis-related genes in petals. Asterisks indicate significant differences (Student's *t*-test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). ns indicates no significant difference. Error bars represent mean $\pm$ SD ($n = 3$). (b) Expression in KP and KW petals. (c) Expression in ST and SW petals. (d) Expression in KP and KW leaves. 3GT, 3-Glucosyltransferase; ANS, Anthocyanidin synthase; CHI, Chalcone isomerase; CHS, Chalcone synthase; DFR, Dihydroflavonol 4-reductase; F3H, Flavanone 3-hydroxylase; FNS, Flavone synthase; GSTF, Glutathione S-transferase phi; KP, 'Kilauea' Pink; KW, 'Kilauea' White; ST, 'Stephanie'; SW, 'Stephanie' White.

pink pigmentation in the L1 layer and white (i.e. no pigmentation) in the L2 layer and below, all shoots originating from the L2 layer and below would be expected to display a white monochromatic phenotype. Tissues, including the epidermal cell layer (L1) of KWS petioles, were excised using a razor and forceps, leaving only the subepidermal cell layer(s) (L2 or below). The resulting explants were cultured in MS medium for 3 months to obtain adventitious shoots. All regenerated plants were grown as described above until they reached the flowering stage.

Flavonoid analysis

To identify which stage of the anthocyanin biosynthetic pathway was interrupted in KW petals, the anthocyanin and flavone compositions of KP and KW petals were measured by high-performance liquid chromatography (HPLC). Petals of KP and KW were collected 2–3 d after flower opening and frozen at -80°C . The frozen petals were quickly crushed using a mortar and pestle at room temperature to avoid volume increases due to absorption of additional moisture from the air. The powdered petals were then extracted with 10-fold weight of 50% acetic acid-methanol (acetic acid : methanol : distilled water, 1 : 4 : 5, v/v/v), and filtered through a 0.45- μm syringe filter (SLNY1345NB; Nippon Genetics, Tokyo, Japan). The filtered

solution was then subjected to HPLC for qualitative and quantitative determination of flavonoids. Analytical HPLC was performed using a C18 (4.6 $\times$ 250 mm) column (Waters, Milford, MA, USA) at 40°C with a flow rate of 1 ml min^{-1} , monitored by a photodiode array detector on an LC 10A system (Shimadzu Corp., Kyoto, Japan). Linear gradient elution was performed using an acetic acid-methanol mobile phase, from 20% to 85% solvent B (1.5% H_3PO_4 , 20% HOAc, 25% MeCN in H_2O) in solvent A (1.5% H_3PO_4 in H_2O) over 40 min, followed by a 5-min re-equilibration at 20% solvent B for flavonoids (530 and 350 nm). Pigments were identified to be pelargonidin-3-acetyl-rutinoside-5-glucoside (as anthocyanin 1; A1, Fig. 2s), peonidin-3-acetyl-rutinoside-5-glucoside (A2), cyanidin-3-acetyl-sambubioside (A3), apigenin-4'-glucuronide (flavone 1; F1), and luteolin-4'-glucuronide (F2), respectively, when compared with authentic samples obtained from *S. 'Thamires'* (Tatsuzawa *et al.*, 2012), *'Tomoko'* (Tatsuzawa *et al.*, 2015), and *'Concord'* (Nabeshima *et al.*, 2017) by HPLC analysis.

Absorbance at 530 nm (Abs_{530}) was measured using a Shimadzu UV-1800 spectrophotometer (Shimadzu Corp.) for the anthocyanin extracts derived from the petals of KP, KW, ST, and SW. A calibration curve was constructed from a fivefold dilution series of cyanidine-3-glucoside. Petal anthocyanin content was

quantified as cyanidine-3-glucoside equivalents using the calibration curve. The 5th to 7th expanded leaves of KP, KW, SH, and m-SH were collected to measure their anthocyanin content, then extracted and measured as described above.

Quantitative reverse transcription PCR (qRT-PCR) and reverse transcription PCR (RT-PCR)

The expression of anthocyanin biosynthesis-related genes was investigated by quantitative RT-PCR (qRT-PCR). Petal development was divided into four stages, designated as S1–S4 (Fig. S2). Petals covered by sepals and showing no coloration were designated as S1, the beginning of coloration as S2, the development of deeper coloration and just before flowering as S3, and fully colored petals around the time of flower opening as S4. Petal buds of KP, KW, ST, and SW buds of S2 and the youngest leaves of KP and KW were collected, frozen in liquid nitrogen, and total RNA was extracted using Sepasol[®]-RNA I Super G (Nacalai Tesque, Kyoto, Japan). First-strand cDNA was synthesized from total RNA using oligo dT primers and ReverTra Ace[®] (TOYOBO, Osaka, Japan). Reverse transcription reactions were performed at 42°C for 30 min and 99°C for 5 min. qRT-PCR was performed using cDNA and SYBR THUNDER BIRD[®] qPCR Mix (TOYOBO) by CFX Connect[™] Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). Reaction conditions were 95°C for 10 s and 60°C for 30 s for 40 cycles. Primers specific for anthocyanin biosynthesis-related genes were designed using Primer 3 Plus (Untergasser *et al.*, 2012) based on BLAST and annotation information of predicted gene sequences (Table S1). A calibration curve with an $R^2 \geq 0.98$ was constructed, and relative quantification was performed. *SiActin* (Sio_r2.0_p0205.1.g00590.1) was used as the housekeeping gene for normalization.

Petals of KWS and SWS were collected on the d of flower opening. Pigmented and nonpigmented regions of the petals at S4 were separated with a razor and immediately frozen in liquid nitrogen. RNA was then extracted, and qRT-PCR was performed as described above.

RT-PCR was performed to detect the expression of *SiMYB1* in petals and leaves. Petals (at S2) and leaf RNAs were reverse transcribed as described above. PCR reactions were performed using KOD One[®] PCR Master Mix (TOYOBO). Reactions were performed at 98°C for 10 s, 60°C for 5 s, and 68°C for 1 s for 35 cycles and amplified to a plateau. Primers were the same as those used for qRT-PCR (Table S1). *SiActin* was used as the internal control.

RNA-seq analysis

RNA-seq was performed to analyze the transcriptome of KP and KW. Petals were collected from KP and KW buds (S2) and frozen in liquid nitrogen. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Venlo, the Netherlands) according to the manufacturer's instructions. RNA-seq library preparation and sequencing were performed by the Rhelixa genomic and epigenomic research services company (Tokyo, Japan). RNA-seq

libraries were prepared using the NEBNext[®] Poly(A) mRNA Magnetic Isolation Module and the NEBNext[®] Ultra[™] II Directional RNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA). A 150-bp paired-end sequencing was performed on an Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA), yielding 4 Gb of reads (SRA accession: DRR462696–462701). Adaptor sequences and low-quality reads were removed computationally using TRIMMOMATIC v.0.39 (Bolger *et al.*, 2014).

The *Saintpaulia* SIO_r2.0 genome data (derived from the KP genome, with 37 682 predicted genes. BioProject PRJDB15710) was used as the reference sequence (Kurata *et al.*, 2024). Mapping to SIO_r2.0 was done using HISAT2 v.2.2.1 (Kim *et al.*, 2019). Read count data and transcripts per million (TPM) values were calculated by STRINGTIE v.2.2.1 (Pertea *et al.*, 2015). Differentially expressed genes (DEG) analysis using the TCC package (Sun *et al.*, 2013) and edgeR v.4.0.3 (Robinson *et al.*, 2010) was performed using read count data with the TMM method (FDR = 0.05). Mapping results obtained by RNA-seq analysis were visualized by pyGENOMETRACKS v.3.8 (Lopez-Delisle *et al.*, 2021). *De novo* assembly of transcript sequences was performed by TRINITY v.2.14.0 (Grabherr *et al.*, 2011) using RNA-seq data.

Phylogenetic tree construction

Molecular phylogenetic trees were constructed for MYB and bHLH transcription factors (TFs), which are important for regulating the expression of anthocyanin biosynthesis genes (ABGs). Amino acid sequences were filtered using PREQUAL v.1.02 (Whelan *et al.*, 2018) and aligned using MAFFT v.7.490 (Katoh & Standley, 2013) with default settings. Multiple aligned sequences were trimmed using TRIMAL v.1.4.1 (Capella-Gutierrez *et al.*, 2009) and used to generate a molecular phylogenetic tree by IQ-TREE2 v.2.2.0.3 (Minh *et al.*, 2020). Internal branch robustness was determined with Ultrafast bootstrapping (UFboot) and SH-aLRT 1000x each, and branches with ≥ 95 and $\geq 80\%$ confidence levels were considered robust (Minh *et al.*, 2013). The classification of MYB and bHLH TFs was based on Stracke *et al.* (2001) and Heim (2003), respectively. The accession IDs of the amino acid sequences are listed in Table S2.

Whole-genome bisulfite sequencing

Unstable pigmentation in petals and leaves may result from changes in the epigenomic state of genes in those organs. Therefore, whole-genome bisulfite sequencing (WGBS) was used to identify differences in the methylated cytosine status of anthocyanin biosynthesis-related genes. The 1st–4th expanded leaves were collected from KP and KW plants. Leaves were flash-frozen in liquid nitrogen, and DNA was extracted using QIAGEN Genomic-tip 100/G (Qiagen). DNA was treated with bisulfite using an EZ DNA Methylation-Gold Kit (Zymo Research, CA, USA). WGBS library preparation and sequencing were performed by Rhelixa, Inc. WGBS libraries were prepared from bisulfite-treated DNA using an Accel-NGS[®] Methyl-Seq DNA

Library Kit (Swift Biosciences, Ann Arbor, MI, USA), and 150-bp paired-end sequences were generated on an Illumina NovaSeq 6000 or NovaSeq X Plus (Illumina), yielding 30 Gb of reads. Adaptor sequences and low-quality reads were removed using TRIM GALORE v.0.6.7 (<https://github.com/FelixKrueger/TrimGalore>). The *Saintpaulia* SIO_r2.0 genome was used as the reference sequence (Kurata *et al.*, 2024). WGBS sequences were mapped to SIO_r2.0 using BWAMETH v.0.2.7 (arxiv.org/abs/1401.1129). Then, CG, CHG, and CHH methylation contexts were extracted using METHYLDACKEL v.0.6.1 (<https://github.com/dpryan79/MethylDackel>). Differentially methylated region analysis in the Sio_r2.0_p0007.1 contig was performed using METHYLKIT v.0.99.2 (Akalin *et al.*, 2012). Methylation levels were visualized by PYGENOMETRACKS v.3.8 (Lopez-Delisle *et al.*, 2021).

Methylation-sensitive McrBC digestion–PCR/quantitative PCR (McrBC-PCR/qPCR)

SiMYB2 methylation levels were compared between pigmented and nonpigmented tissues using McrBC (Takara, Shiga, Japan), a methylation-dependent restriction endonuclease. DNA was extracted from the first through fourth expanded leaves of KP and KW using NucleoBond® HMW DNA (Takara) and from the petals of KP and KW using a FavorPrep™ Plant Genomic DNA Extraction Mini Kit (Favorgen, Ping Tung, Taiwan), and digested according to the method described by Jaligot *et al.* (2014). Briefly, 100 ng of DNA was digested with 3 U McrBC at 37°C overnight. The DNA was then incubated at 65°C for 20 min to stop the reaction and used as a template for PCR. Genomic DNA without McrBC treatment was used as a control. The primers used are listed in Table S1. PCR and qPCR reactions were performed using KOD One® PCR Master Mix (TOYOBO) and KOD SYBR qPCR Mix (TOYOBO), respectively.

Cloning of *SiMYB2-Long*

First-strand cDNA was synthesized from RNA extracted from KP petals using oligo dT primers and ReverTra Ace® enzyme (TOYOBO). The full-length sequence of *SiMYB2-Long* was amplified using cDNA as a template with Blend Taq® enzyme (TOYOBO). Primer sequences are listed in Table S1. Amplicons were ligated to pTAC-1 using the DynaExpress TA PCR Cloning Kit (BioDynamics Laboratory, Tokyo, Japan), transformed into *Escherichia coli* DH5α competent cells, and colonies carrying the target fragments were selected. Plasmids were extracted from the selected colonies using FastGene™ PlasmidMini (NIPPON Genetics, Tokyo, Japan), and the DNA sequences were confirmed with Sanger sequencing by AZENTA (Tokyo, Japan).

CAGE-seq analysis

Cap analysis of gene expression (CAGE)-seq was performed to determine the transcription start site (TSS) of *SiMYB2*. Three petal buds were collected at S1 from KP and KW plant lines and frozen in liquid nitrogen. Total RNA was extracted using

Sepasol®-RNA I Super G (Nacalai Tesque). CAGE library preparation, sequencing, and mapping were performed by DNA-FORM (Kanagawa, Japan). cDNA was synthesized from total RNA using random primers. Ribose diols in the 5' cap structures of RNA molecules were oxidized and biotinylated. The biotinylated RNA/cDNA complexes were then isolated using streptavidin beads (cap-trapping method). Following RNA digestion with RNase ONE/H and adaptor ligation at both cDNA ends, double-stranded cDNA libraries (CAGE libraries) were generated and sequenced using 75-nt single-end reads (NextSeq 500, Illumina). The resulting reads were mapped to the *Saintpaulia* SIO_r2.0 genome using STAR v.2.7.9a (Dobin *et al.*, 2013).

Agroinfiltration assays

Nicotiana tabacum was grown for transient assay by sowing seeds in Kumiai Nippi Gardening Soil No. 1 (Nihon Hiryo, Tokyo, Japan) and germinating them under 20°C, 14 h : 10 h, light : dark conditions. The germinated seedlings were transferred to 220 ml plastic pots filled with the same soil and grown under the same conditions for 4–5 wk.

Four distinct vectors (*SiMYB1*, *SiMYB2-Long*, *SiMYB2-Short*, *SibHLH1*) were designed for transient assays. Vectors were constructed by artificial synthesis using VectorBuilder (VectorBuilder Japan, Kanagawa, Japan) or the in-fusion method. Each vector was designed to have the CDS of each TF driven by the CaMV35S promoter and the kanamycin marker gene NPT II. The VectorBuilder ID of each vector is listed: *SiMYB1* ID: VB210520-1064sfu, *SiMYB2-Long* ID: VB210520-1068fhv, *SibHLH1* ID: VB210520-1053zmv.

The pBI121 vector was digested with *Bam*HI (Takara) and *Sad* (Takara) to remove the *GUS* coding region. The CDS fragment of *SiMYB2-Short* was amplified using cDNA obtained by reverse transcription of petal RNA. CDS fragments were then amplified using primers for the in-fusion system (Table S2) (In-Fusion® HD Cloning Kit; Takara Bio USA, CA, USA) by combining pBI121 digested fragments and *SiMYB2-Short* amplicons. In-fusion cloning was performed according to the manufacturer's protocol.

Vector constructs were introduced into *Agrobacterium (Rhizobium radiobacter)* strain EHA105. Transformed *Agrobacterium* were grown on LB medium (0.5% yeast extract, 1% tryptone, 1% NaCl, 100 ppm kanamycin, 100 ppm rifampicin) for 2 d, and a single colony was selected for the transient assay. Vectors carrying *green fluorescent protein (GFP)* were introduced into *Agrobacterium* and used as negative controls.

Agroinfiltration was performed with a slight modification of the method of Outchkourov *et al.* (2014). Transformed *Agrobacterium* were suspended in liquid YEP medium amended with 50 ppm kanamycin and incubated overnight at 28°C with shaking. The culture medium was centrifuged (3500 g, 10 min), pellets were resuspended in infiltration buffer (10 mM MES-NaOH (pH 5.7), 10 mM MgCl₂, 150 μM acetosyringone), and adjusted to OD₆₀₀ = 0.4–0.6. The mixture was incubated for 2–3 h at room temperature. The mixture was then injected into

Table 1 Investigation of the flower color phenotype of regenerated plants obtained by tissue culture.

Explant	KP like		KW like		Random pigmentation		KWS like		Total
	No. of plants	Frequency (%)	No. of plants	Frequency (%)	No. of plants	Frequency (%)	No. of plants	Frequency (%)	
KP leaf	484	96.6	5	1	4	0.8	2	0.4	495
KW leaf	0	0	50	100	0	0	0	0	50
KWS L2 layer	1	2.9	33	94.2	0	0	1	2.9	35

KP, 'Kilauea' Pink; KW, 'Kilauea' White; KWS, 'Kilauea' White Stripe.

4- to 5-wk-old tobacco leaves. Leaves were photographed, collected, and stored frozen at -80°C 3 d post-infiltration. The experiment included three biological replicates of two leaves per tobacco plant, and the experiment was repeated in triplicate. Anthocyanin content was then quantified according to the procedure described above.

Leaf disks were collected from the injection site using a biopsy punch (BP-60F, 6 mm dia.; Kai Industries, Gifu, Japan), avoiding the puncture point of the syringe, and frozen in liquid nitrogen. qRT-PCR was performed according to the procedure described above using primers listed in Table S1. *NtActin* was used as the reference gene.

SiMYB2 genome structure

The selection of *SiMYB2* variants was investigated in several cultivars ('Akira', 'Tomoko', 'Georgia', 'Barbara', 'Manitoba', 'Taro', 'Saturn', 'Stephanie', KP) using RT-PCR. RNA extraction and RT-PCR were performed as described above. Reactions were performed at 98°C for 10 s, 60°C for 5 s, and 68°C for 1 s for 40 cycles and amplified to a plateau. Primers were the same as those used for qRT-PCR (Table S1). *SiActin* was used as the internal control.

The genomic structure of *SiMYB2* was analyzed through Amplicon-seq using multiple lines and cultivars ('Akira', 'Tomoko', 'Georgia', 'Barbara', 'Manitoba', KP). First, genomic DNA was extracted from the leaves of these cultivars using the FavorPrep™ Plant Genomic DNA Extraction Mini Kit (Favorgen). Primers were designed to amplify the ORF of *SiMYB2* (Table S1). PCR reactions were performed using KOD One® PCR Master Mix (TOYOBO) under the following conditions: 98°C for 10 s, and 68°C for 185 s for 40 cycles. Subsequently, a library was prepared with the Native Barcoding Kit 24 V14 (SQK-NBD114.24; Oxford Nanopore Technologies, Didcot, UK) and sequenced on a FLO-MIN114 flow cell (Oxford Nanopore Technologies).

Based on the results of Amplicon-seq, primers were designed to detect insertions or deletions (In/Dels) within the *SiMYB2* locus (Table S1). Genomic DNA was extracted from the leaves of ST, SA, SH, and m-SH in addition to the cultivars listed above using a FavorPrep™ Plant Genomic DNA Extraction Mini Kit (Favorgen). PCR reactions were performed using KOD One® PCR Master Mix (TOYOBO). Reactions were performed at 98°C for 10 s, 60°C for 10 s, and 68°C for 3 s for 40 cycles.

The expression levels of *SiMYB2-Long* and *SiMYB2-Short* in the leaves of SH and m-SH were investigated by qRT-PCR. RNA extraction and qRT-PCR were performed as described above.

Statistical analyses

All statistical processing was performed using R software (v.4.3.2, <https://www.R-project.org/>). Student's *t*-test was performed using the STATS package, and Dunnett's range test was performed using the MULTCOMP package.

Results

Variations in anthocyanin accumulation as a result of tissue culture

A total of 495 regenerated plants were obtained from 'Kilauea' Pink (KP) leaf adventitious shoots, of which 96.6% (484/495) produced pink petals and reddish leaves, the same phenotype as KP (Table 1). 1.0% (5/495) did not accumulate anthocyanin and therefore had white petals and green leaves ('Kilauea' White; KW). 0.8% (4/495) had randomly-colored petals, and 0.4% (2/495) had white-striped petals ('Kilauea' White Stripe; KWS) (Fig. 1; Table 1). As the five KW plants were grown, two of them began to produce white striped petals like KWS, and the remaining three had randomly pigmented petals (Fig. 1h,j). In addition, five KWs and two KWSs eventually grew petals and leaves like the parental KP (Fig. 1k). Fifty plants were regenerated from KW leaves, all of which had white petals and green leaves (50/50, Table 1). As these plants grew, some began producing striped or fully pigmented petals like those of KWS or KP, respectively.

If KWS is a periclinal chimera, plants derived from subepidermal tissue (L2 layer or below) should have only white petals and green leaves. On the other hand, if KWS is the result of epigenetic regulation, petal coloration may vary among individual plants. Based on this assumption, 35 plants were generated from subepidermal tissue. Petal coloration was distributed as follows: one plant (1/35) had pink petals, one (1/35) had white striped petals, and all the remaining plants had white petals (33/35, Table 1). As the plants grew, white petals began to produce pink pigment, white stripes, or randomly-colored petals. The other cultivar used in this study, 'Stephanie' (ST), was also as unstable as 'Kilauea' for pigment patterning. For example, 'Stephanie' White (SW), arising as a mutant from ST, also became striped

(‘Stephanie’ White Stripe; SWS) and eventually had the same petal coloring as ST (Fig. 1i).

Loss of anthocyanin biosynthesis in white petals and green leaves

Pelargonidin-3-acetyl-rutinoside-5-glucoside (anthocyanin 1; A1, Fig. S3a) and peonidin-3-acetyl-rutinoside-5-glucoside (A2, Fig. S3a) were readily detected in KP petals, but were at low levels in KW petals (Fig. S3b). While apigenin-4'-glucuronide (flavone 1; F1, Fig. S3e) was identified as a flavone in KP petals, it was hardly detected in KW petals (Fig. S3f). In addition, cyanidin-3-acetyl-sambubioside (A3, Fig. S3c) and luteolin-4'-glucuronide (F2, Fig. S3g) were detected in KP leaves, but were not detected in KW leaves (Fig. S3d,h). Thus, the functions of *CHALCONE SYNTHASE* (*CHS*), *CHALCONE ISOMERASE* (*CHI*), *FLAVONE SYNTHASE* (*FNS*), or all of them, may be disturbed in KW petals and leaves (Fig. 2a). Furthermore, total anthocyanin content was significantly lower in KW than KP petals and in SW than ST petals (Fig. 2a), and the anthocyanin content of KW leaves was lower than that of KP leaves (Fig. 2a).

To investigate at which stage anthocyanin metabolism is interrupted, the expression levels of ABGs, *CHS* (*SiCHS1/2*), *CHI* (*SiCHI1*), and *FNS* (*SiFNS1*) in KP and KW petals were compared by qRT-PCR. The expression of *SiCHS1/2*, *SiCHI1*, and *SiFNS1* was significantly lower in KW petals than in KP petals (Fig. 2b). The expression levels of *SiF3H1*, *SiDFR1*, *SiANS1/2*, *SiGSTF1*, and *Si3GT1* were also significantly lower in KW petals than in KP petals (Fig. 2b). Consistent with the qRT-PCR results, RNA-seq showed that the expression levels of several ABGs (*SiCHS1/2*, *SiCHI1*, *SiF3H1*, *SiDFR1*, *SiANS1/2*, *SiGSTF1*, *Si3GT1/2*, *SiMATE1/2/3*, etc.) were reduced in KW petals relative to KP petals (Table 2). SW petals, which have similar characteristics to KW, have several ABG expression levels significantly lower than in ST petals (Fig. 2c). In addition, the expression of several ABGs (*SiCHS1/2*, *SiCHI1*, *SiF3H1*, *SiDFR1*, and *SiANS1/2*) was lower in the white regions of KWS and SWS petals than in their pigmented regions (Fig. S4a,b). There were no significant differences in the expression levels of *SiCHS1*, *SiCHI1*, *SiF3H1*, and *SiFNS1* between KP and KW leaves (Fig. 2d), but the expression levels of *SiCHS2*, *SiDFR1*, *SiANS1/2*, *SiGSTF1*, and *Si3GT1* were significantly lower in KW leaves than in KP leaves (Fig. 2d). Based on qRT-PCR and RNA-seq results, the low anthocyanin accumulation levels in white petals and green leaves are likely due to the downregulation of several ABGs.

Downregulation of anthocyanin biosynthesis-related TFs

The reduced expression of multiple ABGs in anthocyanin-deficient tissues suggests that TFs may be responsible for this downregulation, rather than transcript turnover or translation deficiencies. Based on the RNA-seq analysis of KP and KW petals, we focused on the key regulators of anthocyanin biosynthesis: that is *MYB*, *bHLH*, and *WDR* (Ramsay & Glover, 2005; Xu *et al.*, 2015).

DEG analysis identified *SiMYB2* and *SibHLH2* as DEGs in petals, both of which are predicted to be TFs related to anthocyanin biosynthesis according to the BLAST results (Table 2). Expression levels of *SibHLH2* were significantly lower in KW than in KP petals, whereas *SiMYB2* expression levels were significantly higher in KW than in KP petals (Table 2). Consistent with the RNA-seq results, qRT-PCR indicated that *SibHLH2* expression was significantly lower in KW than in KP petals, and *SiMYB2* expression was significantly higher in KW than in KP petals (Fig. 3a). This expression pattern was also observed in a comparison between ST and SW petals (Fig. 3b). *SiMYB2* expression levels were also significantly higher in KW than in KP leaves, whereas *SibHLH2* expression levels were significantly lower in KW than in KP leaves (Fig. 3c), suggesting that either *SiMYB2*, *SibHLH2*, or both are involved in the unstable pigmentation of *Saintpaulia* petals and leaves.

Although *SiMYB1* and *SibHLH1* were not DEGs, these TFs were anthocyanin biosynthesis-related, ancillary to *SiMYB2* and *SibHLH2*. The qRT-PCR assay revealed that *SibHLH1* had equivalent expression levels in pigmented and nonpigmented tissues (Fig. 3a–c), and that *SiMYB1* expression was also about the same in KP and KW petals (Fig. 3a). In the nonpigmented mutant SW, which arose from ST, *SiMYB1* expression was significantly lower in SW than in ST (Fig. 3b). However, *SiMYB1* expression was nearly undetectable in the leaves of either KP or KW (Fig. 3d). These results suggest that *SiMYB1* and *SibHLH1* are unlikely to be the main factors in unstable anthocyanin accumulation.

A molecular phylogenetic tree was constructed to model the evolutionary history of MYB and bHLH TFs in *Saintpaulia*. The tree indicates that *SiMYB1* groups within the SG6 clade (Stracke *et al.*, 2001), which is known to promote anthocyanin biosynthesis (Fig. 3e). Although *SiMYB2* was highly expressed in nonpigmented tissues and was hypothesized to belong to a clade associated with anthocyanin biosynthesis repression, it was instead clustered within the SG6 clade (Fig. 3e). Both *SibHLH1* (IIIb) and *SibHLH2* (IIIa) belonged to clades known to promote anthocyanin biosynthesis (Heim, 2003) (Fig. 3f).

Hypermethylation of *SiMYB2* in anthocyanin-deficient tissues

From the expression analysis, functional annotation, and phylogenetic tree results, *SiMYB2* and *SibHLH2* were identified as candidate genes involved in the unstable pigmentation of *Saintpaulia* petals and leaves. The hypothesis that loss and reversion of anthocyanin accumulation in flowers and leaves was associated with changes in the epigenome led us to consider that the epigenetic state of either *SiMYB2* or *SibHLH2* may have undergone modification. Therefore, WGBS (whole genome bisulfite sequence) was conducted to examine the methylation levels of these candidate genes in the leaves of KP and KW. Differences in methylation levels were observed in the second exon of *SiMYB2* (Fig. 4a). In particular, methylation levels in the CHG and CHH contexts were higher in KW than in KP (Fig. 4a; Table S3). Additionally, CHH context methylation was higher in KW in

Table 2 Summary of DEG analysis for anthocyanin biosynthesis-related genes.

	Gene ID	Gene name	q-Value	Rank	BLAST top hit vs UniProtKB		Average of TPM (n = 3)	
					Protein name	Organism	KP	KW
DEG (q-value < 0.05)	Sio_r2.0_p0372.1.g00610.1	SiMATE1	2.60E-03	1	Protein DETOXIFICATION	<i>Sesamum indicum</i>	149.58612	31.852703
	Sio_r2.0_p0376.1.g00340.1	SiMATE2	2.60E-03	2	Protein DETOXIFICATION	<i>Sesamum indicum</i>	100.73756	28.77183
	Sio_r2.0_p0045.1.g02800.1	SiF3H1	2.60E-03	3	Flavone 3-hydroxylase	[Saintpaulia] hybrid cultivar	206.91722	41.695614
	Sio_r2.0_p0290.1.g00150.1	SiAT1	3.06E-03	4	Malonyl-coenzyme A: anthocyanin 3-O-glucoside-6'-O-malonyltransferase-like Chalcone synthase	<i>Doroceras hygrometricum</i>	12.371132	0.1122113
	Sio_r2.0_p0222.1.g00520.1	SiCHS1	3.06E-03	5	Chalcone synthase	<i>Sesamum indicum</i>	537.37294	6.2335113
	Sio_r2.0_p0085.1.g02260.1	SiMATE3	5.98E-03	7	Protein DETOXIFICATION	<i>Doroceras hygrometricum</i>	62.14773	4.992234
	Sio_r2.0_p0007.1.g08820.1	SiMYB2	6.93E-03	8	R2R3-MYB1	<i>Primulina swinglei</i>	11.315021	287.37764
	Sio_r2.0_p0408.1.g00410.1	SiANS1	1.23E-02	12	Leucoanthocyanidin dioxygenase-like	<i>Sesamum indicum</i>	155.04968	11.442141
	Sio_r2.0_p0007.1.g03330.1	SiBHLH2	1.45E-02	18	Transcription factor BHLH42-like	<i>Punica granatum</i>	26.254791	3.056827
	Sio_r2.0_p0041.1.g00830.1	SiANS2	1.45E-02	20	Leucoanthocyanidin dioxygenase-like	<i>Sesamum indicum</i>	181.95548	11.581487
	Sio_r2.0_p0021.1.g02260.1	Si3GT1	1.60E-02	27	Glycosyltransferase	<i>Sesamum indicum</i>	385.96082	68.025324
	Sio_r2.0_p0004.1.g00630.1	SiDFR1	1.67E-02	40	Dihydroflavonone reductase	<i>Erythranthe lewisii</i>	312.94233	17.74549
	Sio_r2.0_p0044.1.g01100.1	SiCHI1	1.77E-02	50	Chalcone-flavonone isomerase family protein	<i>Primulina swinglei</i>	1359.7302	271.05054
	Sio_r2.0_p0002.1.g09030.1	SiFNS1	2.35E-02	72	Flavone synthase II	<i>Doroceras hygrometricum</i>	285.47231	32.438459
	Sio_r2.0_p0008.1.g04160.1	Si3GT2	2.95E-02	100	Glycosyltransferase	<i>Doroceras hygrometricum</i>	83.702533	33.027269
	Sio_r2.0_p0005.1.g02330.1	SiCHS2	3.15E-02	113	Chalcone synthase	<i>Sesamum indicum</i>	1761.9983	358.21016
	Sio_r2.0_p0043.1.g03850.1	SiGSTF1	3.90E-02	165	Glutathione transferase	<i>Sesamum indicum</i>	1405.1501	257.18032
	Sio_r2.0_p0002.1.g10740.1	SiF3'5'H	4.06E-02	172	Flavonoid 3',5'-hydroxylase	[Saintpaulia] hybrid cultivar	106.33262	5.3640643

Color shades indicate differential expression between KP and KW: red represents upregulation and blue represents downregulation.

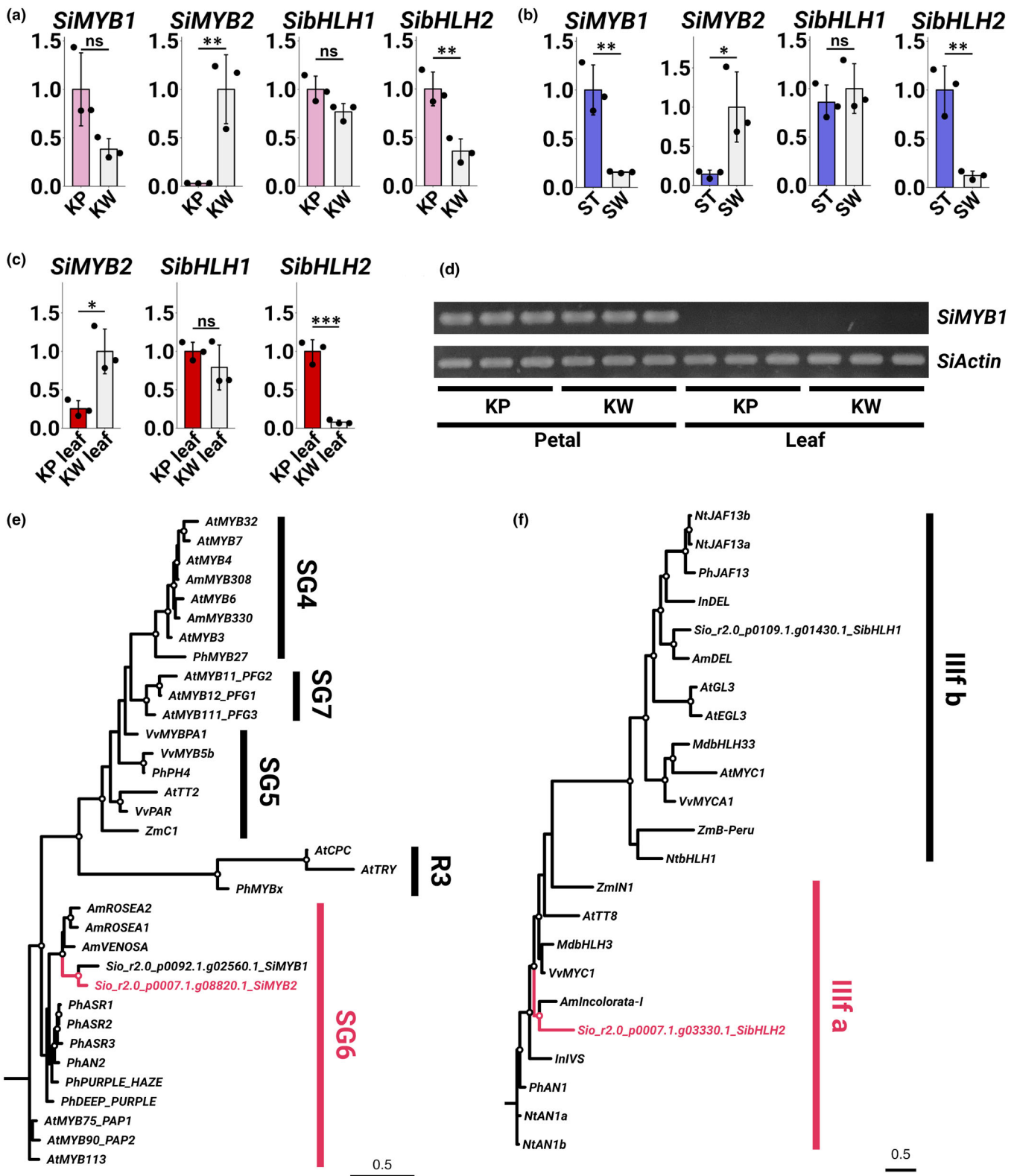


Fig. 3 (a–c) Quantitative reverse transcription PCR (qRT-PCR) analysis of anthocyanin biosynthesis-related transcription factors. Asterisks indicate significant differences (Student's *t*-test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). Error bars represent mean $\pm$ SD ($n = 3$). (a) Expression levels in KP and KW petals. (b) Expression levels in ST and SW petals. (c) Expression levels in KP and KW leaves. (d) Reverse transcription PCR (RT-PCR) results of *SiMYB1* and *SiActin* in petals and leaves of KP and KW ($n = 3$). The amplification reaction was carried out for 35 cycles (plateau). (e, f) Molecular phylogenetic tree of MYB TF (e) and bHLH TF (f) amino acid sequences. Robust branches are indicated by open circles. KP, 'Kilauea' Pink; KW, 'Kilauea' White; ST, 'Stephanie'; SW, 'Stephanie' White.

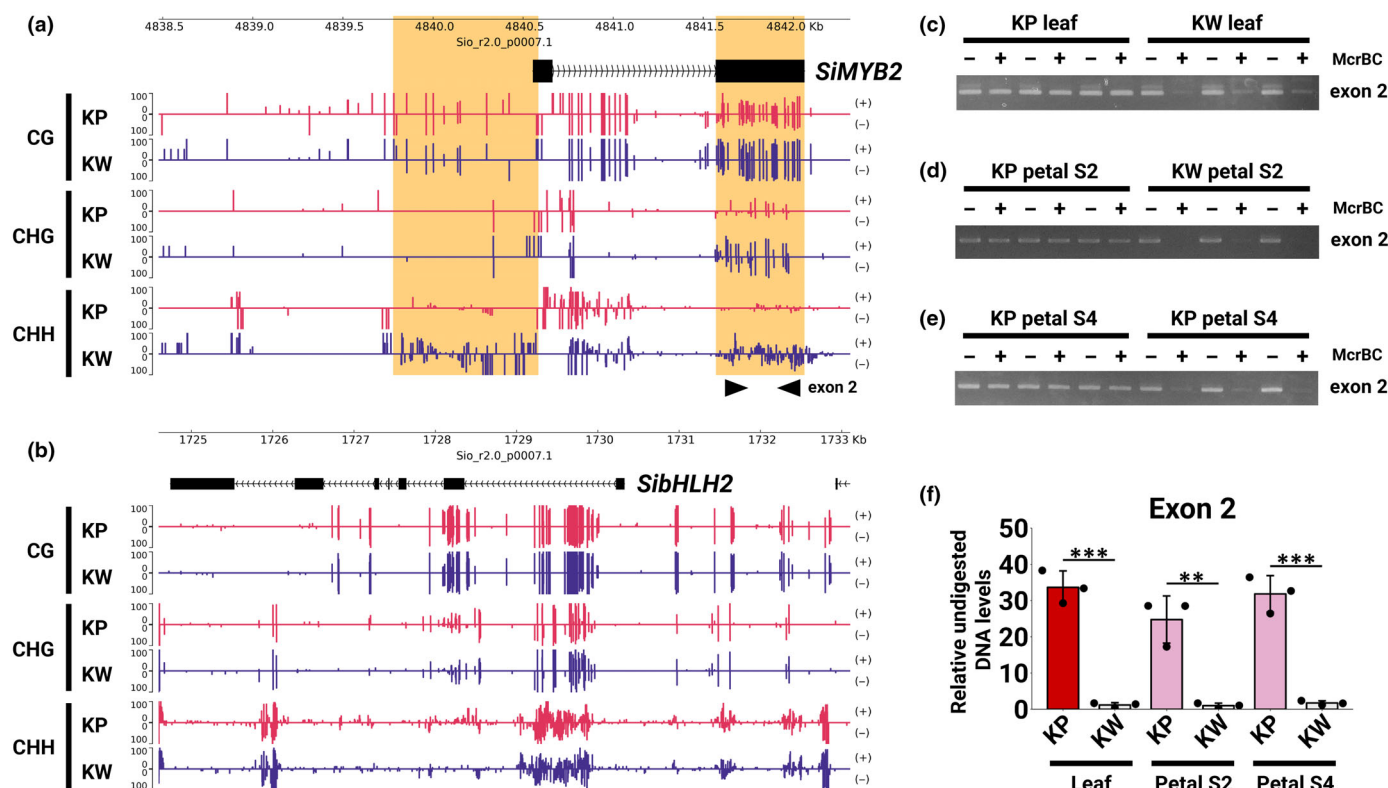


Fig. 4 (a, b) Methylation levels of candidate genes responsible for unstable leaf pigmentation. The vertical axis indicates the methylation level, and the horizontal axis indicates the position in the *Sio_r2.0_p0007.1* contig. Methylation levels are shown for each cytosine context: CG, CHG (H represents A, T or C), CHH and each strand (plus and minus). Red bars indicate methylation levels in KP, and blue bars in KW. (a) Methylation levels in *SiMYB2*. Regions where differences in methylation levels were observed are highlighted in orange. (b) Methylation levels in *SibHLH2*. (c–e) MCRBC-PCR results of *SiMYB2* exon 2 in leaf (c), petal at S2 (d), and petal at S4 (e) of KP and KW. (f) MCRBC-qPCR results of *SiMYB2* exon 2 in KP and KW tissues. Asterisks indicate significant differences (Student's *t*-test, **, *P* < 0.01; ***, *P* < 0.001). The error bar represents mean $\pm$ SD (*n* = 3). KP, 'Kilauea' Pink; KW, 'Kilauea' White.

the promoter region of *SiMYB2* than in KP (Fig. 4a; Table S3). For *SibHLH2*, there were no differences in methylation levels in either the ORF or promoter regions (Fig. 4b; Table S3). The methylation levels of several other genes related to anthocyanin biosynthesis (*SiCHS1/2*, *SiCHI1*, *SiF3H1*, *SiDFR1*, *SiANS1/2*, *SiGSTF1*, *Si3GT1/2*, *Si3GRT1*, *SiMATE1/2/3*, *SiAT*, *SiF3'5'H*, *SiMYB1*, and *SibHLH1*) were also examined, but no differences were detected in the coding gene itself or its promoter regions (Fig. S5). MCRBC-PCR/qPCR results also showed a difference in methylation levels in the second exon of *SiMYB2* between KP and KW leaves (Fig. 4c,f). This difference was also observed in KP and KW petals (Fig. 4d–f). Thus, *SiMYB2* was identified as an epiallele associated with unstable pigment accumulation. The methylation level of the *SiMYB2* promoter region could not be assessed due to repetitive sequences, as discussed later, and a high number of MCRBC digestion sites.

Two transcripts selectively generated from the *SiMYB2* locus

SiMYB2 (*Sio_r2.0_p0007.1.g08820.1*) was significantly upregulated and accompanied by hypermethylation in anthocyanin-deficient tissues, suggesting its involvement in anthocyanin accumulation instability. Mapping of the *SiMYB2*

genomic region showed a splice junction (sj) between *SiMYB2* and *Sio_r2.0_p0007.1.g08810.1* in KP (Fig. 5a; sj2), suggesting that the *SiMYB2* gene region extends from *Sio_r2.0_p0007.1.g08810.1* to *Sio_r2.0_p0007.1.g08820.1*. By Sanger sequencing of the TA cloning products and *de novo* assembly of RNA-seq reads, transcripts consisted of exon peak 1 (p1) + p2 + p4, and those of p3 + p4 were identified (Dataset S1). This result indicates that the *SiMYB2* locus is a single gene that ranges over *c.* 18 kb and produces at least two RNAs, presumably from two different TSSs (Fig. 5b,c). Exon 1 (124 bp), exon 2 (130 bp), exon 3 (110 bp), and exon 4 (493 bp) were determined from the sequences of p1, p2, p3, and p4 (Fig. 5b). The two 5' UTRs (1 and 2) and one 3' UTR were estimated. Transcripts consisting of 5' UTR 1, exon 1, exon 2, exon 4, and 3' UTR were designated as *SiMYB2-Long*, and 5' UTR 2, exon 3, exon 4, and 3' UTR as *SiMYB2-Short* (Fig. 5c). The gene referred to as *SiMYB2* in the earlier sections was, in fact, the *SiMYB2-Short* variant. *SiMYB2-Long* was preferentially generated in KP and *SiMYB2-Short* in KW.

The expression levels of *SiMYB2-Long* and *SiMYB2-Short* were confirmed by qRT-PCR. In petals, *SiMYB2-Long* expression levels were significantly lower in KW than in KP petals, and in SW than in ST petals (Fig. 5d), whereas *SiMYB2-Short* expression levels were significantly higher in KW than in KP petals, and

in SW than in ST petals (Fig. 5d). This expression pattern for *SiMYB2-Long* and *SiMYB2-Short* was also observed in KP and KW leaves (Fig. 5d). The expression levels of *SiMYB2-Long* were

significantly higher in KP than in KW leaves, whereas the expression levels of *SiMYB2-Short* were significantly higher in KW than in KP leaves (Fig. 5d). The expression levels of *SiMYB2-Long*

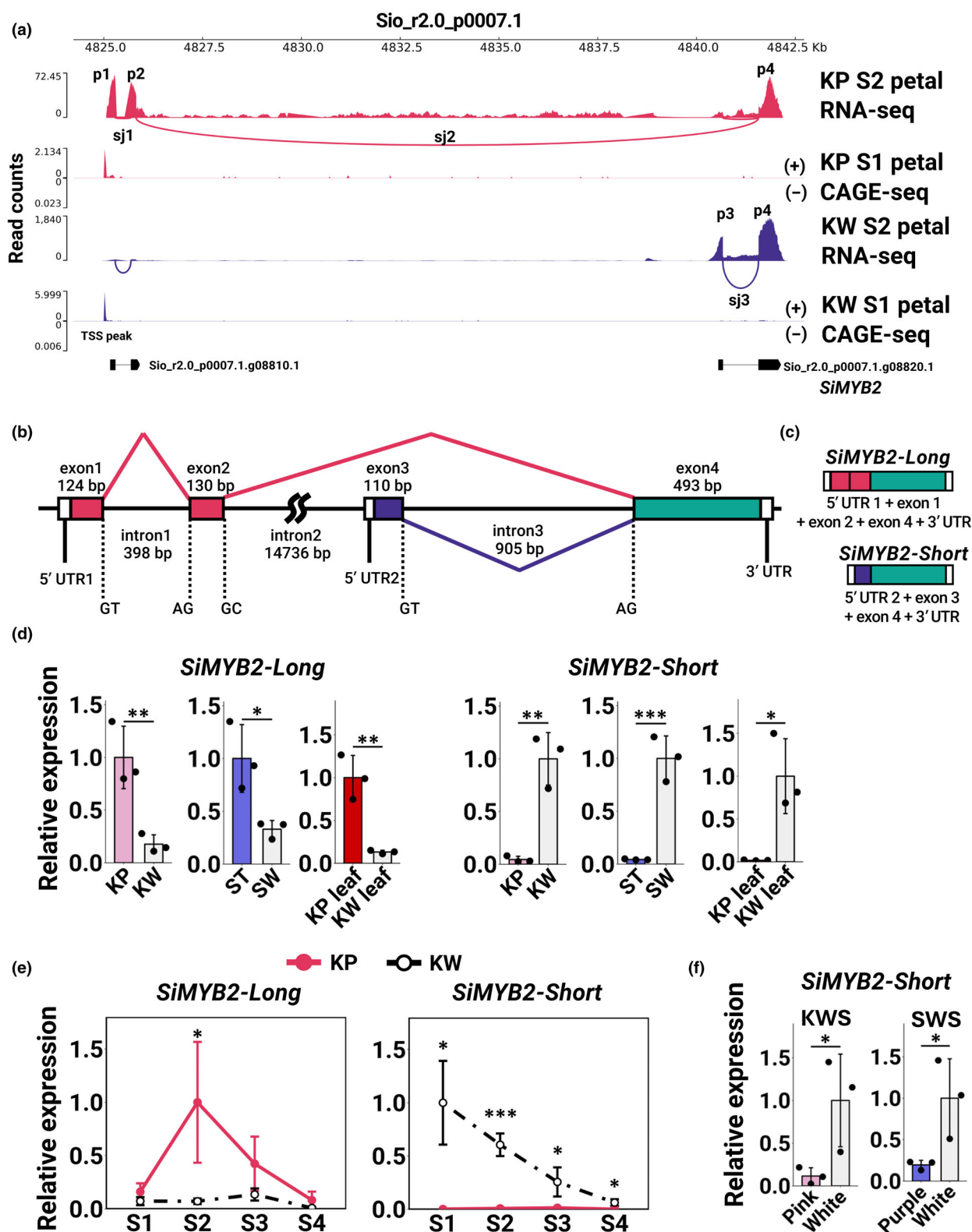


Fig. 5 Characteristics of *SiMYB2*. (a) RNA-seq (from S2 petal) and CAGE-seq (from S1 petal) mapping within the *SiMYB2* locus (Sio_r2.0_p0007.1.g08820.1) and its upstream region. Vertical axes represent read counts. Horizontal axes represent a position on the contig Sio_r2.0_p0007.1. Peaks in RNA-seq indicate exon sequences. Splicing junctions (sj1/2/3) were located between peak 1 (p1) and p2, p2 and p4, and p3 and p4. Although p1 + p2 and p3 + p4 were predicted as independent genes (Sio_r2.0_p0007.1.g08810.1, Sio_r2.0_p0007.1.g08820.1), they were the same gene. Peaks in CAGE-seq indicate a transcription start site. (b, c) Expected structure of the *SiMYB2* locus based on mapping and cloning results. Transcripts consisting of exons 1, 2, and 4 were designated as *SiMYB2-Long*, and those consisting of exons 3 and 4 were designated as *SiMYB2-Short*. (d) Quantitative reverse transcription PCR (qRT-PCR) analysis of *SiMYB2* variants in KP and KW petals, ST and SW petals, KP and KW leaves; pink and white regions in KWS petals, and purple and white regions in SWS petals. Asterisks indicate significant differences (Student's *t*-test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). Error bars represent mean $\pm$ SD ($n = 3$). (e) qRT-PCR analysis of *SiMYB2* variants in KP (solid red line) and KW (black dashed-dotted line) petals at different developmental stages. Asterisks indicate significant differences (Student's *t*-test, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Error bars represent mean $\pm$ SD ($n = 3$). (f) qRT-PCR analysis of *SiMYB2-Short* in pink and white regions in KWS petals, and purple and white regions in SWS petals. Asterisks indicate significant differences (Student's *t*-test, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Error bars represent mean $\pm$ SD ($n = 3$). KP, 'Kilauea' Pink; KW, 'Kilauea' White; KWS, 'Kilauea' White Stripe; ST, 'Stephanie'; SW, 'Stephanie' White; SWS, 'Stephanie' White Stripe.

and *SiMYB2-Short* in KP and KW petals were also examined at different developmental stages. Anthocyanin accumulation was detected between S2 and S3, with the expression levels of *SiMYB2-Long* being significantly higher at S2 in KP petals (Fig. 5e). In KW petals, the expression level was lower at all stages (Fig. 5e). *SiMYB2-Short* was hardly expressed at any stage in KP petals, while it was highly expressed in KW petals (Fig. 5e). Expression levels of *SiMYB2-Short* in the pigmented regions of striped flowers was significantly lower than in the nonpigmented regions of KWS and SWS petals (Fig. 5f), indicating that *SiMYB2-Long* is specifically expressed in colored tissues, whereas *SiMYB2-Short* is selectively transcribed in noncolored tissues.

SiMYB2-Long and *SiMYB2-Short* have different 5' UTR sequences, suggesting that these variants arise from alternative promoter usage (APU, Landry *et al.*, 2003). CAGE-seq analysis was performed to determine the TSSs of *SiMYB2* variants. If the two variants were produced via APU, peaks would be expected to detect the upstream regions of exon 1 and exon 3 of *SiMYB2*, respectively. However, only a single TSS was detected upstream of exon 1 in both KP and KW samples (Fig. 5a). Consequently, the mechanism by which these two transcription variants are generated could not be determined in this study. Note that the upstream sequence of *SiMYB2-Short* contains a TA-rich region with several potential TATA boxes (Fig. S6; Mukumoto *et al.*, 1993).

Functional analysis of *SiMYB2* by transient assays

The amino acid sequences of *SiMYB1*, *SiMYB2-Long*, and *SiMYB2-Short* were compared with other R2R3-MYB TFs belonging to the SG6 clade. *SiMYB2-Short* does not contain the entire R2 domain or most of the R3 domain, which are important conserved regions for MYB genes (Fig. 6a), but the R2–R3 domains of *SiMYB1* and *SiMYB2-Long* were nearly identical to those of other MYBs, and the bHLH binding domain ([D/E] Lx₂[R/K]x₃Lx₆Lx₃R) and SG6 motif (KPRPR[S/T]F) were intact (Fig. 6a; Stracke *et al.*, 2001; Zimmermann *et al.*, 2004).

The functions of *SibHLH1*, *SiMYB1*, *SiMYB2-Long*, and *SiMYB2-Short* were examined by transient assays in tobacco leaves. Active expression of the gene of interest was confirmed by qRT-PCR (Fig. S7). Expression of *SiMYB2-Long* alone resulted in very little accumulation of anthocyanins (Fig. 6b). When *SibHLH1*, *SiMYB1*, or *SiMYB2-Short* were expressed individually, no

anthocyanin accumulation was observed (Fig. 6b). Subsequently, *SibHLH1* was selected as a potential cofactor for *SiMYB1* and *SiMYB2*, as its expression levels did not differ between colored and noncolored tissues. Anthocyanin accumulated when *SiMYB1* plus *SibHLH1* or *SiMYB2-Long* plus *SibHLH1* were expressed (Fig. 6c). There was no accumulation of anthocyanin in tissues where *SiMYB2-Short* was co-expressed with *SibHLH1* (Fig. 6c). Anthocyanin content was also measured when MYB and bHLH were simultaneously expressed. The anthocyanin content of *SiMYB2-Long* co-expressed with *SibHLH1* was the highest among all transient assays, while the anthocyanin content of tissues co-expressing *SiMYB2-Short* and *SibHLH1* was not different from the *GFP* control (Fig. 6d). The anthocyanin content of *SiMYB1* plus *SibHLH1* was lower than that of *SiMYB2-Long* plus *SibHLH1*, as reflected in its fainter coloring (Fig. 6d). The expression levels of ABGs were confirmed when these TFs were expressed. Several ABGs were upregulated in *SiMYB2-Long* only, *SiMYB1* plus *SibHLH1*, and *SiMYB2-Long* plus *SibHLH1*, whereas there was no upregulation of ABGs in *SiMYB2-Short* plus *SibHLH1* co-expression assays (Fig. 6e). Compared to *SiMYB1* plus *SibHLH1*, *SiMYB2-Long* plus *SibHLH1* had higher ABG expression, which is consistent with its anthocyanin content (Fig. 6e).

To investigate the possibility that *SiMYB2-Short*, whose expression is higher in nonpigmented tissues, indirectly inhibits R2R3-MYB function in a dosage-dependent manner, similar to CPC type R3-MYB (Li *et al.*, 2020). Bacterial suspensions expressing gradually increasing ratios of *SiMYB2-Short* to *SiMYB1* and *SibHLH1* were injected into tobacco leaves. Anthocyanin accumulation was not affected by changing the ratio of *SiMYB2-Short* (Fig. S8).

The origin of *SiMYB2-Short* is the insertion of a transposon-like sequence

RT-PCR was performed to investigate whether the transcriptional selection of *SiMYB2-Long* and *SiMYB2-Short* also occurs in other cultivars. Two *SiMYB2* variants were found to be transcribed in 'Taro' and 'Saturn', which are similar to KP and ST (Fig. S9a). By contrast, *SiMYB2-Long* was amplified in 'Akira', 'Tomoko', 'Georgia', 'Barbara', and 'Manitoba', but *SiMYB2-Short* was not (Fig. S9a). 'Taro' and 'Saturn', like KP and ST, are cultivars in which anthocyanin accumulation is unstable, showing

loss and reversion during tissue culture propagation (Fig. S1f–i). ‘Akira’, ‘Tomoko’, ‘Georgia’, ‘Barbara’, and ‘Manitoba’ are cultivars with consistent anthocyanin accumulation in their leaves and flowers (Fig. S1a–e).

The genomic structure of *SiMYB2* was examined in ‘Akira’, ‘Tomoko’, ‘Georgia’, ‘Barbara’, and ‘Manitoba’ using Amplicon-seq, which showed that there is a *c.* 2.5 kb deletion, including exon 3, which is required for the expression of *SiMYB2-Short*, in all five cultivars (Figs 7a, S9b). To validate this finding, genomic PCR was performed using primers designed to flank the deletion region (Fig. 7a). As with the Amplicon-seq results, only the deletion-type band was amplified in ‘Akira’, ‘Tomoko’, ‘Georgia’, ‘Barbara’, and ‘Manitoba’ (Fig. 7b). In KP, only the insertion-type band was amplified, whereas in ST, ‘Taro’ and ‘Saturn’, both insertion- and deletion-type bands were detected (Fig. 7b). These results indicate that the absence of *SiMYB2-Short* expression in ‘Akira’, ‘Tomoko’, ‘Georgia’, ‘Barbara’, and ‘Manitoba’ is due to deletion.

The deleted region DNA sequence contains features of a transposable element-like (TE-like), including terminal inverted repeats (TIRs) (Fig. 7a,c). The 5' TIR was 171 bp, the 3' TIR was 170 bp, and the target site duplication (TSD) was 10 bp. A BLAST search for homologous transposon sequences failed to reveal any highly similar matches. Using the TIR sequences as queries, a BLAST search against the *Saintpaulia* SIO_r2.0 genome identified numerous homologous sequences (Table S4), suggesting that *SiMYB2-Short* may have arisen due to the insertion of an endogenous TE-like element.

Like KW and SW, ‘Shimai’ (SH) is a cultivar that does not accumulate anthocyanins in its leaves. When SH was propagated via tissue culture, the resulting plants also did not accumulate anthocyanins in their leaves, consistent with the original SH phenotype (Fig. S1j). Similarly, flowers of regenerated plants were green and lacked anthocyanin accumulation (Fig. S1k). However, a single mutant, referred to as m-SH, emerged with red-tinted leaves and flowers (1 out of 50 regenerated plants; Fig. S1j,k). The anthocyanin content of m-SH leaves was significantly higher than that in wild-type SH leaves (Fig. 7d). The expression of *SiMYB2-Long* was barely detectable in SH, but was highly expressed in m-SH (Fig. 7e). *SiMYB2-Short* was expressed in SH but was not detected in m-SH (Fig. 7e).

The genomic structure of *SiMYB2* was examined in these lines using In/Del primers, as in previous analyses. SH showed amplification of both insertion- and deletion-type bands (Fig. 7f), whereas in m-SH, only the deletion-type band was detected (Fig. 7f). These results suggest that the presence of *SiMYB2-Short* expression or the TE-like region may be essential for the loss of anthocyanin accumulation or its instability.

Discussion

In this study, we identified an epiallele involved in unstable pigmentation patterning in *Saintpaulia* plants. Anthocyanins did not accumulate, and the expression levels of several ABGs were low in nonpigmented tissues. RNA-seq analysis was performed to determine candidate TFs, and a MYB TF, *SiMYB2*, was identified as

an epiallele through WGBS analysis. Two transcripts were generated from the *SiMYB2* locus, one of which was highly expressed in pigmented colored tissues, and the other in nonpigmented tissues. These transcripts were designated as *SiMYB2-Long* and *SiMYB2-Short*, respectively, reflecting their relative lengths. Transient assays showed that *SiMYB2-Long* is an anthocyanin biosynthesis activator. *SiMYB2-Short* was highly expressed in nonpigmented tissues, suggesting that it is a repressor-type MYB TF. Paradoxically, the transient assay results also revealed that *SiMYB2-Short* does not have either a positive or negative impact on anthocyanin expression. Sequence analysis of the *SiMYB2* gene among several cultivars suggested that *SiMYB2-Short* resulted from the insertion of a transposon-like sequence. These observations provide a compelling model to help explain why mutants generated by tissue culture from an anthocyanin-deficient cultivar, in which *SiMYB2-Short* is expressed, were able to accumulate anthocyanins.

Formation of white striped petals due to periclinal chimeras vs epigenetics

The white-striped flowers produced by *Saintpaulia* are thought to be the result of a periclinal chimeric structure (Lineberger & Druckenbrod, 1985; Kazemian *et al.*, 2019). However, there is currently no evidence that they are due to the peripheral chimera structure. In this study, the tissue culture progeny of *Saintpaulia* ‘Kilauea’ Pink (KP) and ‘Kilauea’ White Stripe flowers varied in their petal and leaf color patterns (Fig. 1a; Table 1). A previous study reported that the phenotypic variation of petal color varied greatly in plants regenerated from the tissue culturing of cultivar ‘Kilauea’ (Yang *et al.*, 2017). The characteristic of the *Saintpaulia* phenotype that changes in tissue culture and then reverts to the original phenotype suggests that this phenotype is the result of changes in epigenetic regulation.

White-striped petals closely resemble striped petals caused by periclinal chimeras, which may lead to the misconception that they too are periclinal chimeras. Periclinal chimeras occur when the cell layers of the shoot apical meristem have distinct genetic backgrounds and coexist within a single plant (Satina *et al.*, 1940). When such genetic differences involve anthocyanin biosynthesis-related genes, *Saintpaulia* flowers can exhibit striped petals. For example, the periclinal chimera cultivar ‘Kaname’ has a mutation in the *F3'5'H* gene restricted to the L1 layer, resulting in pink and blue striped petals (Yang *et al.*, 2017). Because adventitious shoots regenerate almost invariably from a single L1 cell (Naylor & Johnson, 1937; Broertjes & Keen, 1980; Ohki, 1994), such chimeras give rise to genetically uniform plants with monochromatic flowers when tissue culture (Yang *et al.*, 2017). In this study, however, *Saintpaulia* with white striped petals eventually flowered with fully colored petals. Moreover, white-striped petals arose post-developmentally from *Saintpaulia* plants that had initially exhibited entirely white petals. Thus, it cannot be explained by periclinal chimerism alone, since single-cell origin should eliminate layer-specific genetic mosaics. How, then, are the white-striped petals formed? In white-striped petals, the presence or absence of *SiMYB2-Short* expression correlated with a lack of pigmentation (Fig. 5g). Additionally, the

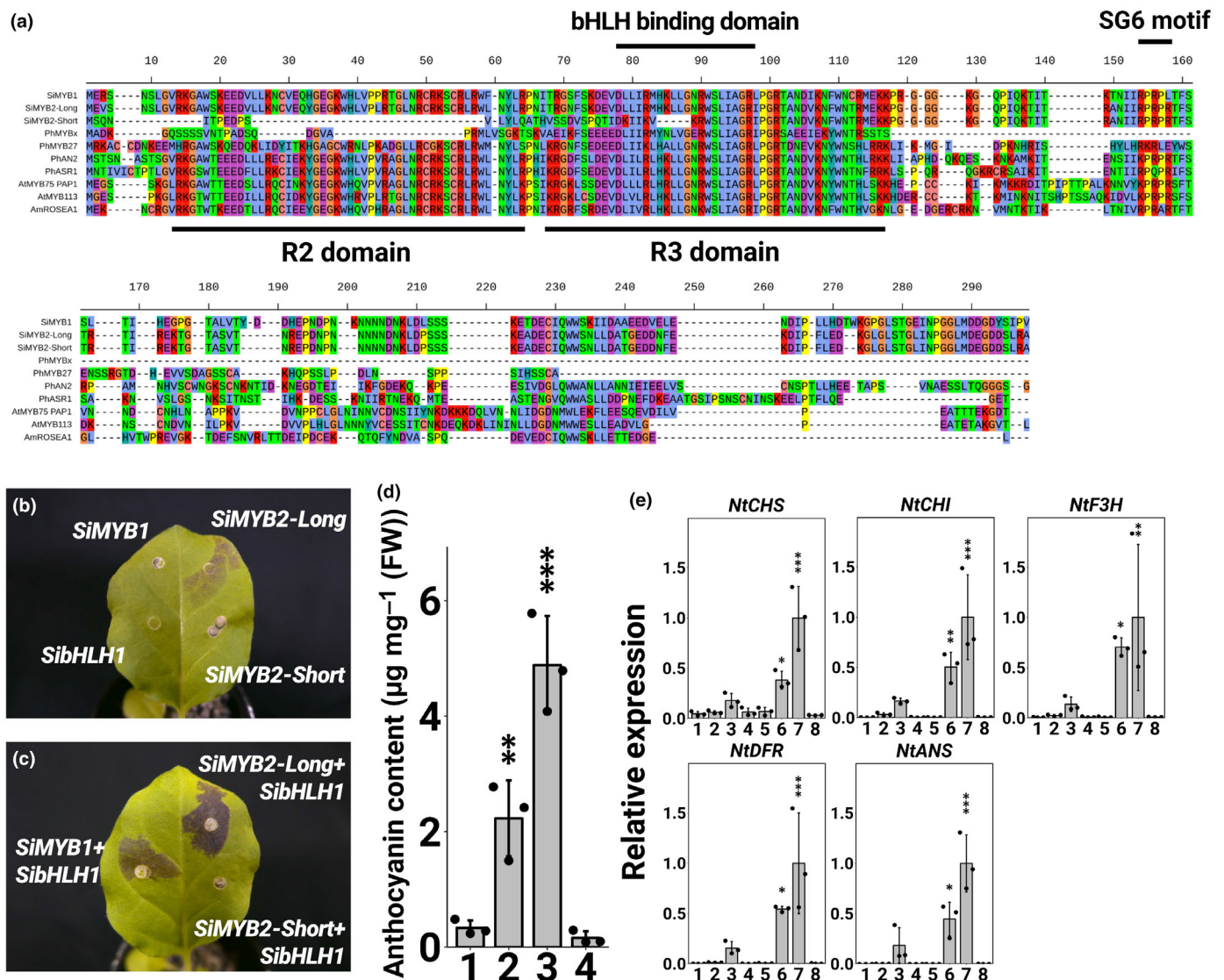


Fig. 6 (a) Multiple alignment of MYB amino acids from *Saintpaulia ionantha* and other plants. The R2/R3 domain and SG6 motif were based on Stracke *et al.* (2001), and the bHLH binding domain was based on Zimmermann *et al.* (2004). (b–e) Results of transient assays using tobacco leaves. (b) Tobacco leaves when the target transcription factor (TF) was expressed alone. *SiMYB1* (top left), *SiMYB2-Long* (top right), and *SiMYB2-Short* (bottom right). (c) Tobacco leaves when MYB TFs were expressed simultaneously with *SibHLH1*. *SiMYB1* + *SibHLH1* (left), *SiMYB2-Long* + *SibHLH1* (top right), and *SiMYB2-Short* + *SibHLH1* (bottom right). Tobacco leaves were photographed 4 d post-infiltration. (d) Anthocyanin content in tobacco leaves when MYB and bHLH TFs were expressed simultaneously. Anthocyanin content was determined as cyanidin-3-glucoside equivalents. Error bars represent mean $\pm$ SD ($n = 3$). 1: GFP; 2: *SiMYB1* + *SibHLH1*; 3: *SiMYB2-Long* + *SibHLH1*; 4: *SiMYB2-Short* + *SibHLH1*. Asterisks indicate significant differences according to Dunnett's range test compared to GFP (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). (e) Quantitative reverse transcription PCR (qRT-PCR) analysis of ABGs in infiltrated tobacco leaves. 1: GFP; 2: *SiMYB1*; 3: *SiMYB2-Long*; 4: *SiMYB2-Short*; 5: *SibHLH1*; 6: *SiMYB1* + *SibHLH1*; 7: *SiMYB2-Long* + *SibHLH1*; 8: *SiMYB2-Short* + *SibHLH1*. Error bars represent mean $\pm$ SD ($n = 3$). Asterisks indicate significant differences according to Dunnett's range test compared to GFP (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

transcriptional selection of *SiMYB2* variants appears to be regulated epigenetically. Based on these observations, we propose an alternative hypothesis that white-striped petals may result from the coexistence of cell layers with distinct epigenetic backgrounds (Fig. 8). In other words, it may be an epigenetic periclinal chimera. Indeed, genes with cell layer-specific expression have been reported in *A. thaliana* (Lu *et al.*, 1996). Tissue culture may alter the epigenomic state of the L2 layer or below tissues, potentially contributing to this layered epigenetic pattern (Fig. 8).

Identification of the epiallele linked to unstable pigmentation

The traits of plants obtained through tissue culture or vegetative propagation are not always consistent. Such variations are occasionally attributable to alterations in epigenetic status. For example, during the tissue culture of *E. guineensis*, a reduction in the methylation level of the *Karma* transposable element within the *Deficiens1* (*DEF1*) gene leads to splicing errors in the *DEF1* gene,

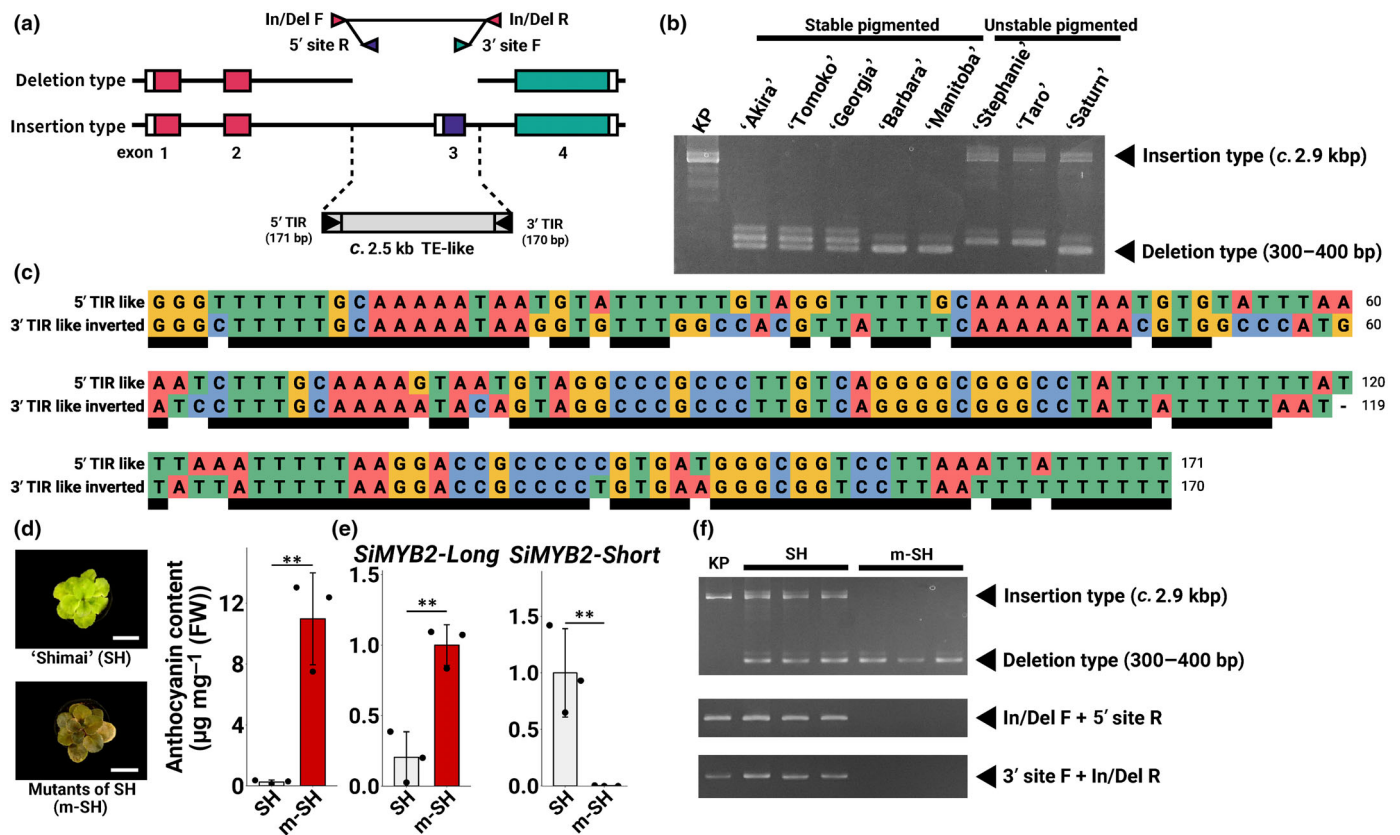


Fig. 7 (a) The position of primers designed from the *SiMYB2* locus sequence is indicated by arrowheads. (b) Confirmation of indels by PCR in some cultivars. The abbreviated names of the cultivars are shown across the top of the gel lanes. The positions of the primers (In/Del F and R) used are shown in (a). (c) Multiple sequence alignment results of TIR-like sequences. Black bars indicate corresponding bases between 5' TIR and 3' TIR inverted sequences. (d–f) Characterization of the mutants of 'Shimai' (m-SH) obtained by tissue culture and parental cultivar 'Shimai' (SH). (d) Anthocyanin content of SH and m-SH leaves. The content was determined as cyanidin-3-glucoside equivalents. Error bars represent mean $\pm$ SD ($n = 3$). Asterisks indicate significant differences (Student's *t*-test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). (e) Quantitative reverse transcription PCR (qRT-PCR) quantification of *SiMYB2* variants in leaves. Error bars represent mean $\pm$ SD ($n = 3$). Asterisks indicate significant differences (Student's *t*-test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). (f) Confirmation of the presence of a transposon-like element by PCR in SH and m-SH. The positions of the primers used are shown in (a).

resulting in malformed fruit (Ong-Abdullah *et al.*, 2015). In *C. morifolium*, yellow-flowered plants arise from the lateral buds of red-flowered plants. This phenomenon is caused by hypermethylation in the promoter region of *CmMYB6* (Tang *et al.*, 2022). In our case, tissue culture may have altered the epigenetic state of *Saintpaulia*, leading to phenotypic variations.

Based on the hypothesis that unstable anthocyanin accumulation in *Saintpaulia* is caused by epigenetic regulation, we determined that *SiMYB2* is the epiallele associated with coloration by a combination of RNA-seq and WGBS analyses. *CmMYB6* in *C. morifolium* was identified as an epiallele involved in flower color changes using a similar approach as in the present study (Tang *et al.*, 2022). *EgDEF1* in *E. guineensis* was also identified by epigenome-wide association analysis and WGBS (Ong-Abdullah *et al.*, 2015). Exploration of other epialleles that control unstable traits, such as unstable anthocyanin accumulation, by this combination of methods may yield additional useful candidates for the identification of causal genes. *Saintpaulia* has been propagated and maintained in large quantities by vegetative propagation, such as leaf cuttings and tissue culture. Epigenetic mutants, such as the anthocyanin-depleted individuals in this study, are often

produced by vegetative propagation that is maintained until the next generation. For example, variation in transposon transposition frequency and altered floral patterns have been reported in tissue culture progeny of *Saintpaulia* 'Thamires' (Sato *et al.*, 2011). It is also postulated that these traits are the result of epimutation. It may therefore be feasible to utilize the epigenetic traits observed in *Saintpaulia* for genetic identification purposes.

In *Petunia*, it has been reported that gene expression can be induced by DNA methylation (Shibuya *et al.*, 2009). The expression of *SiMYB2-Short* may also be induced by the methylation of its promoter region. However, the promoter region contains transposon-like sequences dispersed throughout the genome, making it difficult to design specific primers. Consequently, PCR-based validation could not be performed. Further detailed evaluation of this region remains a subject for future investigation.

The present study identified *SiMYB1*, *SibHLH1*, and *SibHLH2* as TFs associated with anthocyanin accumulation in petals, in addition to *SiMYB2* (Fig. 3). The transient assay results demonstrated that *SiMYB1* functions as an anthocyanin biosynthesis activator (Fig. 6c–e). Therefore, *SiMYB1* expression

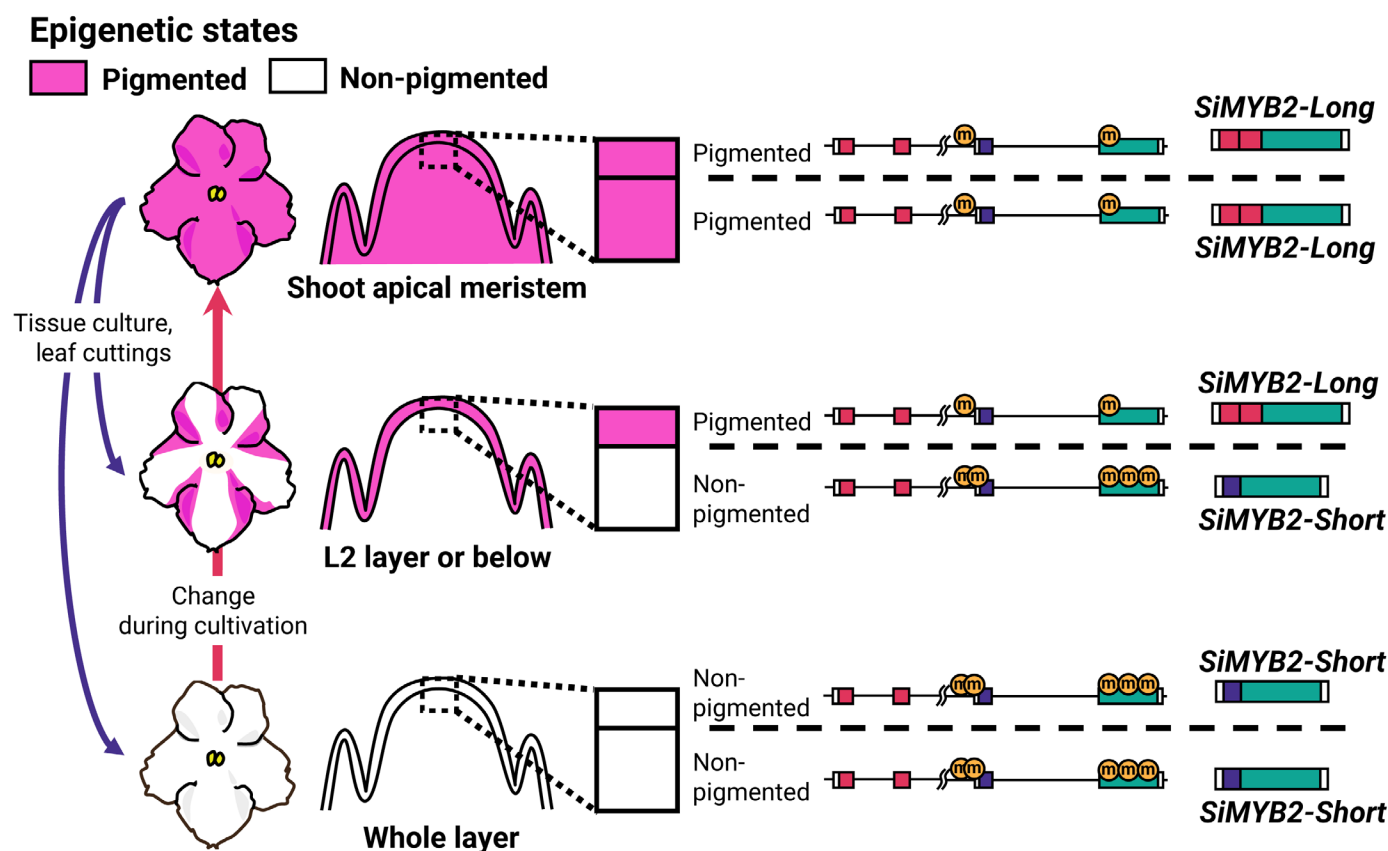


Fig. 8 Summary of this study and hypothetical model to explain white stripe patterns. Pigmented petals have an epigenomic state in which *SiMYB2-Long* is expressed. Nonpigmented petals have an epigenomic state in which *SiMYB2-Short* is expressed. White striped petals may be the result of an epigenomic state in which *SiMYB2-Long* is expressed in the L1 layer and *SiMYB2-Short* in the L2 layer and below.

may be the determinant of unstable pigmentation. However, *SiMYB1* was barely expressed in leaves exhibiting unstable pigment accumulation like petals (Fig. 3d). By contrast, the transcriptional selectivity of *SiMYB2-Long* and *SiMYB2-Short* was linked to pigment accumulation in the leaves (Figs 3c, 7d,e). These results indicate that *SiMYB1* is not the gene responsible for the phenomenon of unstable pigmentation.

The expression levels of *SibHLH1*, which is thought to regulate the transcription of ABGs, were essentially the same in white and colored petals. Phylogeny placed *SibHLH1* in the IIIf b clade, which includes *PhJAF13* (Fig. 3f; Quattrocchio *et al.*, 1998; Heim, 2003). In *Petunia*, the expression of *PhJAF13* is not influenced by the expression of MYB TFs (Quattrocchio *et al.*, 1998; Spelt *et al.*, 2000), which is consistent with our observation that the expression of *SibHLH1* was unchanged in the white parts of *Saintpaulia*, where the expression of the MYB TFs was downregulated. *SibHLH2* clustered with the IIIf a clade, which includes *PhAN1* (Fig. 3f; Spelt *et al.*, 2000; Heim, 2003). The loss of *PhAN1* function results in the loss of anthocyanin biosynthesis in *Petunia* (Spelt *et al.*, 2000). Thus, the reduction in *SibHLH2* expression may contribute to the loss of anthocyanin in *Saintpaulia*. It has been reported that *NtJAF13*, a member of the IIIf b

clade, interacts with MYB TFs and promotes the expression of *NtAN1*, a member of the IIIf a clade in *N. tabacum* (Montefiori *et al.*, 2015). Moreover, *AmDelila*, a member of clade IIIf b, binds to the *Rosea1* MYB TF and positively regulates the transcription of *AmIncolorata 1*, a member of clade IIIf a, in *Antirrhinum majus* (Albert *et al.*, 2021). Therefore, the downregulation of *SibHLH2* in the nonpigmented tissues is most likely due to decreased *SiMYB2-Long* expression levels. Moreover, *SiMYB2* methylation levels differed between KP and 'KW', while *SibHLH2* did not (Fig. 4), suggesting that differences in *SibHLH2* expression are not the underlying cause of unstable pigmentation.

Transcriptional selectivity of *SiMYB2* variants and pattern formation

Alternative splicing and APU are two regulatory mechanisms by which different transcripts can be produced from a single locus (Reviewed in Landry *et al.*, 2003). An example of unstable pigmentation caused by different variants has been reported in *C. morifolium*. Alternative splicing of *CmbHLH2* produces different variants, with nonfunctional *CmbHLH2* in white petals and functional *CmbHLH2* in colored petals (Lim *et al.*, 2021).

However, the present work is the first report that two or more transcripts from a single locus are involved in flower pattern formation. It was thought that *SiMYB2* variants were generated by APU since the 5' UTRs of *SiMYB2-Long* and *SiMYB2-Short* were unique to each, particularly since multiple TATA boxes were observed in the predicted promoter region of *SiMYB2-Short*. However, we were unable to detect the TSS of *SiMYB2-Short* by CAGE-seq, thus preventing a conclusive determination about whether *SiMYB2* variants are generated through APU. CAGE-seq specifically sequences the 5' ends of RNAs with a 5' cap structure (m7G). Therefore, the absence of a detectable TSS for *SiMYB2-Short* may be attributed to the lack of a 5' cap structure in *SiMYB2-Short* transcripts. Although this study could not identify the precise mechanism by which *SiMYB2* variants are generated, the findings from CAGE-seq provide valuable insights for future *SiMYB2* variant studies.

SiMYB2-Short was preferentially transcribed in both petals and leaves where anthocyanins were absent, and was already highly expressed by stage S1 in KW petals, which was earlier than the developmental stage at which *SiMYB2-Long* was expressed in KP petals. Mutants lacking the TE-like sequence accumulated anthocyanins at high levels. In other anthocyanin-accumulating cultivars, the loss of the TE-like sequence may result in anthocyanin production because transcription of functional *SiMYB2-Long* is the only possibility. However, anthocyanin accumulation can still occur in some cases, even in the presence of the TE-like sequence. Notably, in tissues where anthocyanins accumulate, *SiMYB2-Short* is not transcribed, while *SiMYB2-Long* is actively transcribed. These findings suggest that nonproduction of anthocyanin may result from the transcription of *SiMYB2-Short* rather than the mere presence of the TE-like sequence.

The specific function of *SiMYB2-Short* could not be identified in this study, so further research is necessary to determine its role. If stable anthocyanin accumulation is desired, deleting the TE-like sequence may be beneficial. On the other hand, the presence of the TE-like sequence may enable transcriptional selectivity between *SiMYB2* variants, potentially giving rise to patterned flowers, albeit with instability. It is important to note that the presence of the TE-like insertion alone does not necessarily result in anthocyanin loss; additional factors may influence the transcriptional selectivity of *SiMYB2-Long* and *SiMYB2-Short*.

Acknowledgements

We are grateful to Dr Atsushi Hoshino of the National Institute for Basic Biology and Associate Professor Takuji Tsukiyama of Kindai University for their support in identifying transposons. We also wish to acknowledge the assistance provided by Assistant Professor Akira Yamazaki of Kindai University in the preparation and use of experimental equipment. We also thank Mr Ryutaro Kunimune and Kazuki Kobayashi of Kindai University for their help in cultivation management. We would like to express our gratitude to Dr Yu Kinoshita of Kyoto University for her invaluable discussion on the subject of epigenetics and for her assistance with cultivation. This work was partly supported by the Sasakawa Scientific Research Grant from the Japan Science Society

(2024-4050) and a grant from the Agricultural Technology and Innovation Research Institute (ATIRI), Kindai University. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.

Competing interests

None declared.

Author contributions

MH conceptualized the study. DK, TT, KS and FT carried out the investigation. DK, TT, KS, HH and FT performed the data curation. DK, TT and MH provided the resources. DK, KS and HH conducted the formal analysis. DK performed visualization. DK, HH and MH secured the funding. MH provided the supervision. MH handled the project administration. DK wrote the original manuscript. KS, HH, FT and MH reviewed and edited the original manuscript.

ORCID

Hideki Hirakawa  <https://orcid.org/0000-0002-7128-7390>

Munetaka Hosokawa  <https://orcid.org/0000-0002-1021-0381>

Daichi Kurata  <https://orcid.org/0009-0009-7440-8280>

Kenta Shirasawa  <https://orcid.org/0000-0001-7880-6221>

Fumi Tatsuzawa  <https://orcid.org/0009-0005-4597-3273>

Data availability

The newly generated raw read data were deposited in the Sequence Read Archive (SRA) database of the DNA Data Bank of Japan (DDBJ) under the BioProject accession no. PRJDB22858.

References

- Akalın A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. 2012. METHYLKIT: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology* 13: R87.
- Albert NW, Butelli E, Moss SMA, Piazza P, Waite CN, Schwinn KE, Davies KM, Martin C. 2021. Discrete bHLH transcription factors play functionally overlapping roles in pigmentation patterning in flowers of *Antirrhinum majus*. *New Phytologist* 231: 849–863.
- Bolger AM, Lohse M, Usadel B. 2014. TRIMMOMATIC: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Broertjes C, Keen A. 1980. Adventitious shoots: do they develop from one cell? *Euphytica* 29: 73–87.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. TRIMAL: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.
- Eyerdom H. 1981. Flower color sports and variations in *Saintpaulia* hybrids. *African Violet Magazine* 29: 33–37.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* 2011. Full-length transcriptome

- assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
- Heim MA. 2003. The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Molecular Biology and Evolution* 20: 735–747.
- Jaligot E, Hooi WY, Debladis E, Richard F, Beulé T, Collin M, Agbessi MDT, Sabot F, Garsmeur O, D'Hont A *et al.* 2014. DNA methylation and expression of the EgDEF1 gene and neighboring retrotransposons in mantled somaclonal variants of oil palm. *PLoS ONE* 9: e91896.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Kazemian M, Ghasemi omran V, Mohajel Kazemi E, Kolahi M. 2019. Regeneration of pinwheel phenotype and evaluation of anthocyanin in African violet (*Saintpaulia ionantha* Wendl.) periclinal chimera. *Journal of Plant Molecular Breeding* 7: 39–49.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* 37: 907–915.
- Kurata D, Fukutomi K, Kubo K, Shirasawa K, Hirakawa H, Hosokawa M. 2024. Comprehensive expression analysis of ERF transcription factors during chilling acclimation in *Saintpaulia*. *Plant Growth Regulation* 104: 745–759.
- Landry J-R, Mager DL, Wilhelm BT. 2003. Complex controls: the role of alternative promoters in mammalian genomes. *Trends in Genetics* 19: 640–648.
- Li Y, Shan X, Gao R, Han T, Zhang J, Wang Y, Kimani S, Wang L, Gao X. 2020. MYB repressors and MBW activation complex collaborate to fine-tune flower coloration in *Freesia hybrida*. *Communications Biology* 3: 396.
- Lim S-H, Kim D-H, Jung J-A, Lee J-Y. 2021. Alternative splicing of the basic helix-loop-helix transcription factor gene *CmbHLH2* affects anthocyanin biosynthesis in ray florets of chrysanthemum (*Chrysanthemum morifolium*). *Frontiers in Plant Science* 12: 669315.
- Lineberger RD, Druckenbrod M. 1985. Chimeral nature of the pinwheel flowering african violets (*Saintpaulia*, Gesneriaceae). *American Journal of Botany* 72: 1204–1212.
- Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, Ramírez F, Manke T. 2021. pyGENOMETRACKS: reproducible plots for multivariate genomic datasets. *Bioinformatics* 37: 422–423.
- Lu P, Porat R, Nadeau JA, O'Neil SD. 1996. Identification of a meristem L1 layer-specific gene in Arabidopsis that is expressed during embryonic pattern formation and defines a new class of Homeobox genes. *Plant Cell* 8: 2155–2168.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.
- Minh BQ, Nguyen MAT, Von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30: 1188–1195.
- Möller M, Cronk QCB. 1997. Phylogeny and disjunct distribution: evolution of *Saintpaulia* (Gesneriaceae). *Proceedings of the Royal Society of London. Series B: Biological Sciences* 264: 1827–1836.
- Montefiori M, Brendolise C, Dare AP, Lin-Wang K, Davies KM, Hellens RP, Allan AC. 2015. In the Solanaceae, a hierarchy of bHLHs confer distinct target specificity to the anthocyanin regulatory complex. *Journal of Experimental Botany* 66: 1427–1436.
- Mukumoto F, Hirose S, Imaseki H, Yamazaki K. 1993. DNA sequence requirement of a TATA element-binding protein from Arabidopsis for transcription *in vitro*. *Plant Molecular Biology* 23: 995–1003.
- Murashige T, Skoog F. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum* 15: 473–497.
- Nabeshima T, Yang S-J, Ohno S, Honda K, Deguchi A, Doi M, Tatsuzawa F, Hosokawa M. 2017. Histogen layers contributing to adventitious bud formation are determined by their cell division activities. *Frontiers in Plant Science* 8: 1749.
- Naylor EE, Johnson B. 1937. A histological study of vegetative reproduction in *Saintpaulia ionantha*. *American Journal of Botany* 24: 673–678.
- Ohki S. 1994. Scanning electron microscopy of shoot differentiation *in vitro* from leaf explants of the African violet. *Plant Cell, Tissue and Organ Culture* 36: 157–162.
- Ong-Abdullah M, Ordway JM, Jiang N, Ooi S-E, Kok S-Y, Sarpan N, Azimi N, Hashim AT, Ishak Z, Rosli SK *et al.* 2015. Loss of *Karma* transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* 525: 533–537.
- Outchkourov NS, Carollo CA, Gomez-Roldan V, De Vos RCH, Bosch D, Hall RD, Beekwilder J. 2014. Control of anthocyanin and non-flavonoid compounds by anthocyanin-regulating MYB and bHLH transcription factors in *Nicotiana benthamiana* leaves. *Frontiers in Plant Science* 5: 519.
- Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. STRINGTIE enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 33: 290–295.
- Quattrocchio F, Wing JF, Va K, De W, Mol JNM, Koes R. 1998. Analysis of bHLH and MYB domain proteins: species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. *The Plant Journal* 13: 475–488.
- Ramsay NA, Glover BJ. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science* 10: 63–70.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. EDGER: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.
- Satina S, Blakeslee AF, Avery AG. 1940. Demonstration of the three germ layers in the shoot apex of datura by means of induced polyploidy in periclinal chimeras. *American Journal of Botany* 27: 895–905.
- Sato M, Kawabe T, Hosokawa M, Tatsuzawa F, Doi M. 2011. Tissue culture-induced flower-color changes in *Saintpaulia* caused by excision of the transposon inserted in the flavonoid 3', 5' hydroxylase (F3'5'H) promoter. *Plant Cell Reports* 30: 929–939.
- Shibuya K, Fukushima S, Takatsui H. 2009. RNA-directed DNA methylation induces transcriptional activation in plants. *Proceedings of the National Academy of Sciences, USA* 106: 1660–1665.
- Spelt C, Quattrocchio F, Mol JNM, Koes R. 2000. *anthocyanin1* of Petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. *Plant Cell* 12: 1619–1631.
- Stracke R, Werber M, Weisshaar B. 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology* 4: 447–456.
- Sun J, Nishiyama T, Shimizu K, Kadota K. 2013. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics* 14: 219.
- Tang M, Xue W, Li X, Wang L, Wang M, Wang W, Yin X, Chen B, Qu X, Li J *et al.* 2022. Mitotically heritable epigenetic modifications of *CmMYB6* control anthocyanin biosynthesis in chrysanthemum. *New Phytologist* 236: 1075–1088.
- Tatsuzawa F, Hosokawa M, Saito N, Honda T. 2012. Three acylated anthocyanins and a flavone glycoside in violet-blue flowers of *Saintpaulia* 'Thamires'. *South African Journal of Botany* 79: 71–76.
- Tatsuzawa F, Matsuda S, Kato K, Hosokawa M. 2015. Acetylated anthocyanidin 3-O-di-glycosides in red-purple flowers and gray-purple leaves of *Saintpaulia* 'Tomoko'. *The Horticulture Journal* 84: 77–82.
- Untergasser A, Cutcutache I, Korossar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3 – new capabilities and interfaces. *Nucleic Acids Research* 40: e115.
- Whelan S, Irsarri I, Burki F. 2018. PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* 34: 3929–3930.
- Xu W, Dubos C, Lepiniec L. 2015. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends in Plant Science* 20: 176–185.
- Yang SJ, Ohno S, Deguchi A, Sato M, Goto M, Doi M, Ohnishi M, Tatsuzawa F, Hosokawa M. 2017. The histological study in sympetalous corolla development of pinwheel-type flowers of *Saintpaulia*. *Scientia Horticulturae* 223: 10–18.
- Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF. 2004. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal* 40: 22–34.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Dataset S1 *SiMYB2-Long* sequences.

Fig. S1 Cultivars and mutants used in this study.

Fig. S2 Developmental stages of petals used in this study.

Fig. S3 HPLC analysis of KP and KW petals.

Fig. S4 Expression levels of ABGs in white-striped petals.

Fig. S5 Methylation levels of anthocyanin biosynthesis-related genes in KP and KW leaves.

Fig. S6 Predicted promoter sequences of *SiMYB2-Short*.

Fig. S7 qRT-PCR analysis of the agroinfiltrated genes in tobacco leaves.

Fig. S8 Testing of *SiMYB2-Short* repression activity as R3-MYB.

Fig. S9 Transcript variants expression and genome structure of the *SiMYB2* gene in several cultivars.

Table S1 List of primers used in this study.

Table S2 List of GenPept accession IDs used in phylogenetic tree.

Table S3 Result of DMR analysis in Sio_r2.0_p0007.1 contig.

Table S4 Location and length of other TE-like sequences that were hit as a result of BLAST using the TIR sequence of MYB2 TE-like as query.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Disclaimer: The New Phytologist Foundation remains neutral with regard to jurisdictional claims in maps and in any institutional affiliations.