Title:

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

Authors:

Daichi Kurata[1] · Tomohisa Tsuzaki[1] · Fumi Tatsuzawa[2] · Kenta Shirasawa[3] · Hideki Hirakawa[3] · Munetaka Hosokawa[1,4,*]

Affilinations:

[1]Graduate School of Agriculture, Kindai University, 3327-204, Nakamachi, Nara, Nara 631-8505, Japan

[2]Faculty of Agriculture, Iwate University, 3-18-8, Ueda, Morioka, Iwate 606-8550, Japan

[3]Kazusa DNA Research Institute, 2-6-7, Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

[4]Agricultural Technology and Innovation Research Institute Kindai University (ATIRI), 3327-204, Hanamachi, Nara, Nara 631-8505, Japan

*Correspondence

Munetaka Hosokawa

E-mail: mune@nara.kindai.ac.jp

ORCID: 0000-0002-1021-0381

Total word count: 7599 (Introduction: **408**, Materials and Methods: **2748**, Results: **2786**, Discussion: **1657)**

Number of figures: 8 (all in color)

Number of tables: 1

Number of supporting information: 12

Summary

- *Saintpaulia* (African violet) pigmentation is notoriously unstable, particularly following passage through tissue culture, but the instability mechanism is unknown. White-striped petals were thought to be due to epigenetics, but the genes involved are unknown. We confirmed that white stripes result from epigenetic regulation 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 non-colored tissues. Functional analyses revealed that SiMYB2-Long is an anthocyanin biosynthesis activator and SiMYB2-Short is non-functional. Exon 3 of *SiMYB2* was generated by the insertion of a transposon-like sequence. A mutant lacking the element was obtained from cultivars with non-colored 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 thorough the transcriptional selectivity of epigenetically regulated *SiMYB2*.

Key words:

Epigenetics, Periclinal chimera, Alternative Promoter Usage, Anthocyanin

Introduction

African violet (*Streptocarpus* sect. *Saintpaulia*) 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 pericarpal chimaeras are propagated through adventitious buds (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, Tang et al. (2022) identified a gene (*CmMYB6*) responsible for floral color variation emerging from lateral buds in Chrysanthemum (*Chrysanthemum morifolium*). Because both of these studies successfully used 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

70 *Saintpaulia*, based on the hypothesis that they result from epigenetic mutations. Using an integrated approach
71 combining gene expression analysis and epigenomic profiling, we aimed to identify the gene or genes underlying
72 this instability.
73
74 **Materials and methods**
75 **Plant materials and tissue culture**
76 *Saintpaulia* cultivar 'Kilauea' has white-striped petals, and we obtained plants with pink petals via tissue culture
77 of this cultivar (Fig. 1a, b). Propagated plants with pink petals were labeled as 'KP'. Leaves collected from KP
78 plants were surface sterilized with 1% sodium hypochlorite supplemented with 10 µl Tween 20 for 15 min. The
79 leaves were then rinsed with double-distilled water and cut into 1–2 cm^2 pieces using a razor. To obtain adventitious
80 shoots, the pieces were cultured on full-strength MS medium (Murashige & Skoog, 1962) containing 2% sucrose
81 (g/v), 0.3% gellan gum (g/v), 0.1 ppm 1-naphthalene acetic acid (NAA), and 0.5 ppm 6-benzyladenine (BA) for
82 four weeks. Regenerated plants with white petals were labeled 'KW' (Fig. 1c), and those with white-striped petals
83 were labeled 'KWS' (Fig. 1d). The cultivar 'Stephanie' (ST) was used as a comparator line. Stephanie plants
84 emerging from leaf cuttings with white petals (SW), and with white-striped petals (SWS) were labeled accordingly
85 (Fig. 1e, f, g). Cultivars 'Akira' (AK), 'Tomoko' (TK), 'Georgia' (GG), 'Barbara' (BB), 'Manitoba' (MT), 'Taro'
86 (TR), 'Saturn' (SA), 'Shimai' (SH), and mutants of SH (m-SH) were also included for anthocyanin assays and gene
87 analyses (Fig. 7d, e, Fig. S8a-i). All plants were grown at 21±2 °C, 14 h light/10 h dark, under fluorescent light
88 (BioLux, 20W, HotaluX, approximately 50 µmol cm^{-2} s^{-1}). The plants were planted in 220 ml plastic pots filled with
89 Metro-Mix 350J soil (HYPONeX, Osaka, Japan), and fertilizer (HYPONeX N-P-K = 6-10-5, HYPONeX) was
90 applied once every two weeks.
91 KW leaves were cultured on MS medium as above for four weeks to obtain adventitious shoots. If KWS plants
92 were chimeric with pink pigmentation in the L1 layer and white (*i.e.* no pigmentation) in the L2 layer and below,
93 all shoots originating from the L2 layer and below would be expected to display a white monochromatic phenotype.
94 Tissues, including the epidermal cell layer (L1) of KWS petioles were excised using a razor and forceps, leaving

95 only the subepidermal cell layer(s) (L2 or below). The resulting explants were cultured in MS medium for three
96 months to obtain adventitious shoots. All regenerated plants were grown as described above until they reached the
97 flowering stage.
98
99 **Flavonoid analysis**
100 To identify which stage of the anthocyanin biosynthetic pathway was interrupted in KW petals, the anthocyanin
101 and flavone compositions of KP and KW petals were measured by high-performance liquid chromatography
102 (HPLC). Petals of KP and KW were collected 2–3 days after flower opening and frozen at -80 °C. The frozen petals
103 were quickly crushed using a mortar and pestle at room temperature to avoid volume increases due absorption of
104 additional moisture from the air. The powdered petals were then extracted with ten-fold weight of 50% acetic acid-
105 methanol (acetic acid: methanol: distilled water = 1:4:5, v/v/v), and filtered through a 0.45 μm syringe filter
106 (SLNY1345NB, Nippon Genetics, Tokyo, Japan). The filtered solution was then subjected to HPLC for qualitative
107 and quantitative determination of flavonoids. Analytical HPLC was performed using a C18 (4.6 × 250 mm) column
108 (Waters, Milford, Massachusetts, USA) at 40 °C with a flow rate of 1 mL/min, monitored by a photodiode array
109 detector on an LC 10A system (Shimadzu Corporation, Kyoto, Japan). Linear gradient elution was performed using
110 an acetic acid-methanol mobile phase, from 20 to 85% solvent B (1.5% H_3PO_4, 20% HOAc, 25% MeCN in H_2O)
111 in solvent A (1.5% H_3PO_4 in H_2O) over 40 minutes, followed by a five-minute re-equilibration at 20% solvent B
112 for flavonoids (530 and 350 nm).
113 Absorbance at 530 nm (Abs_{530}) was measured using a Shimadzu UV-1800 spectrophotometer (Shimadzu, Japan)
114 for the anthocyanin extracts derived from the petals of KP, KW, ST, and SW. A calibration curve was constructed
115 from a five-fold dilution series of cyanidine-3-glucoside. Petal anthocyanin content was quantified as cyanidine-3-
116 glucoside equivalents using the calibration curve. The 5th to 7th expanded leaves of KP, KW, SH and m-SH were
117 collected to measure their anthocyanin content, then extracted and measured as described above.
118
119 **qRT-PCR and 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 to S4 (Fig. S1). 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 ConnectTM Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). Reaction conditions were 95 °C 10 sec and 60 °C 30 sec 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 day of flower opening. Pigmented and non-pigmented 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 sec, 60°C for 5 sec, and 68°C for 1 sec 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

144 RNA-seq was performed to analyze the transcriptome of KP and KW. Petals were collected from KP and KW
145 buds (S2) and frozen in liquid nitrogen. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Venlo,
146 The Netherlands) according to the manufacturer's instructions. RNA-seq library preparation and sequencing were
147 performed by the Rhelixa genomic and epigenomic research services company (Tokyo, Japan). RNA-seq libraries
148 were prepared using the NEBNext® Poly(A) mRNA Magnetic Isolation Module and the NEBNext® Ultra™ II
149 Directional RNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA). 150 bp paired-end sequencing was
150 performed on an Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA), yielding 4 Gb reads. Adaptor sequences
151 and low-quality reads were removed computationally using Trimmomatic v0.39 (Bolger *et al.*, 2014).
152 The *Saintpaulia* SIO_r2.0 genome data (derived from the KP genome, with 37,682 predicted genes. BioProject
153 PRJDB15710) was used as the reference sequence (Kurata et al., 2024). Mapping to SIO_r2.0 was done using
154 HISAT2 v2.2.1 (Kim *et al.*, 2019). Read count data and transcripts per million (TPM) values were calculated by
155 Stringtie v2.2.1 (Pertea *et al.*, 2015). Differentially expressed genes (DEG) analysis using edgeR v4.0.3 (Robinson
156 *et al.*, 2010) was performed using read count data with the TMM method (FDR = 0.05). Mapping results obtained
157 by RNA-seq analysis were visualized by pyGenomeTracks v3.8 (Lopez-Delisle *et al.*, 2021). De novo assembly of
158 transcript sequences was performed by Trinity v2.14.0 (Grabherr *et al.*, 2011) using RNA-seq data.
159
160 **Phylogenetic tree construction**
161 Molecular phylogenetic trees were constructed for MYB and bHLH transcription factors (TFs), which are
162 important for regulating the expression of anthocyanin biosynthesis genes (ABGs). Amino acid sequences were
163 filtered using PREQUAL v1.02 (Whelan *et al.*, 2018), and aligned using MAFFT v.7.490 (Katoh & Standley, 2013)
164 with default settings. Multiple aligned sequences were trimmed using trimAl v1.4.1 (Capella-Gutierrez *et al.*, 2009),
165 and used to generate a molecular phylogenetic tree by IQ-TREE2 v2.2.0.3 (Minh *et al.*, 2020). Internal branch
166 robustness was determined with Ultrafast bootstrapping (UF-boot) and SH-aLRT 1,000x each, and branches with ≥
167 95% and ≥ 80% confidence levels were considered robust (Minh *et al.*, 2013). The classification of MYB and bHLH

168 TFs was based on Stracke et al. (2001) and Heim (2003), respectively. The accession IDs of the amino acid
169 sequences are listed in Table S2.
170
171 **Whole-genome bisulfite sequencing (WGBS)**
172 Unstable pigmentation in petals and leaves may result from changes in the epigenomic state of genes in those
173 organs. Therefore, WGBS was used to identify differences in the methylated cytosine status of anthocyanin
174 biosynthesis-related genes. The first through fourth expanded leaves were collected from KP and KW plants. Leaves
175 were flash-frozen in liquid nitrogen and DNA was extracted using QIAGEN Genomic-tip 100/G (Qiagen). The
176 DNA was treated with bisulfite using an EZ DNA Methylation Gold Kit (Zymo Research, CA, USA). WGBS library
177 preparation and sequencing were performed by Rhelixa, Inc. WGBS libraries were prepared from bisulfite-treated
178 DNA using an Accel-NGS® Methyl-Seq DNA Library Kit (Swift Biosciences, MI, USA), and 150 bp paired-end
179 sequences were generated on an Illumina NovaSeq 6000 (Illumina), yielding 30 Gb reads. Adaptor sequences and
180 low-quality reads were removed using Trim Galore v0.6.7 (https://github.com/FelixKrueger/TrimGalore). The
181 *Saintpaulia* SIO_r2.0 genome was used as the reference sequence (Kurata et al., 2024). WGBS sequences were
182 mapped to SIO_r2.0 using BWAmeth v0.2.7 (arxiv.org/abs/1401.1129). Then, CG, CHG, and CHH methylation
183 contexts were extracted using MethylDackel v0.6.1 (https://github.com/dpryan79/MethylDackel). Methylation
184 levels were visualized by pyGenomeTracks v3.8 (Lopez-Delisle *et al.*, 2021).
185
186 **McrBC-PCR/qPCR**
187 *SiMYB2* methylation levels were compared between pigmented and non-pigmented tissues using McrBC, a
188 methylation-dependent restriction endonuclease. DNA was extracted from the first through fourth expanded leaves
189 of KP and KW using NucleoBond® HMW DNA (Takara, Shiga, Japan) and from the petals of KP and KW using a
190 FavorPrep™ Plant Genomic DNA Extraction Mini Kit (Favorgen, Ping Tung, Taiwan), and digested according to
191 the method described by Jaligot *et al*. (2014). Briefly, 100 ng of DNA was digested with 3 U McrBC (Takara) at
192 37 °C overnight. The DNA was then incubated at 65 °C for 20 min to stop the reaction and used as template for

193 PCR. Genomic DNA without McrBC treatment was used as a control. The primers used are listed in Table S1. PCR
194 and qPCR reactions were performed using KOD One® PCR Master Mix (TOYOBO) and SYBR THUNDER BIRD®
195 qPCR Mix (TOYOBO), respectively.
196
197 **Cloning of *SiMYB2-Long***
198 First-strand cDNA was synthesized from RNA extracted from KP petals using oligo dT primers and ReverTra
199 Ace® enzyme (TOYOBO). The full-length sequence of *SiMYB2-Long* was amplified using cDNA as a template
200 with Blend Taq® enzyme (TOYOBO). Primer sequences are listed in Table S1. Amplicons were ligated to pTAC-
201 1 using the DynaExpress TA PCR Cloning Kit (BioDynamics Laboratory, Tokyo, Japan), transformed into
202 *Escherichia coli* DH5α competent cells, and colonies carrying the target fragments were selected. Plasmids were
203 extracted from the selected colonies using FastGene™ PlasmidMini (NIPPON Genetics, Tokyo, Japan), and the
204 DNA sequences were confirmed with Sanger sequencing by AZENTA (Tokyo, Japan).
205
206 **CAGE-seq analysis**
207 Cap Analysis of Gene Expression (CAGE)-seq was performed to determine the transcription start site (TSS) of
208 *SiMYB2*. Three petal buds were collected at S1 from KP and KW plant lines and frozen in liquid nitrogen. Total
209 RNA was extracted using Sepasol®-RNA I Super G (Nacalai Tesque). CAGE library preparation, sequencing, and
210 mapping were performed by DNAFORM (Kanagawa, Japan). cDNA was synthesized from total RNA using random
211 primers. Ribose diols in the 5' cap structures of RNA molecules were oxidized and biotinylated. The biotinylated
212 RNA/cDNA complexes were then isolated using streptavidin beads (cap-trapping method). Following RNA
213 digestion with RNaseONE/H and adaptor ligation at both cDNA ends, double-stranded cDNA libraries (CAGE
214 libraries) were generated and sequenced using 75 nt single-end reads (NextSeq 500, Illumina). The resulting reads
215 were mapped to the *Saintpaulia* SIO_r2.0 genome using STAR v2.7.9a (Dobin *et al.*, 2013).
216
217 **Agroinfiltration assays**

218 *Nicotiana tabacum* was grown for transient assay by sowing seeds in Kumiai Nippi Gardening Soil No. 1 (Nihon
219 Hiryo, Tokyo, Japan) and germinating them under 20 °C, 14 h light/10 h dark conditions. The germinated seedlings
220 were transferred to 220 ml plastic pots filled with the same soil and grown under the same conditions for 4–5 weeks.
221 Four distinct vectors (*SiMYB1*, *SiMYB2-Long*, *SiMYB2-Short*, *SibHLH1*) were designed for transient assays.
222 Vectors were constructed by artificial synthesis using VectorBuilder (VectorBuilder Japan, Kanagawa, Japan) or
223 the in-fusion method. Each vector was designed to have the CDS of each TF driven by the CaMV35S promoter and
224 the kanamycin marker gene NPT II. The VectorBuilder ID of each vector is listed: *SiMYB1* ID; VB210520-1064sfu,
225 *SiMYB2-Long* ID; VB210520-1068fhv, *SibHLH1* ID; VB210520-1053zmv.
226 The pBI121 vector was digested with *Bam*HI (Takara) and *Sac*I (Takara) to remove the *GUS* coding region. The
227 CDS fragment of *SiMYB2-Short* was amplified using cDNA obtained by reverse transcription of petal RNA. CDS
228 fragments were then amplified using primers for the in-fusion system (Table S2) (In-Fusion® HD Cloning Kit,
229 Takara Bio USA, CA, USA) by combining pBI121 digested fragments and *SiMYB2-Short* amplicons. In-fusion
230 cloning was performed according to the manufacturer's protocol.
231 Vector constructs were introduced into Agrobacterium (*Rhizobium radiobactor*) strain EHA105. Transformed
232 Agrobacterium were grown on LB medium (0.5% Yeast Extract, 1% Tryptone, 1% NaCl, 100 ppm Kanamycin,
233 100 ppm Rifampicin) for two days, and a single colony was selected for the transient assay. Vectors carrying *Green*
234 *fluorescence protein* (*GFP*) were introduced into Agrobacterium and used as negative controls.
235 Agroinfiltration was performed with a slight modification of the method of Outchkourov et al. (2014).
236 Transformed Agrobacterium were suspended in liquid YEP medium amended with 50 ppm kanamycin and
237 incubated overnight at 28 °C with shaking. The culture medium was centrifuged (3,500g, 10 min), pellets were
238 resuspended in infiltration buffer [10 mM MES-NaOH (pH = 5.7), 10 mM $MgCl_2$, 150 μM acetosyringone], and
239 adjusted to OD_{600} = 0.4–0.6. The mixture was incubated for 2–3 hours at room temperature. The mixture was then
240 injected into 4–5-week-old tobacco leaves. Leaves were photographed, collected and stored frozen at -80 °C three
241 days post-infiltration. The experiment included three biological replicates of two leaves per tobacco plant, and the

242 experiment was repeated in triplicate. Anthocyanin content was then quantified according to the procedure described
243 above.
244 Leaf discs were collected from the injection site using a biopsy punch (BP-60F, 6 mm dia., Kai Industries, Gifu,
245 Japan), avoiding the puncture point of the syringe, and frozen in liquid nitrogen. qRT-PCR was performed according
246 to the procedure described above using primers listed in Table S1. *NtActin* was used as the reference gene.
247
248 ***SiMYB2* genome structure**
249 The selection of *SiMYB2* variants was investigated in several cultivars (AK, TK, GG, BB, MT, TR, SA, ST, KP)
250 using RT-PCR. RNA extraction and RT-PCR were performed as described above. Reactions were performed at
251 98°C for 10 sec, 60°C for 5 sec, and 68°C for 1 sec for 40 cycles and amplified to a plateau. Primers were the same
252 as those used for qRT-PCR (Table S1). *SiActin* was used as the internal control.
253 The genomic structure of *SiMYB2* was analyzed through Amplicon-seq using multiple lines and cultivars (AK,
254 TK, GG, BB, MT, KP). First, genomic DNA was extracted from the leaves of these cultivars using the FavorPrep™
255 Plant Genomic DNA Extraction Mini Kit (Favorgen). Primers were designed to amplify the ORF of *SiMYB2* (Table
256 S1). PCR reactions were performed using KOD One® PCR Master Mix (TOYOBO) under the following conditions:
257 98°C for 10 sec, and 68°C for 185 sec for 40 cycles. Subsequently, a library was prepared with the Native Barcoding
258 Kit 24 V14 (SQK-NBD114.24, Oxford Nanopore Technologies, Didcot, UK) and sequenced on a FLO-MIN114
259 flow cell (Oxford Nanopore Technologies).
260 Based on the results of Amplicon-seq, primers were designed to detect insertions or deletions (In/Dels) within
261 the *SiMYB2* locus (Table S1). Genomic DNA was extracted from the leaves of ST, SA, SH, and m-SH in addition
262 to the cultivars listed above using a FavorPrep™ Plant Genomic DNA Extraction Mini Kit (Favorgen). PCR
263 reactions were performed using KOD One® PCR Master Mix (TOYOBO). Reactions were performed at 98°C for
264 10 sec, 60°C for 10 sec, and 68 °C for 3 sec for 40 cycles.
265 The expression levels of *SiMYB2-Long* and *SiMYB2-Short* in the leaves of SH and m-SH were investigated by
266 qRT-PCR. RNA extraction and qRT-PCR were performed as described above.

267

268 **Statistical analyses**

269 All statistical processing was performed using R software (ver. 4.3.2, https://www.R-project.org/). Student's *t*-
270 test was performed using the 'stats' package, and Dunnett's range test was performed using the 'multcomp' package.

271

272 **Results**

273 **Variations in anthocyanin accumulation as a result of tissue culture**

274 A total of 495 regenerated plants were obtained from KP leaf adventitious shoots, of which 96.6% (484/495)
275 produced pink petals and reddish leaves, the same phenotype as KP (Table 1). 1.0% (5/495) did not accumulate
276 anthocyanin and therefore had white petals and green leaves (KW). 0.8% (4/495) had randomly-colored petals, and
277 0.4% (2/495) had white-striped petals (KWS) (Fig. 1, Table 1). As the five KW plants were grown, two of them
278 began to produce white striped petals like KWS, and the remaining three had randomly-pigmented petals (Fig. 1h,
279 j). In addition, five KWs and two KWSs eventually grew petals and leaves like the parental KP (Fig. 1k). Fifty
280 plants were regenerated from KW leaves, all of which had white petals and green leaves (50/50, Table 1). As these
281 plants grew, some began producing striped or wholly-pigmented petals like those of KWS or KP, respectively.

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

290

291 **Loss of anthocyanin biosynthesis in white petals and green leaves**

292 Pelargonidin-3-acetylrutinose-5-glucoside (A1, Fig. S2a) and peonidin-3-acetylrutinose-5-glucoside (A2, Fig.
293 S2a) were readily detected in KP petals, but were at low levels in KW petals (Fig. S2b). While apigenin-4'-
294 glucuronide (F1, Fig. S2e) was identified as a flavone in KP petals, it was hardly detected in KW petals (Fig. S2f).
295 In addition, cyanidin-3-acetylsambubioside (A3, Fig. S2c) and luteolin-4'-glucuronide (F2, Fig. S2g) were detected
296 in KP leaves, but were not detected in KW leaves (Fig. S2d,h). Thus, the functions of *Chalcone synthase* (*CHS*),
297 *Chalcone isomerase* (*CHI*), *Flavone synthase* (*FNS*), or all of them, may be disturbed in KW petals and leaves (Fig.
298 2a). Furthermore, total anthocyanin content was significantly lower in KW than KP petals and in SW than ST petals
299 (Fig. 2a), and the anthocyanin content of KW leaves was lower than that of KP leaves (Fig. 2a).
300 To investigate at which stage anthocyanin metabolism is interrupted, the expression levels of ABGs, *CHS*
301 (*SiCHS1/2*), *CHI* (*SiCHI1*), and *FNS* (*SiFNS1*) in KP and KW petals were compared by qRT-PCR. The expression
302 of *SiCHS1/2*, *SiCHI1,* and *SiFNS1* was significantly lower in KW petals than in KP petals (Fig. 2b). The expression
303 levels of *SiF3H1*, *SiDFR1*, *SiANS1/2*, *SiGSTF1,* and *Si3GT1* were also significantly lower in KW petals than in KP
304 petals (Fig. 2b). Consistent with the qRT-PCR results, RNA-seq showed that the expression levels of several ABGs
305 (*SiCHS1/2*, *SiCHI1*, *SiF3H1*, *SiDFR*, *SiANS1/2*, *SiGSTF1*, *Si3GT1/2*, *SiMATE1/2/3*, etc.) were reduced in KW
306 petals relative to KP petals (Table 2). SW petals, which have similar characteristics to KW, have several ABG
307 expression levels significantly lower than in ST petals (Fig. 2c). In addition, the expression of several ABGs
308 (*SiCHS1/2*, *SiCHI1*, *SiF3H1*, *SiDFR1*, and *SiANS1/2*) was lower in the white regions of KWS and SWS petals than
309 in their pigmented regions (Fig. S3a, b). There were no significant differences in the expression levels of *SiCHS1*,
310 *SiCHI1*, *SiF3H1,* and *SiFNS1* between KP and KW leaves (Fig. 2d), but the expression levels of *SiCHS2*, *SiDFR1*,
311 *SiANS1/2*, *SiGSTF1*, and *Si3GT1* were significantly lower in KW leaves than in KP leaves (Fig. 2d). Based on qRT-
312 PCR and RNA-seq results, the low anthocyanin accumulation levels in white petals and green leaves is likely due
313 to the downregulation of several ABGs.
314
315 **Downregulation of anthocyanin biosynthesis-related transcription factors**

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

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

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

336 A molecular phylogenetic tree was constructed to model the evolutionary history of MYB and bHLH TFs in
337 *Saintpaulia*. The tree indicates that *SiMYB1* groups within the SG 6 clade (Stracke *et al.*, 2001), which is known to
338 promote anthocyanin biosynthesis (Fig. 3e). Although *SiMYB2* was highly expressed in non-pigmented tissues, and
339 was hypothesized to belong to a clade associated with anthocyanin biosynthesis repression, it was instead clustered

340 within the SG 6 clade (Fig. 3e). Both *SibHLH1* (IIIf b) and *SibHLH2* (IIIf a) belonged to clades known to promote
341 anthocyanin biosynthesis (Heim, 2003) (Fig. 3f).
342
343 **Hypermethylation of *SiMYB2* in anthocyanin-deficient tissues**
344 From the expression analysis results, *SiMYB2* and *SibHLH2* were identified as candidate genes involved in the
345 unstable pigmentation of *Saintpaulia* petals and leaves. The hypothesis that loss and reversion of anthocyanin
346 accumulation in flowers and leaves was associated with changes in the epigenome led us to consider that the
347 epigenetic state of either *SiMYB2* or *SibHLH2* may have undergone modification. Therefore, WGBS (whole genome
348 bisulfite sequence) was conducted to examine the methylation levels of these candidate genes in the leaves of KP
349 and KW. Differences in methylation levels were observed in the second exon of *SiMYB2* (Fig. 4a). In particular,
350 methylation levels in the CHG and CHH contexts were higher in KW than in KP (Fig. 4a). Additionally, CHH
351 context methylation was higher in KW in the promoter region of *SiMYB2* than in KP (Fig. 4a). For *SibHLH2*, there
352 were no differences in methylation levels in either the ORF or promoter regions (Fig. 4b). The methylation levels
353 of several other genes related to anthocyanin biosynthesis (*SiCHS1/2*, *SiCHI1*, *SiF3H1*, *SiDFR1*, *SiANS1/2*,
354 *SiGSTF1*, *Si3GT1/2*, *Si3GRT1*, *SiMATE1/2/3*, *SiAT*, *SiF3'5'H*, *SiMYB1*, and *SibHLH1*) were also examined, but no
355 differences were detected in the coding gene itself or its promoter regions (Fig. S4). McrBC-PCR/qPCR results also
356 showed a difference in methylation levels in the second exon of *SiMYB2* between KP and KW leaves (Fig. 4c, f).
357 This difference was also observed in KP and KW petals (Fig. 4d–f). Thus, *SiMYB2* was identified as an epiallele
358 associated with unstable pigment accumulation.
359
360 **Two transcripts selectively generated from the *SiMYB2* locus**
361 *SiMYB2* (Sio_r2.0_p0007.1.g08820.1) was significantly up-regulated and accompanied by hypermethylation in
362 anthocyanin-deficient tissues, suggesting its involvement in anthocyanin accumulation instability. Mapping of the
363 *SiMYB2* genomic region showed a splice junction between *SiMYB2* and Sio_r2.0_p0007.1.g08810.1 in KP (Fig. 5a;
364 sj2), suggesting that the *SiMYB2* gene region extends from Sio_r2.0_p0007.1.g08810.1 to

365 Sio_r2.0_p0007.1.g08820.1. By Sanger sequencing of the TA cloning products and *de novo* assembly of RNA-seq
366 reads, transcripts consisted of p1 + p2 + p4 and those of p3 + p4 were identified (Dataset S1). This result indicates
367 that the *SiMYB2* locus is a single gene that ranges over approximately 18 kb, and produces at least two RNAs,
368 presumably from two different transcription start sites (Fig. 5b, c). Exon 1 (124 bp), exon 2 (130 bp), exon 3 (110
369 bp), and exon 4 (493 bp) were determined from the sequences of p1, p2, p3, and p4 (Fig. 5b). The two 5' UTRs (1
370 and 2) and one 3' UTR were estimated. Transcripts consisting of 5' UTR 1, exon 1, exon 2, exon 4, and 3' UTR were
371 designated as *SiMYB2-Long*, and 5' UTR 2, exon 3, exon 4, and 3' UTR as *SiMYB2-Short* (Fig. 5c). *SiMYB2-Long*
372 was preferentially generated in KP and *SiMYB2-Short* in KW.

373 The expression levels of *SiMYB2-Long* and *SiMYB2-Short* were confirmed by qRT-PCR. *SiMYB2-Long*
374 expression levels were significantly lower in KW than in KP petals, and in SW than in ST petals (Fig. 5d). In
375 contrast, *SiMYB2-Short* expression levels were significantly higher in KW than in KP petals, and in SW than in ST
376 petals (Fig. 5d). This expression pattern for *SiMYB2-Long* and *SiMYB2-Short* was also observed in KP and KW
377 leaves (Fig. 5d). The expression levels of *SiMYB2-Long* were significantly higher in KP compared to KW leaves,
378 whereas the expression levels of *SiMYB2-Short* were significantly higher in KW than in KP leaves (Fig. 5d). The
379 expression levels of *SiMYB2-Long* and *SiMYB2-Short* in KP and KW petals were also examined at different
380 developmental stages. Anthocyanin accumulation was detected between S2–S3, with the expression levels of
381 *SiMYB2-Long* being significantly higher at S2 in KP petals (Fig. 5e). In KW petals, the expression level was lower
382 at all stages (Fig. 5e). *SiMYB2-Short* was hardly expressed at any stage in KP petals, while it was highly expressed
383 in KW petals (Fig. 5e). Expression levels of *SiMYB2-Short* in the pigmented regions of striped flowers was
384 significantly lower than in the non-pigmented regions of KWS and SWS petals (Fig. 5f), indicating that *SiMYB2-*
385 *Long* is specifically expressed in colored tissues, whereas *SiMYB2-Short* is selectively transcribed in non-colored
386 tissues.

387 *SiMYB2-Long* and *SiMYB2-Short* have different 5' UTR sequences, suggesting that these variants arise from
388 Alternative Promoter Usage (APU, Landry *et al.*, 2003). CAGE-seq analysis was performed to determine the
389 transcription start sites (TSSs) of *SiMYB2* variants. If the two variants were produced via APU, peaks would be

390 expected to detect the upstream regions of exon 1 and exon 3 of *SiMYB2*, respectively. However, only a single TSS
391 was detected upstream of exon 1 in both KP and KW samples (Fig. 5a). Consequently, the mechanism by which
392 these two transcription variants are generated could not be determined in this study. Note that the upstream sequence
393 of *SiMYB2-Short* contains a TA-rich region with several potential TATA boxes. (Fig. S5, Mukumoto et al., 1993).
394

395 **Functional analysis of *SiMYB2* by transient assays**

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

402 The functions of SibHLH1, SiMYB1, SiMYB2-Long and SiMYB2-Short were examined by transient assays in
403 tobacco leaves. Active expression of the gene of interest was confirmed by qRT-PCR (Fig. S6). Expression of
404 *SiMYB2-Long* alone resulted in very little accumulation of anthocyanins (Fig. 6b). When *SibHLH1*, *SiMYB1*, or
405 S*iMYB2-Short* were expressed individually, no anthocyanin accumulation was observed (Fig. 6b). Subsequently,
406 *SibHLH1* was selected as a potential cofactor for *SiMYB1* and *SiMYB2*, as its expression levels did not differ
407 between colored and non-colored tissues. Anthocyanin accumulated when *SiMYB1* plus *SibHLH1*, or *SiMYB2-Long*
408 plus *SibHLH1* were expressed (Fig. 6c). There was no accumulation of anthocyanin in tissues where *SiMYB2-Short*
409 was co-expressed with *SibHLH1* (Fig. 6c). Anthocyanin content was also measured when MYB and bHLH were
410 simultaneously expressed. The anthocyanin content of *SiMYB2-Long* co-expressed with *SibHLH1* was the highest
411 among all transient assays, while anthocyanin content of tissues co-expressing *SiMYB2-Short* and *SibHLH1* was not
412 different from the GFP control (Fig. 6d). The anthocyanin content of *SiMYB1* plus *SibHLH1* was lower than that of
413 *SiMYB2-Long* plus *SibHLH1*, as reflected in its fainter coloring (Fig. 6d). The expression levels of ABGs were
414 confirmed when these TFs were expressed. Several ABGs were upregulated in *SiMYB2-Long* only, *SiMYB1* plus

415 *SibHLH1*, and *SiMYB2-Long* plus *SibHLH1*, whereas there was no upregulation of ABGs in S*iMYB2-Short* plus
416 *SibHLH1* co-expression assays (Fig. 6e). Compared to *SiMYB1* plus *SibHLH1*, *SiMYB2-Long* plus *SibHLH1* had
417 higher ABG expression, which is consistent with its anthocyanin content (Fig. 6e).
418 To investigate the possibility that SiMYB2-Short, whose expression is higher in non-pigmented tissues, indirectly
419 inhibits R2R3-MYB function in a dosage-dependent manor, similar to CPC type R3-MYB (Li *et al.*, 2020), bacterial
420 suspensions expressing gradually increasing ratios of SiMYB2-Short to SiMYB1 and SibHLH1 were injected into
421 tobacco leaves. Anthocyanin accumulation was not affected by changing the ratio of SiMYB2-Short (Fig. S7).
422
423 **The origin of *SiMYB2-Short* is the insertion of a transposon-like sequence**
424 RT-PCR was performed to investigate whether the transcriptional selection of *SiMYB2-Long* and *SiMYB2-Short*
425 also occurs in other cultivars. Two *SiMYB2* variants were found to be transcribed in TR and SA, which are similar
426 to KP and ST (Fig. S8j). In contrast, *SiMYB2-Long* was amplified in AK, TK, GG, BB, and MT, but *SiMYB2-Short*
427 was not (Fig. S8j). TR and SA, like KP and ST, are cultivars in which anthocyanin accumulation is unstable,
428 showing loss and reversion during tissue culture propagation (Fig. S8f-i). AK, TK, GG, BB, and MT are cultivars
429 with consistent anthocyanin accumulation in their leaves and flowers (Fig. S8a-e).
430 The genomic structure of *SiMYB2* was examined in AK, TK, GG, BB and MT using Amplicon-seq, which showed
431 that there is a ~2.5 kb deletion, including exon 3, which is required for the expression of *SiMYB2-Short*, in all five
432 cultivars (Fig. 7a, S7k). To validate this finding, genomic PCR was performed using primers designed to flank the
433 deletion region (Fig. 7a). As with the Amplicon-seq results, only the deletion-type band was amplified in AK, TK,
434 GG, BB, and MT (Fig. 7b). In KP, only the insertion-type band was amplified, whereas in ST, TR, and SA, both
435 insertion- and deletion-type bands were detected (Fig. 7b). These results indicate that the absence of *SiMYB2-Short*
436 expression in AK, TK, GG, BB, and MT is due to deletion.
437 The deleted region DNA sequence contains features of a transposon-like element (TE-like), including terminal
438 inverted repeats (TIRs) (Fig. 7a,c). The 5' TIR was 171 bp, the 3' TIR was 170 bp, and the target site duplication
439 (TSD) was 10 bp. A BLAST search for homologous transposon sequences failed to reveal any highly similar

440 matches. Using the TIR sequences as queries, a BLAST search against the *Saintpaulia* SIO_r2.0 genome identified
441 numerous homologous sequences (Table S3), suggesting that *SiMYB2-Short* may have arisen due to the insertion
442 of an endogenous TE-like element.

443 Like KW and SW, SH is a cultivar that does not accumulate anthocyanins in its leaves. When SH was propagated
444 via tissue culture, the resulting plants also did not accumulate anthocyanin in their leaves, like the original SH
445 phenotype (Fig. 7d), and flowers of regenerated plants are green and do not accumulate anthocyanins (Fig. 7e).
446 However, a single mutant, referred to as m-SH, emerged with red-tinted leaves and flowers (1 of 50 regenerated
447 plants. Fig. 7d,e). The anthocyanin content of m-SH leaves was significantly higher than in wild type SH leaves
448 (Fig. 7f). The expression of *SiMYB2-Long* was barely detectable in SH, but was highly expressed in m-SH (Fig. 7g).
449 *SiMYB2-Short* was expressed in SH but was not detected in m-SH (Fig. 7g).

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

454

455 **Discussion**

456 In this study, we identified an epiallele involved in unstable pigmentation patterning in *Saintpaulia* plants.
457 Anthocyanins did not accumulate, and the expression levels of several ABGs were low in non-pigmented tissues.
458 RNA-seq analysis was performed to determine candidate TFs, and a MYB TF, *SiMYB2*, was identified as an
459 epiallele through WGBS analysis. Two transcripts were generated from the *SiMYB2* locus, one of which was highly
460 expressed in pigmented colored tissues, and the other in non-pigmented tissues. These transcripts were designated
461 as *SiMYB2-Long* and *SiMYB2-Short*, respectively, reflecting their relative lengths. Transient assays showed that
462 *SiMYB2-Long* is an anthocyanin biosynthesis activator. *SiMYB2-Short* was highly expressed in non-pigmented
463 tissues, suggesting that it is a repressor-type MYB TF. Paradoxically, the transient assay results also revealed that
464 *SiMYB2-Short* does not have either a positive or negative impact on anthocyanin expression. Sequence analysis of

465 the *SiMYB2* gene among several cultivars suggested that *SiMYB2-Short* resulted from the insertion of a transposon-
466 like sequence. These observations provide a compelling model to help explain why mutants generated by tissue
467 culture from an anthocyanin-deficient cultivar in which *SiMYB2-Short* is expressed accumulated anthocyanins.
468
469 **Formation of white striped petals due to periclinal chimeras versus epigenetics**
470 The white striped flowers produced by *Saintpaulia* are thought to be the result of a periclinal chimeric structure
471 (Lineberger & Druckenbrod, 1985; Kazemian *et al.*, 2019). However, there is currently no evidence that they are
472 due to the peripheral chimera structure. In this study, the tissue culture progeny of *Saintpaulia* KP and KWS flowers
473 varied in their petal and leaf color patterns (Fig. 1a, Table 1). A previous study reported that the phenotypic variation
474 of petal color varied greatly in plants regenerated from the tissue culturing of cultivar 'Kilauea' (Yang *et al.*, 2017).
475 The characteristic of *Saintpaulia* phenotype that changes in tissue culture and then reverts to the original phenotype
476 suggests that this phenotype is the result of changes in epigenetic regulation.
477 White-striped petals closely resemble striped petals caused by periclinal chimeras, which may lead to the
478 misconception that they too are periclinal chimeras. Periclinal chimeras occur when the cell layers of the shoot
479 apical meristem (SAM) have distinct genetic backgrounds and coexist within a single plant (Satina, 1940). When
480 such genetic differences involve anthocyanin biosynthesis-related genes, *Saintpaulia* flowers can exhibit striped
481 petals. In this study, *Saintpaulia* with white striped petals eventually flowered with fully colored petals. Moreover,
482 white-striped petals arose post-developmentally from *Saintpaulia* plants which had initially exhibited entirely white
483 petals. Thus, it is unlikely that white-striped petals are due to periclinal chimeras. How, then, are the white-striped
484 petals formed? In white-striped petals, the presence or absence of *SiMYB2-Short* expression correlated to lack of
485 pigmentation (Fig. 5g). Additionally, the transcriptional selection of *SiMYB2* variants appears to be regulated
486 epigenetically, suggesting that white-striped petals may result from cell layers with distinct epigenetic backgrounds.
487 In other words, it may be an epigenetic periclinal chimera. Indeed, genes with cell layer-specific expression have
488 been reported in *A. thaliana* (Lu *et al.*, 1996). The epigenomic states of the L2 layer or below may be altered by

489 tissue culture (Fig. 8). In white-striped petals, the epigenomic states of different cell layers in SAM may be
490 independently regulated, allowing these distinct states to coexist within the same tissue (Fig. 8).
491
492 **Identification of the epiallele linked to unstable pigmentation**
493 The traits of plants obtained through tissue culture or vegetative propagation are not always consistent. Such
494 variations are occasionally attributable to alterations in epigenetic status. For example, during the tissue culture of
495 *E. guineensis*, a reduction in the methylation level of the *Karma* transposable element within the *Deficiens1* (*DEF1*)
496 gene leads to splicing errors in the *DEF1* gene, resulting in malformed fruit (Ong-Abdullah *et al.*, 2015). In *C.*
497 *morifolium*, yellow-flowered plants arise from the lateral buds of red-flowered plants. This phenomenon is caused
498 by hypermethylation in the promoter region of *CmMYB6* (Tang *et al.*, 2022). In our case, tissue culture may have
499 altered the epigenetic state of *Saintpaulia*, leading to phenotypic variations.
500 Based on the hypothesis that unstable anthocyanin accumulation in *Saintpaulia* is caused by epigenetic regulation,
501 we determined that *SiMYB2* is the epiallele associated with coloration by a combination of RNA-seq and WGBS
502 analyses. *CmMYB6* in *C. morifolium* was identified as an epiallele involved in flower color changes using a similar
503 approach as in the present study (Tang *et al.*, 2022). *EgDEF1* in *E. guineensis* was also identified by epigenome-
504 wide association analysis (EWAS) and WGBS (Ong-Abdullah *et al.*, 2015). Exploration of other epialleles that
505 control unstable traits, such as unstable anthocyanin accumulation, by this combination of methods may yield
506 additional useful candidates for the identification of causal genes. *Saintpaulia* has been propagated and maintained
507 in large quantities by vegetative propagation such as leaf cuttings and tissue culture. Epigenetic mutants, such as
508 the anthocyanin-depleted individuals in this study, are often produced by vegetative propagation that are maintained
509 until the next generation. For example, variation in transposon transposition frequency and altered floral patterns
510 have been reported in tissue culture progeny of *Saintpaulia* 'Thamires' (Sato *et al.*, 2011). It is also postulated that
511 these traits are the result of epi-mutation. It may therefore be feasible to utilize the epigenetic traits observed in
512 *Saintpaulia* for genetic identification purposes.

513 The present study identified *SiMYB1* and *SibHLH2* as TFs associated with anthocyanin accumulation in petals,
514 in addition to *SiMYB2* (Fig. 3). The transient assay results demonstrated that *SiMYB1* functions as an anthocyanin
515 biosynthesis activator (Fig. 6c-e). Therefore, *SiMYB1* expression may be the determinant of unstable pigmentation.
516 However, *SiMYB1* was barely expressed in leaves exhibiting unstable pigment accumulation like petals (Fig. 3d).
517 In contrast, the transcriptional selectivity of *SiMYB2-Long* and *SiMYB2-Shor*t was linked to pigment accumulation
518 in the leaves (Fig. 3c, 7f). These results indicate that *SiMYB1* is not the gene responsible for the phenomenon of
519 unstable pigmentation.

520 The expression levels of *SibHLH1*, which is thought to regulate the transcription of ABGs, were essentially the
521 same in white and colored petals. Phylogeny placed *SibHLH1* in the IIIf b clade, which includes *PhJAF13* (Fig. 3f.
522 Quattrocchio et al., 1998; Heim, 2003). In Petunia, the expression of *PhJAF13* is not influenced by the expression
523 of MYB TFs (Quattrocchio *et al.*, 1998; Spelt *et al.*, 2000), which is consistent with our observation that the
524 expression of *SibHLH1* was unchanged in the white parts of *Saintpaulia*, where the expression of the MYB TFs
525 was downregulated. *SibHLH2* clustered with the IIIf a clade, which includes *PhAN1* (Fig. 3f. Spelt et al., 2000;
526 Heim, 2003). The loss of *PhAN1* function results in the loss of anthocyanin biosynthesis in Petunia (Spelt *et al.*,
527 2000). Thus, the reduction of *SibHLH2* expression may contribute to the loss of anthocyanin in *Saintpaulia*. It has
528 been reported that *NtJAF13*, a member of the IIIf b clade, interacts with MYB TFs and promotes the expression of
529 *NtAN1*, a member of the IIIf a clade in *N. tabacum* (Montefiori *et al.*, 2015). Moreover, *AmDelila*, a member of
530 clade IIIf b, binds to the *Rosea1* MYB TF and positively regulates the transcription of *AmIncolorata I*, a member
531 of clade IIIf a, in *Antirrhinum majus* (Albert *et al.*, 2021). Therefore, the downregulation of *SibHLH2* in the non-
532 pigmented tissues is most likely due to decreased *SiMYB2-Long* expression levels. Moreover, *SiMYB2* methylation
533 levels differed between KP and KW, while *SibHLH2* did not (Fig. 4), suggesting that differences in *SibHLH2*
534 expression are not the underlying cause of unstable pigmentation.

535

536 **Transcriptional selectivity of *SiMYB2* variants and pattern formation**

537 Alternative splicing (AS) and APU are two regulatory mechanisms by which different transcripts can be produced
538 from a single locus (Reviewed in Landry et al., 2003). An example of unstable pigmentation caused by different
539 variants has been reported in *C. morifolium*. AS of *CmbHLH2* produces different variants, with non-functional
540 *CmbHLH2* in white petals and functional *CmbHLH2* in colored petals (Lim *et al.*, 2021). However, the present work
541 is the first report that two or more transcripts from a single locus are involved in flower pattern formation. It was
542 thought that *SiMYB2* variants were generated by APU since the 5' UTRs of *SiMYB2-Long* and *SiMYB2-Short* were
543 unique to each, particularly since multiple TATA boxes were observed in the predicted promoter region of *SiMYB2-*
544 *Short*. However, we were unable to detect the TSS of *SiMYB2-Short* by CAGE-seq, thus preventing a conclusive
545 determination about whether *SiMYB2* variants are generated through APU. CAGE-seq specifically sequences the 5'
546 ends of RNAs with a 5' cap structure (m7G). Therefore, the absence of a detectable TSS for *SiMYB2-Short* may be
547 attributed to the lack of a 5' cap structure in *SiMYB2-Short* transcripts. Although this study could not identify the
548 precise mechanism by which *SiMYB2* variants are generated, the findings from CAGE-seq provide valuable insights
549 for future *SiMYB2* variant studies.

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

559 The specific function of *SiMYB2-Short* could not be identified in this study, so further research is necessary to
560 determine its role. If stable anthocyanin accumulation is desired, deleting the TE-like sequence may be useful. On
561 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*.

Acknowledgments

We are grateful to Dr. Atsushi Hoshino of the National Institute for Basic Biology for his guidance 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 his invaluable guidance on the subject of epigenetics and for her assistance with cultivation. This work partly was supported by Sasakawa Scientific Research Grant from the Japan Science Society. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.

Competing Interest Statement

The authors have declared no competing interest.

Author contributions

DK: Data curation, Investigation, Resources, Software, Visualization, Formal analysis, Conceptualization, Project administration, Writing-original draft. TT: Data curation, Investigation, Resources KS: Funding acquisition, Data curation, Formal analysis, Writing-review & editing. HH: Funding acquisition, Data curation, Formal analysis, Writhing-review & editing. FT: Data curation, Investigation, Writhing-review & editing. MH: Funding acquisition, Resources, Supervision, Project administration, Conceptualization, Writing-review & editing.

587 **References**

588 **Albert NW, Butelli E, Moss SMA, Piazza P, Waite CN, Schwinn KE, Davies KM, Martin C. 2021. Discrete**
589 **bHLH transcription factors play functionally overlapping roles in pigmentation patterning in flowers**
590 **of *Antirrhinum majus*. *New Phytologist* 231: 849–863.**

591 **Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.**
592 ***Bioinformatics* 30: 2114–2120.**

593 **Broertjes C, Keen A. 1980. Adventitious shoots: Do they develop from one cell? *Euphytica* 29: 73–87.**

594 **Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming**
595 **in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.**

596 **Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013.**
597 **STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.**

598 **Eyerdom H. 1981. Flower color sports and variations in Saintpaulia hybrids. *African Violet Magazine*: 33–**
599 **37.**

600 **Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury**
601 **R, Zeng Q, *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference**
602 **genome. *Nature Biotechnology* 29: 644–652.**

603 **Heim MA. 2003. The basic helix-loop-helix transcription factor family in plants: A genome-wide study of**
604 **protein structure and functional diversity. *Molecular Biology and Evolution* 20: 735–747.**

605 **Jaligot E, Hooi WY, Debladis E, Richaud F, Beulé T, Collin M, Agbessi MDT, Sabot F, Garsmeur O, D'Hont**
606 **A, *et al.* 2014. DNA methylation and expression of the *EgDEF1* gene and neighboring**

607 **retrotransposons in *mantled* somaclonal variants of oil palm (M Bendahmane, Ed.). *PLoS ONE* 9:**
608 **e91896.**

609 **Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in**
610 **performance and usability. *Molecular Biology and Evolution* 30: 772–780.**

611 **Kazemian M, Ghasemi omran V, Mohajel Kazemi E, Kolahi M. 2019. Regeneration of pinwheel phenotype**
612 **and evaluation of anthocyanin in African violet (*Saintpaulia ionantha* Wendl.) periclinal chimera.**
613 ***Journal of Plant Molecular Breeding* 7: 39–49.**

614 **Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping**
615 **with HISAT2 and HISAT-genotype. *Nature Biotechnology* 37: 907–915.**

616 **Kurata D, Fukutomi K, Kubo K, Shirasawa K, Hirakawa H, Hosokawa M. 2024. Comprehensive expression**
617 **analysis of ERF transcription factors during chilling acclimation in Saintpaulia. *Plant Growth***
618 ***Regulation*.**

619 **Landry J-R, Mager DL, Wilhelm BT. 2003. Complex controls: the role of alternative promoters in**
620 **mammalian genomes. *Trends in Genetics* 19: 640–648.**

621 **Li Y, Shan X, Gao R, Han T, Zhang J, Wang Y, Kimani S, Wang L, Gao X. 2020. MYB repressors and MBW**
622 **activation complex collaborate to fine-tune flower coloration in *Freesia hybrida*. *Communications***
623 ***Biology* 3: 396.**

624 **Lim S-H, Kim D-H, Jung J-A, Lee J-Y. 2021. Alternative splicing of the basic Helix–Loop–Helix transcription**
625 **factor gene *CmbHLH2* affects anthocyanin biosynthesis in ray florets of chrysanthemum**
626 **(*Chrysanthemum morifolium*). *Frontiers in Plant Science* 12: 669315.**

627 **Lineberger RD, Druckenbrod M. 1985. Chimeral nature of the pinwheel flowering african violets**
628 **(Saintpaulia, Gesneriaceae). *American Journal of Botany* 72: 1204–1212.**

629 **Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, Ramírez F, Manke T. 2021.**
630 **pyGenomeTracks: reproducible plots for multivariate genomic datasets (R Peter, Ed.). *Bioinformatics***
631 **37: 422–423.**

632 **Lu P, Porat R, Nadeau JA, O'Neil SD. 1996. Identification of a meristem L1 layer-specific gene in Arabidopsis**
633 **that is expressed during embryonic pattern formation and defines a new class of Homeobox genes. 8:**
634 **2155–2168.**

635 **Minh BQ, Nguyen MAT, Von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap.**
636 ***Molecular Biology and Evolution* 30: 1188–1195.**

637 **Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-**
638 **TREE 2: New models and efficient methods for phylogenetic inference in the genomic era (E Teeling,**
639 **Ed.). *Molecular Biology and Evolution* 37: 1530–1534.**

640 **Möller M, Cronk QCB. 1997. Phylogeny and disjunct distribution: evolution of *Saintpaulia* (Gesneriaceae).**
641 ***Proceedings of the Royal Society of London. Series B: Biological Sciences* 264: 1827–1836.**

642 **Montefiori M, Brendolise C, Dare AP, Lin-Wang K, Davies KM, Hellens RP, Allan AC. 2015. In the**
643 **Solanaceae, a hierarchy of bHLHs confer distinct target specificity to the anthocyanin regulatory**
644 **complex. *Journal of Experimental Botany* 66: 1427–1436.**

645 **Mukumoto F, Hirose S, Imaseki H, Yamazaki K. 1993. DNA sequence requirement of a TATA element-**
646 **binding protein from Arabidopsis for transcription in vitro. *Plant Molecular Biology* 23: 995–1003.**

647 **Murashige T, Skoog F. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures.**
648 ***Physiologia Plantarum* 15: 473–497.**

649 **Nabeshima T, Yang S-J, Ohno S, Honda K, Deguchi A, Doi M, Tatsuzawa F, Hosokawa M. 2017. Histogen**
650 **layers contributing to adventitious bud formation are determined by their cell division activities.**
651 ***Frontiers in Plant Science* 8: 1749.**

652 **Ohki S. 1994. Scanning electron microscopy of shoot differentiationin vitro from leaf explants of the African**
653 **violet. *Plant Cell, Tissue and Organ Culture* 36: 157–162.**

654 **Ong-Abdullah M, Ordway JM, Jiang N, Ooi S-E, Kok S-Y, Sarpan N, Azimi N, Hashim AT, Ishak Z, Rosli**
655 **SK, *et al.* 2015. Loss of *Karma* transposon methylation underlies the mantled somaclonal variant of**
656 **oil palm. *Nature* 525: 533–537.**

657 **Outchkourov NS, Carollo CA, Gomez-Roldan V, De Vos RCH, Bosch D, Hall RD, Beekwilder J. 2014.**
658 **Control of anthocyanin and non-flavonoid compounds by anthocyanin-regulating MYB and bHLH**
659 **transcription factors in Nicotiana benthamiana leaves. *Frontiers in Plant Science* 5.**

660 **Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables**
661 **improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 33: 290–295.**

662 **Quattrocchio F, Wing JF, Va K, De, Woude, Mol JNM, Koes R. 1998. Analysis of bHLH and MYB domain**
663 **proteins: species-specific regulatory differences are caused by divergent evolution of target**
664 **anthocyanin genes. *The Plant Journal* 13: 475–488.**

665 **Ramsay NA, Glover BJ. 2005. MYB–bHLH–WD40 protein complex and the evolution of cellular diversity.**
666 ***Trends in Plant Science* 10: 63–70.**

667 **Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR : a Bioconductor package for differential expression**
668 **analysis of digital gene expression data. *Bioinformatics* 26: 139–140.**

669 **Satina S. 1940. Demonstration of the Three Germ Layers in the Shoot Apex of Datura by Means of Induced**
670 **Polyploidy in Periclinal Chimeras. : 12.**

671 **Sato M, Kawabe T, Hosokawa M, Tatsuzawa F, Doi M. 2011. Tissue culture-induced flower-color changes in**
672 **Saintpaulia caused by excision of the transposon inserted in the flavonoid 3′, 5′ hydroxylase (F3′5′H)**
673 **promoter. *Plant Cell Reports* 30: 929–939.**

674 **Spelt C, Quattrocchio F, Mol JNM, Koes R. 2000. *anthocyanin1* of Petunia encodes a basic Helix-Loop-Helix**
675 **protein that directly activates transcription of structural anthocyanin genes. *The Plant Cell* 12: 1619–**
676 **1631.**

677 **Stracke R, Werber M, Weisshaar B. 2001. The R2R3-MYB gene family in Arabidopsis thaliana. *Current***
678 ***Opinion in Plant Biology* 4: 447–456.**

679 **Tang M, Xue W, Li X, Wang L, Wang M, Wang W, Yin X, Chen B, Qu X, Li J, *et al.* 2022. Mitotically**
680 **heritable epigenetic modifications of *CmMYB6* control anthocyanin biosynthesis in chrysanthemum.**
681 ***New Phytologist* 236: 1075–1088.**

682 **Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new**
683 **capabilities and interfaces. *Nucleic Acids Research* 40: e115.**

684 **Whelan S, Irisarri I, Burki F. 2018. PREQUAL: detecting non-homologous characters in sets of unaligned**
685 **homologous sequences (J Hancock, Ed.). *Bioinformatics*.**

686 **Xu W, Dubos C, Lepiniec L. 2015. Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR**
687 **complexes. *Trends in Plant Science* 20: 176–185.**

688 **Yang SJ, Ohno S, Deguchi A, Sato M, Goto M, Doi M, Ohnishi M, Tatsuzawa F, Hosokawa M. 2017. The**
689 **histological study in sympetalous corolla development of pinwheel-type flowers of Saintpaulia.**
690 ***Scientia Horticulturae* 223: 10–18.**

691 **Zimmermann IM, Heim MA, Weisshaar Bernd, Uhrig JF. 2004. Comprehensive identification of *Arabidopsis***
692 ***thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal* 40:**
693 **22–34.**

694

695

696 **Figure legends**

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

703

704

705 **Fig. 2** (a) Anthocyanin biosynthesis pathway and anthocyanin content in *Saintpaulia* petals and leaves. Asterisks
706 indicate significant differences (Student's *t*-test, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$). n.s. indicates no significant
707 difference. Error bars represent means ± SD (n = 10 for KP and KW petals, n = 5 for ST and SW petals, and n = 3
708 for KP and KW leaves). (b-d) qRT-PCR analysis of anthocyanin biosynthesis-related genes in petals. Asterisks
709 indicate significant differences (Student's *t*-test, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$). n.s. indicates no significant
710 difference. Error bars represent means ± SD (n = 3). (b) Expression in KP and KW petals. (c) Expression in ST and
711 SW petals. (d) Expression in KP and KW leaves.

712

713

714 **Fig. 3** (a-c) qRT-PCR analysis of anthocyanin biosynthesis related transcription factors. Asterisks indicate
715 significant differences (Student's *t*-test, **P* < 0.05, ***P* < 0.01, ****P* < 0.001). Error bars represent means ± SD (n
716 = 3). (a) Expression levels in KP and KW petals. (b) Expression levels in ST and SW petals. (c) Expression levels
717 in KP and KW leaves. (d) RT-PCR results of *SiMYB1* and *SiActin* in petals and leaves of KP and KW (n = 3). The
718 amplification reaction was carried out for 35 cycles (plateau). (e-f) Molecular phylogenetic tree of MYB TFs (e)
719 and bHLH TFs (f) amino acid sequences. Robust branches are indicated by white circles.

720

721

722

723 **Fig. 4** (a,b) Methylation levels of candidate genes responsible for unstable leaf pigmentation. The vertical axis
724 indicates the methylation level, and the horizontal axis indicates the position in the Sio_r2.0_p0007.1 contig.
725 Methylation levels are shown for each context (CG, CHG, CHH), and each strand (plus and minus). Red bars
726 indicate methylation levels in KP, and blue bars in KW. (a) Methylation levels in *SiMYB2*. Regions where
727 differences in methylation levels were observed are highlighted in orange. (b) Methylation levels in *SibHLH2*. (c-
728 e) McrBC-PCR results of *SiMYB2* exon2 in leaf (c), petal at S2 (d), and petal at S4 (e) of KP and KW. (f) McrBC-
729 qPCR results of *SiMYB2* exon2 in KP and KW tissues. Asterisks indicate significant differences (Student's *t-test*,
730 $^{**}P < 0.01$, $^{***}P < 0.001$). The error bar represents mean ± SD (n = 3).

731

732 **Fig. 5** Characteristics of *SiMYB2*. (a) RNA-seq and CAGE-seq mapping within the *SiMYB2* locus
733 (Sio_r2.0_p0007.1.g08820.1) and its upstream region. Vertical axes represent read counts. Horizontal axes
734 represent a position on the contig Sio_r2.0_p0007.1. Peaks in RNA-seq indicate exon sequences. Splicing junctions
735 (sj1/2/3) were located between p1, and p2, p2 and p4, and p3 and p4. Although p1 + p2 and p3 + p4 were predicted
736 as independent genes (Sio_r2.0_p0007.1.g08810.1, Sio_r2.0_p0007.1.g08820.1), they were the same gene. Peaks
737 in CAGE-seq indicate a transcription start site. (b-c) Expected structure of the *SiMYB2* locus based on mapping and
738 cloning results. Transcripts consisting of exons 1, 2, and 4 were designated as *SiMYB2-Long*, and those consisting
739 of exons 3 and 4 were designated as *SiMYB2-Short*. (d) qRT-PCR analysis of *SiMYB2* variants in KP and KW petals,
740 ST, and SW petals, KP and KW leaves, pink and white regions in KWS petals, and purple and white regions in
741 SWS petals. Asterisks indicate significant differences (Student's *t*-test, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$). Error
742 bars represent means ± SD (n = 3).

743

Fig. 6 (a) Multiple alignment of MYB amino acids from *Saintpaulia* and other plants. The R2/R3 domain and SG 6 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), *SibHLH1* (bottom 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 four days 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 means ± SD (n = 3). 1: GFP, 2: *SiMYB1* + *SibHLH1*, 3: *SiMYB2-Long* + *SibHLH1*, 4: *SiMYB2-Short* + *SibHLH1*. Asterisks indicate significant differences according to the Dunnett's range test compared to *GFP*, (**P* < 0.05, ***P* < 0.01, ****P* < 0.001). (e) qRT-PCR analysis of ABGs in infiltrated tobacco leaves. 1: GFP, 2: *SibHLH1*, 3: *SiMYB1*, 4: *SiMYB2-Long*, 5: *SiMYB2-Short*, 6: *SiMYB1* + *SibHLH1*, 7: *SiMYB2-Long* + *SibHLH1*, 8: *SiMYB2-Short* + *SibHLH1*. Error bars represent means ± SD (n = 3). Asterisks indicate significant differences according to the Dunnett's range test compared to *GFP*~~,~~ (**P* < 0.05, ***P* < 0.01, ****P* < 0.001).

760

761 **Fig. 7** (a) The position of primers designed from the *SiMYB2* locus sequence is indicated by arrowheads. (b)
762 Confirmation of indels by PCR in some cultivars. The abbreviated names of the cultivars are shown across the top
763 of the gel lanes. The positions of the primers (In/Del F and R) used are shown in (a). (c) Multiple sequence alignment
764 results of TIR-like sequences. Black bars indicate corresponding bases between 5' TIR and 3' TIR inverted
765 sequences. (d-h) Characterization of the mutant obtained by tissue culture (m-SH) and parental cultivar (SH). (d)
766 Photo of rosettes. White bars indicate 5 cm scale. (e) Photo of flowers. White bars indicate 1.5 cm scale. (f)
767 Anthocyanin content of SH and m-SH leaves. The content was determined as cyanidin-3-glucoside equivalents.
768 Error bars represent means ± SD (n = 3). Asterisks indicate significant differences (Student's *t*-test, $*P < 0.05$, $**P$
769 < 0.01, $***P < 0.001$). (g) qRT-PCR quantification of SiMYB2 variants in leaves. Error bars represent means ± SD
770 (n = 3). Asterisks indicate significant differences (Student's *t*-test, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$). (h)
771 Confirmation of the presence of a transposon-like element by PCR in SH and m-SH. The positions of the primers
772 used are shown in (a).

773

774 **Fig. 8** Hypothetical model to explain white stripe patterns. Pigmented petals have an epigenomic state in which
775 *SiMYB2-Long* is expressed. Non-pigmented petals have an epigenomic state in which *SiMYB2-Short* is expressed.
776 White striped petals may be the result of an epigenomic state in which *SiMYB2-Long* is expressed in the L1 layer
777 and *SiMYB2-Short* in the L2 layer and below.

778

779

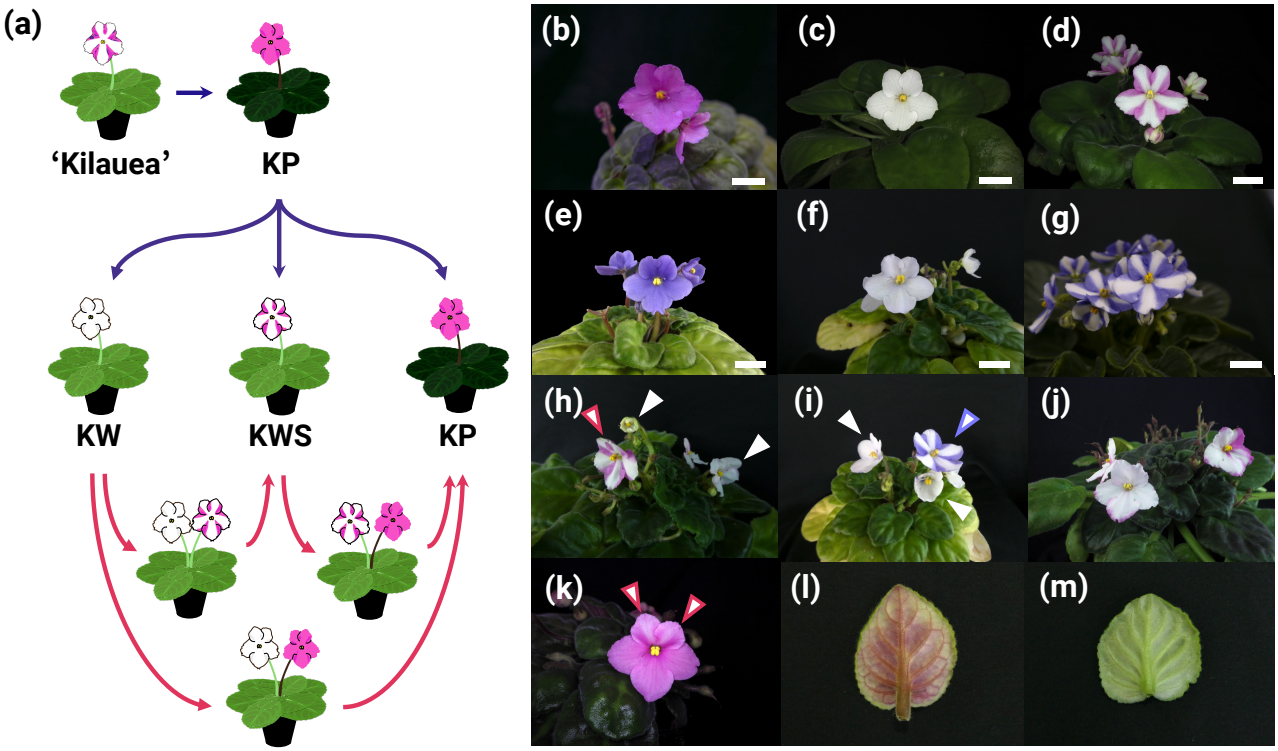
(a)
'Kilauea'
KP
KW
KWS
KP
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)

Fig. 1

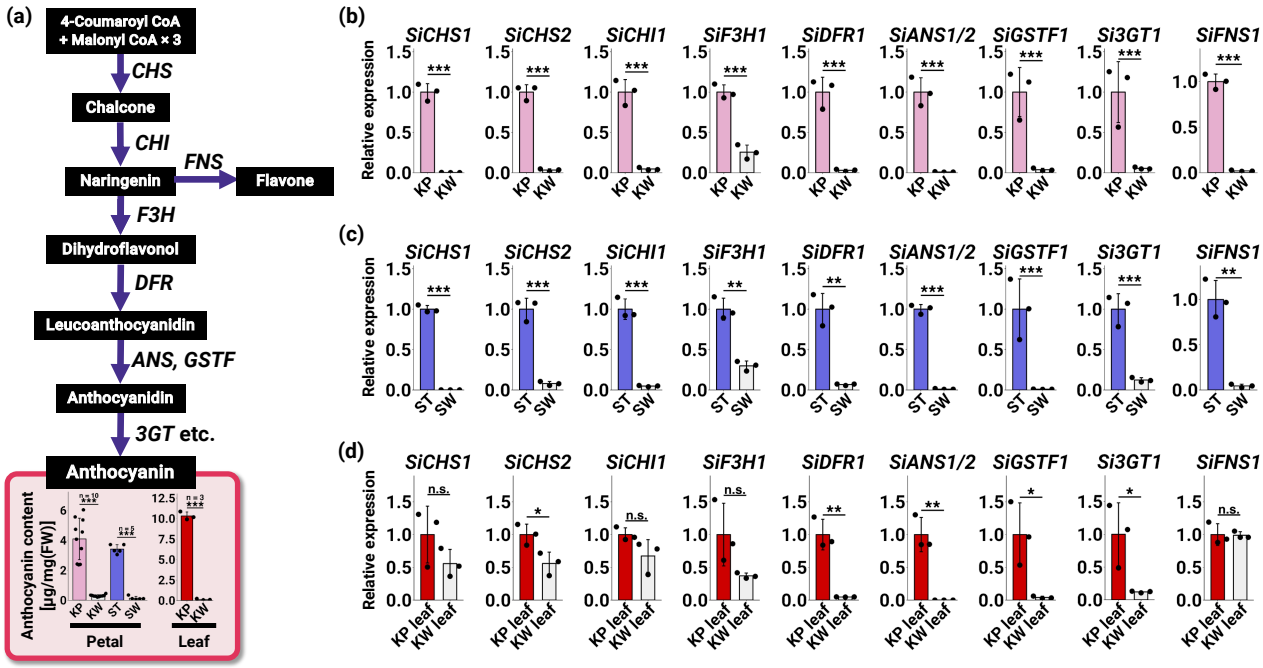
(a)
4-Coumaroyl CoA + Malonyl CoA × 3
CHS
Chalcone
CHI
Naringenin
FNS
Flavone
F3H
Dihydroflavonol
DFR
Leucoanthocyanidin
ANS, GSTF
Anthocyanidin
3GT etc.
Anthocyanin
Anthocyanin content [μg/mg(FW)]
Petal
Leaf
(b)
(c)
(d)
SiCHS1
SiCHS2
SiCHI1
SiF3H1
SiDFR1
SiANS1/2
SiGSTF1
Si3GT1
SiFNS1
Relative expression
KP
KW
ST
SW
KP leaf
KW leaf

Fig. 2

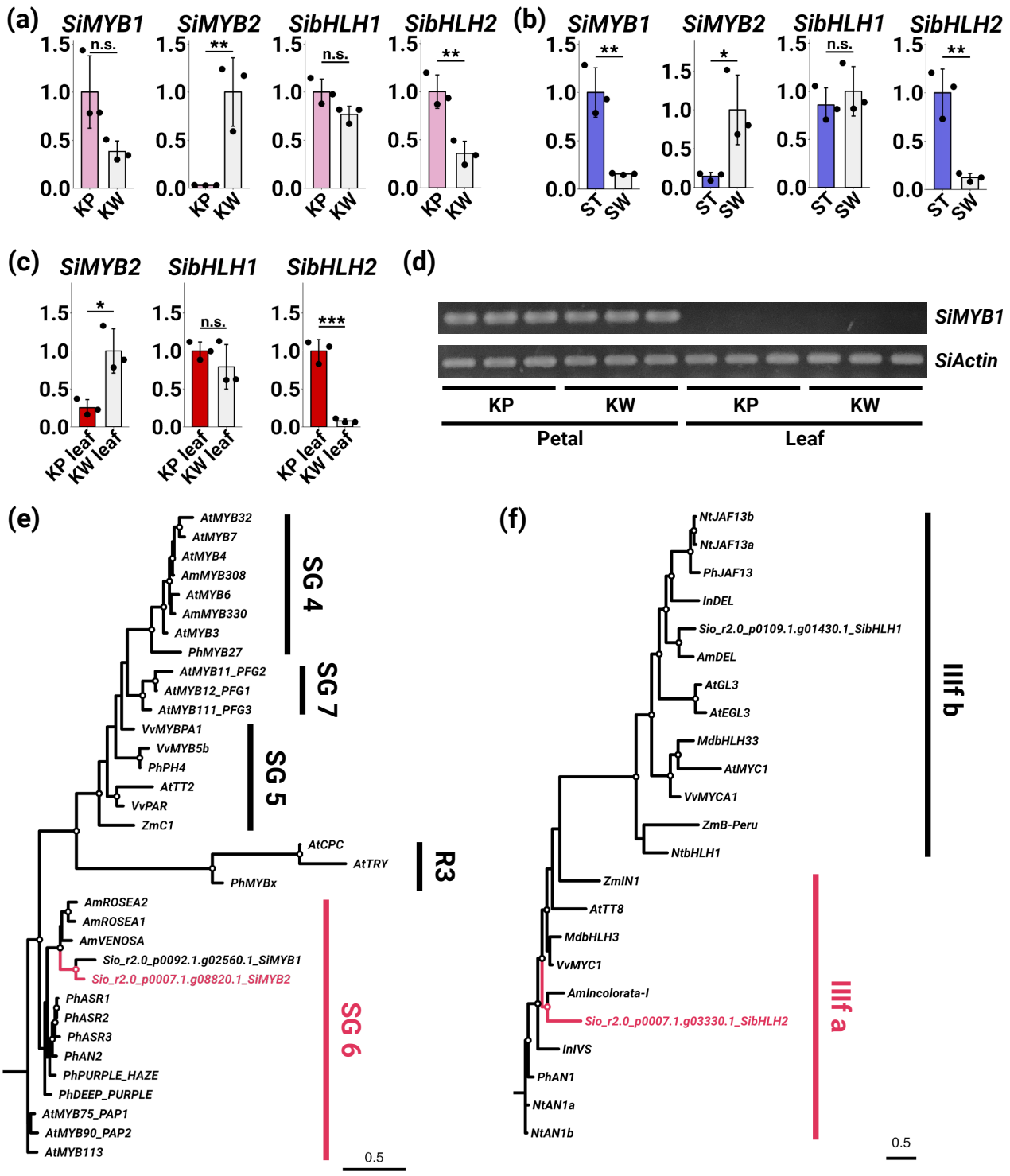
(a)
SiMYB1
SiMYB2
SibHLH1
SibHLH2
n.s.
**
n.s.
**
KP
KW
(b)
SiMYB1
SiMYB2
SibHLH1
SibHLH2
**
*
n.s.
**
ST
SW
(c)
SiMYB2
SibHLH1
SibHLH2
*
n.s.
***
KP leaf
KW leaf
(d)
SiMYB1
SiActin
KP
KW
KP
KW
Petal
Leaf
(e)
AtMYB32
AtMYB7
AtMYB4
AmMYB308
AtMYB6
AmMYB330
AtMYB3
PhMYB27
AtMYB11_PFG2
AtMYB12_PFG1
AtMYB111_PFG3
VvMYBPA1
VvMYB5b
PhPH4
AtTT2
VvPAR
ZmC1
AtCPC
AtTRY
PhMYBx
AmROSEA2
AmROSEA1
AmVENOSA
Sio_r2.0_p0092.1.g02560.1_SiMYB1
Sio_r2.0_p0007.1.g08820.1_SiMYB2
PhASR1
PhASR2
PhASR3
PhAN2
PhPURPLE_HAZE
PhDEEP_PURPLE
AtMYB75_PAP1
AtMYB90_PAP2
AtMYB113
SG 4
SG 7
SG 5
R3
SG 6
0.5
(f)
NtJAF13b
NtJAF13a
PhJAF13
InDEL
Sio_r2.0_p0109.1.g01430.1_SibHLH1
AmDEL
AtGL3
AtEGL3
MdbHLH33
AtMYC1
VvMYCA1
ZmB-Peru
NtbHLH1
ZmIN1
AtTT8
MdbHLH3
VvMYC1
AmIncolorata-I
Sio_r2.0_p0007.1.g03330.1_SibHLH2
InIVS
PhAN1
NtAN1a
NtAN1b
IIIf b
IIIf a
0.5

Fig. 3

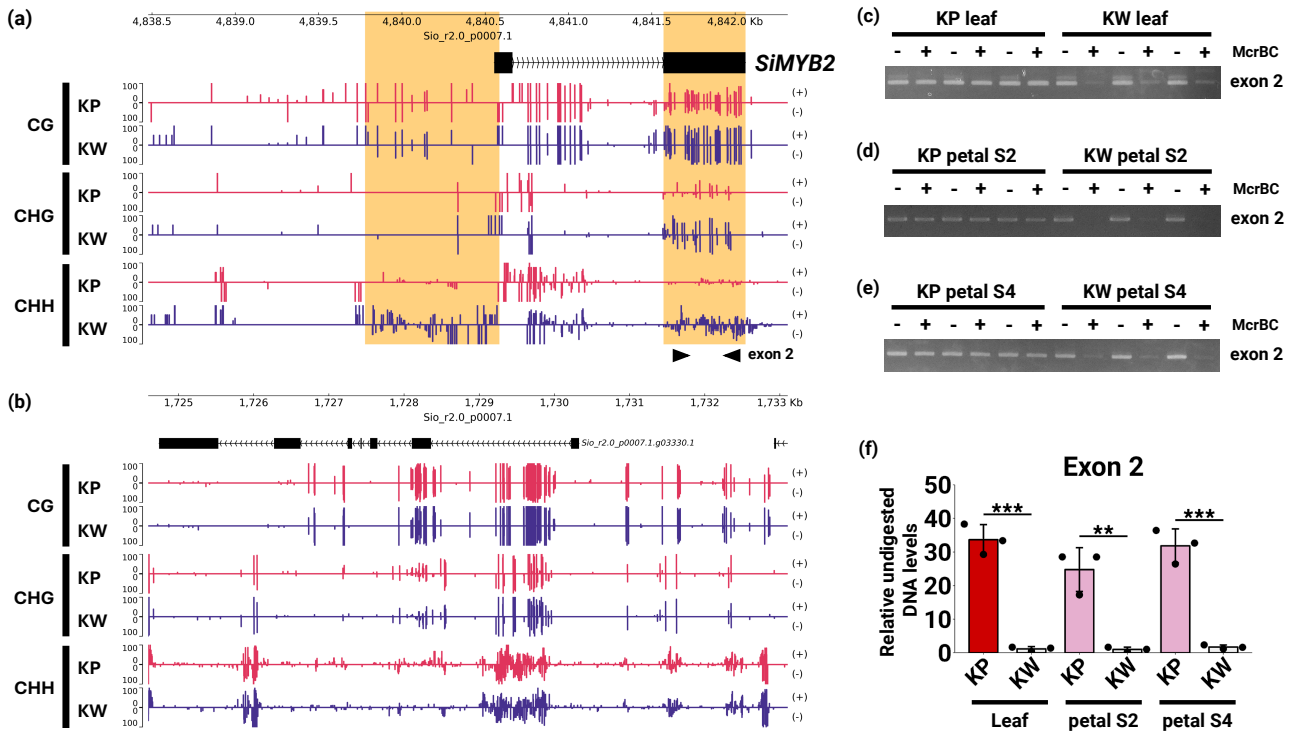
(a)
4,838.5
4,839.0
4,839.5
4,840.0
4,840.5
4,841.0
4,841.5
4,842.0 Kb
Sio_r2.0_p0007.1
SiMYB2
CG
CHG
CHH
KP
KW
exon 2
(b)
1,725
1,726
1,727
1,728
1,729
1,730
1,731
1,732
1,733 Kb
Sio_r2.0_p0007.1
Sio_r2.0_p0007.1.g03330.1
(c)
KP leaf
KW leaf
McrBC
exon 2
(d)
KP petal S2
KW petal S2
McrBC
exon 2
(e)
KP petal S4
KW petal S4
McrBC
exon 2
(f)
Exon 2
Relative undigested DNA levels
50
40
30
20
10
0
***
**
***
KP
KW
Leaf
petal S2
petal S4

Fig. 4

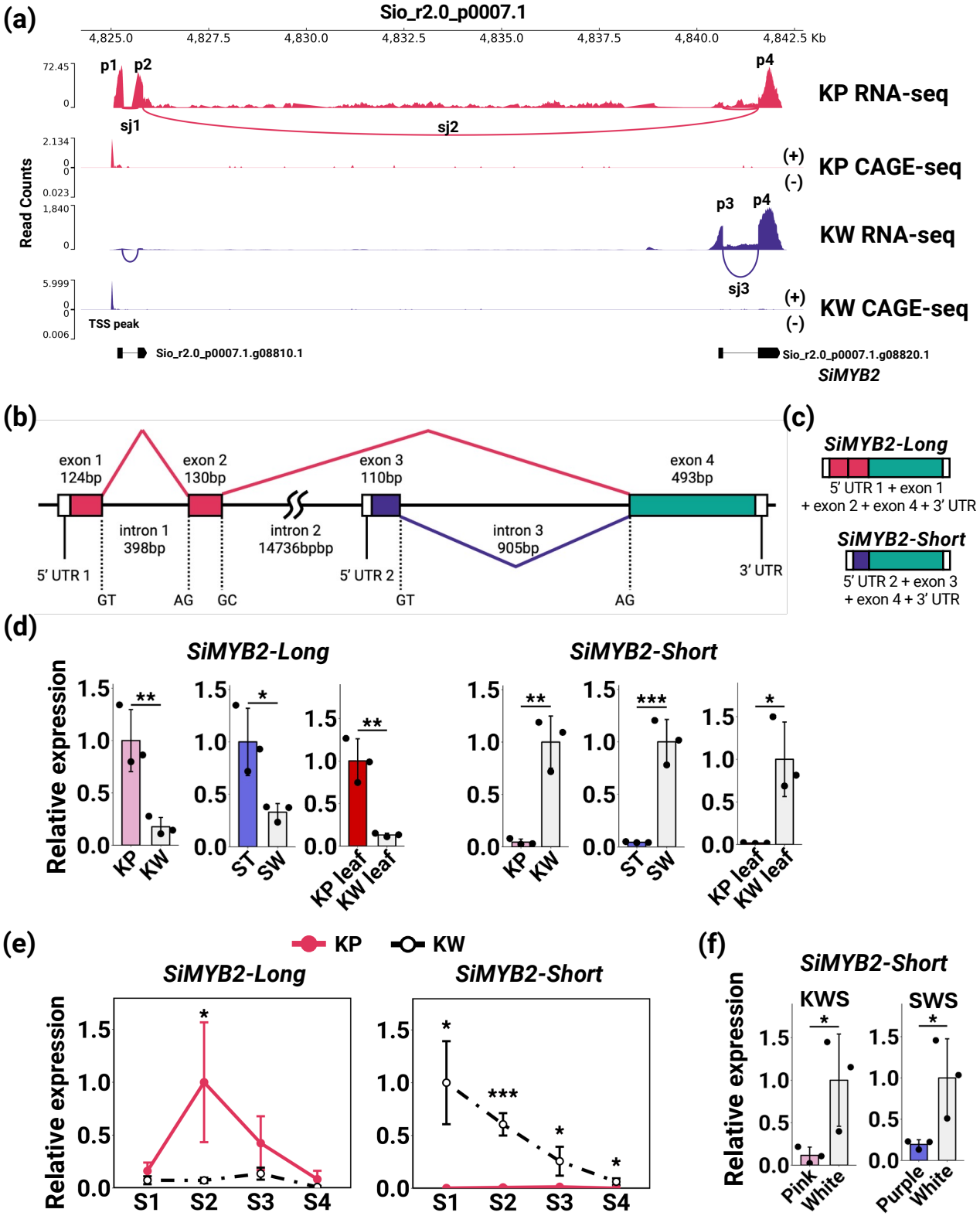
(a)
Sio_r2.0_p0007.1
4,825.0
4,827.5
4,830.0
4,832.5
4,835.0
4,837.5
4,840.0
4,842.5 Kb
Read Counts
p1
p2
p4
sj1
sj2
KP RNA-seq
KP CAGE-seq
(+)
(-)
p3
p4
KW RNA-seq
sj3
KW CAGE-seq
TSS peak
Sio_r2.0_p0007.1.g08810.1
Sio_r2.0_p0007.1.g08820.1
SiMYB2
(b)
exon 1
124bp
exon 2
130bp
exon 3
110bp
exon 4
493bp
intron 1
398bp
intron 2
14736bpbp
intron 3
905bp
5′ UTR 1
5′ UTR 2
3′ UTR
GT
AG
GC
GT
AG
(c)
SiMYB2-Long
5′ UTR 1 + exon 1
+ exon 2 + exon 4 + 3′ UTR
SiMYB2-Short
5′ UTR 2 + exon 3
+ exon 4 + 3′ UTR
(d)
SiMYB2-Long
SiMYB2-Short
Relative expression
KP
KW
ST
SW
KP leaf
KW leaf
(e)
KP
KW
SiMYB2-Long
SiMYB2-Short
Relative expression
S1
S2
S3
S4
(f)
SiMYB2-Short
KWS
SWS
Relative expression
Pink
White
Purple
White

Fig. 5

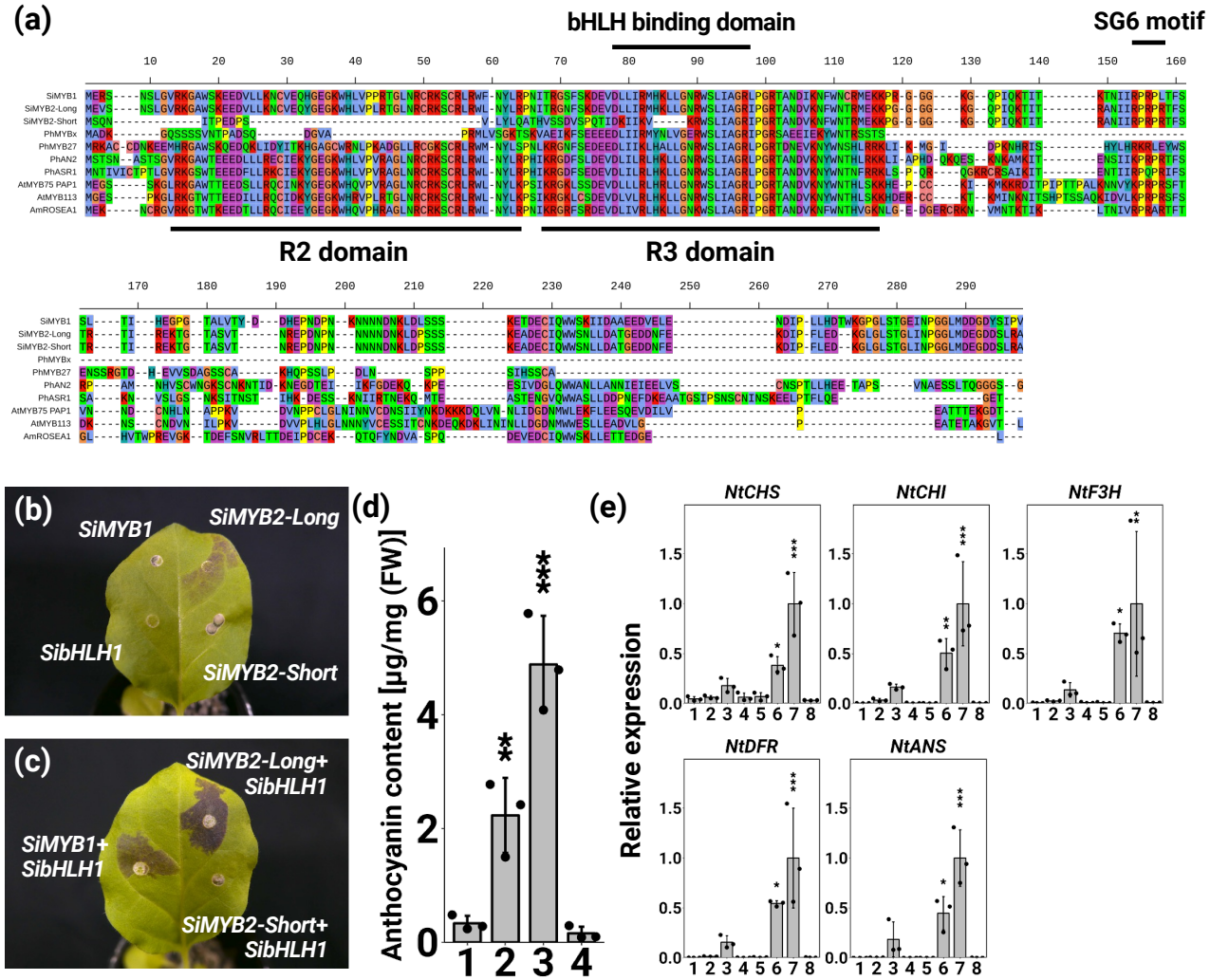
(a)
bHLH binding domain
SG6 motif
R2 domain
R3 domain
SiMYB1
SiMYB2-Long
SiMYB2-Short
PhMYBx
PhMYB27
PhAN2
PhASR1
AtMYB75 PAP1
AtMYB113
AmROSEA1
(b)
SiMYB1
SiMYB2-Long
SibHLH1
SiMYB2-Short
(c)
SiMYB2-Long+
SibHLH1
SiMYB1+
SibHLH1
SiMYB2-Short+
SibHLH1
(d)
Anthocyanin content [µg/mg (FW)]
(e)
NtCHS
NtCHI
NtF3H
NtDFR
NtANS
Relative expression

Fig. 6

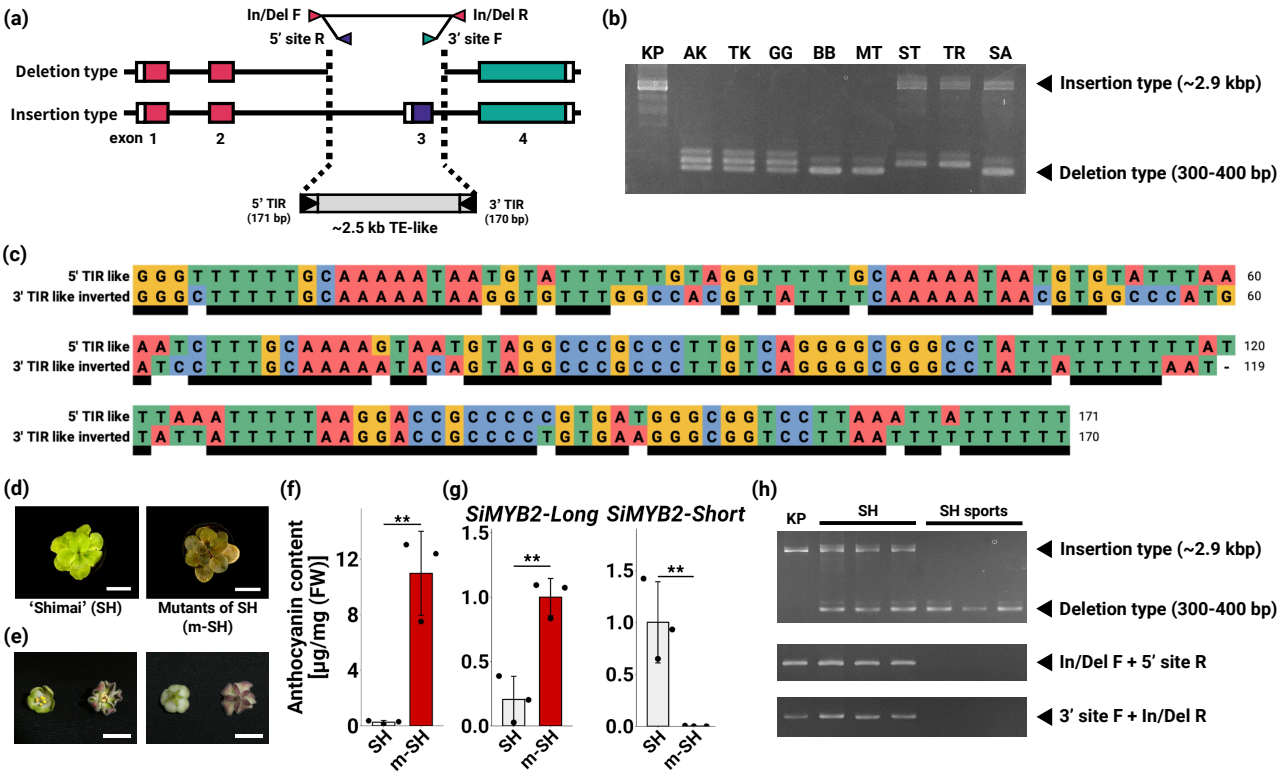
(a)
In/Del F
In/Del R
5' site R
3' site F
Deletion type
Insertion type
exon 1
2
3
4
5' TIR (171 bp)
~2.5 kb TE-like
3' TIR (170 bp)
(b)
KP AK TK GG BB MT ST TR SA
Insertion type (~2.9 kbp)
Deletion type (300-400 bp)
(c)
5' TIR like
3' TIR like inverted
(d)
'Shimai' (SH)
Mutants of SH (m-SH)
(e)
(f)
Anthocyanin content [μg/mg (FW)]
SH
m-SH
(g)
SiMYB2-Long
SiMYB2-Short
(h)
KP
SH
SH sports
Insertion type (~2.9 kbp)
Deletion type (300-400 bp)
In/Del F + 5' site R
3' site F + In/Del R

Fig. 7

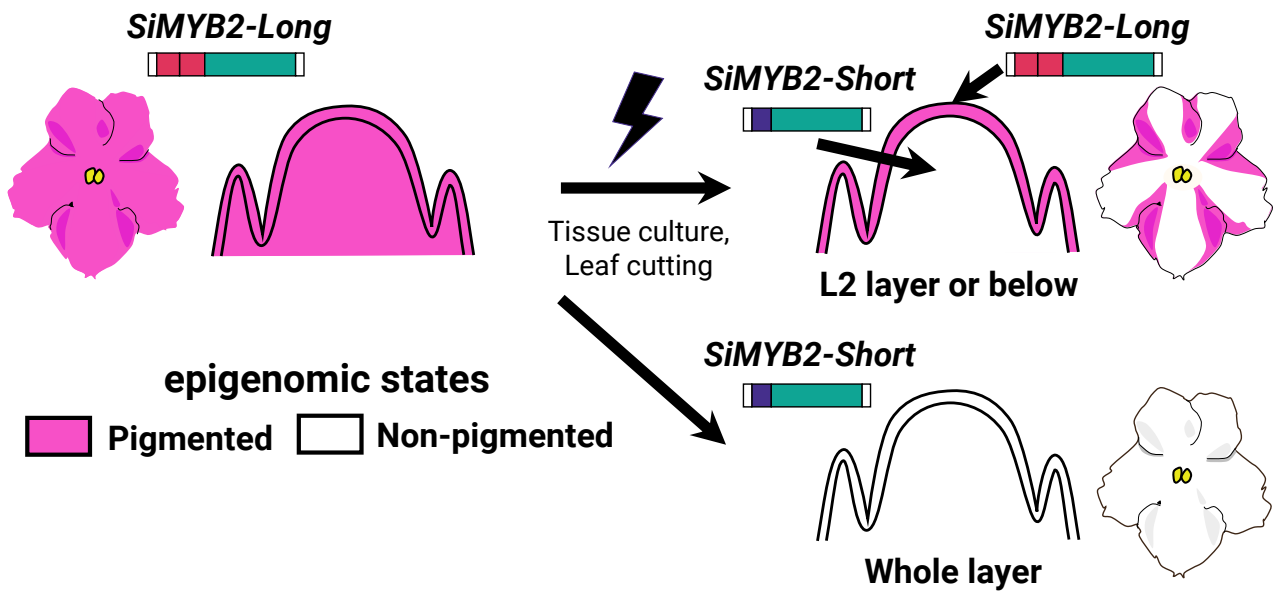
SiMYB2-Long
SiMYB2-Short
SiMYB2-Long
Tissue culture,
Leaf cutting
L2 layer or below
SiMYB2-Short
epigenomic states
Pigmented
Non-pigmented
Whole layer

Fig. 8

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

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*-valu < 0.05)	Sio_r2.0_p0372.1.g00610.1	*SiMATE1*	2.60E-03	1	Protein DETOXIFICATION	*Sesamum indicum*	149.5861203	31.852703
	Sio_r2.0_p0376.1.g00340.1	*SiMATE2*	2.60E-03	2	Protein DETOXIFICATION	*Sesamum indicum*	100.7375617	28.77183033
	Sio_r2.0_p0045.1.g02800.1	*SiF3H1*	2.60E-03	3	Flavone 3-hydroxylase	[Saintpaulia] hybrid cultivar	206.917216	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	*Dorcoceras hygrometricum*	12.37113167	0.112211333
	Sio_r2.0_p0222.1.g00520.1	*SiCHS1*	3.06E-03	5	Chalcone synthase	*Sesamum indicum*	537.372935	6.233511333
	Sio_r2.0_p0085.1.g02260.1	*SiMATE3*	5.98E-03	7	Protein DETOXIFICATION	*Dorcoceras hygrometricum*	62.14773033	4.992234
	Sio_r2.0_p0007.1.g08820.1	*SiMYB2*	6.93E-03	8	R2R3-MYB1	*Primulina swinglei*	11.31502067	287.3776447
	Sio_r2.0_p0408.1.g00410.1	*SiANS1*	1.23E-02	12	leucoanthocyanidin dioxygenase-like	*Sesamum indicum*	155.0496773	11.44214067
	Sio_r2.0_p0007.1.g03330.1	*SibHLH2*	1.45E-02	18	transcription factor BHLH42-like	*Punica granatum*	26.25479133	3.056827
	Sio_r2.0_p0041.1.g00830.1	*SiANS2*	1.45E-02	20	leucoanthocyanidin dioxygenase-like	*Sesamum indicum*	181.95548	11.58148667
	Sio_r2.0_p0021.1.g02260.1	*Si3GT1*	1.60E-02	27	Glycosyltransferase	*Sesamum indicum*	385.9608153	68.02532367
	Sio_r2.0_p0004.1.g00630.1	*SiDFR1*	1.67E-02	40	Dihydroflavanone reductase	*Erythranthe lewisii*	312.942332	17.74549033
	Sio_r2.0_p0044.1.g01100.1	*SiCHI1*	1.77E-02	50	Chalcone-flavonone isomerase family protein	*Primulina swinglei*	1359.730184	271.050537
	Sio_r2.0_p0002.1.g09030.1	*SiFNS1*	2.35E-02	72	Flavone synthase II	*Dorcoceras hygrometricum*	285.4723053	32.438459
	Sio_r2.0_p0008.1.g04160.1	*Si3GT2*	2.95E-02	100	Glycosyltransferase	*Dorcoceras hygrometricum*	83.70253267	33.02726867
	Sio_r2.0_p0005.1.g02330.1	*SiCHS2*	3.15E-02	113	Chalcone synthase	*Sesamum indicum*	1761.998291	358.2101593
	Sio_r2.0_p0043.1.g03850.1	*SiGSTF1*	3.90E-02	165	Glutathione transferase	*Sesamum indicum*	1405.150065	257.1803207
	Sio_r2.0_p0002.1.g10740.1	*SiF3'5'H*	4.06E-02	172	Flavonoid 3',5'-hydroxylase	[Saintpaulia] hybrid cultivar	106.3326237	5.364064333

Fig. S1 Developmental stages of petals used in this study. Black bar indicates 1 cm scale.

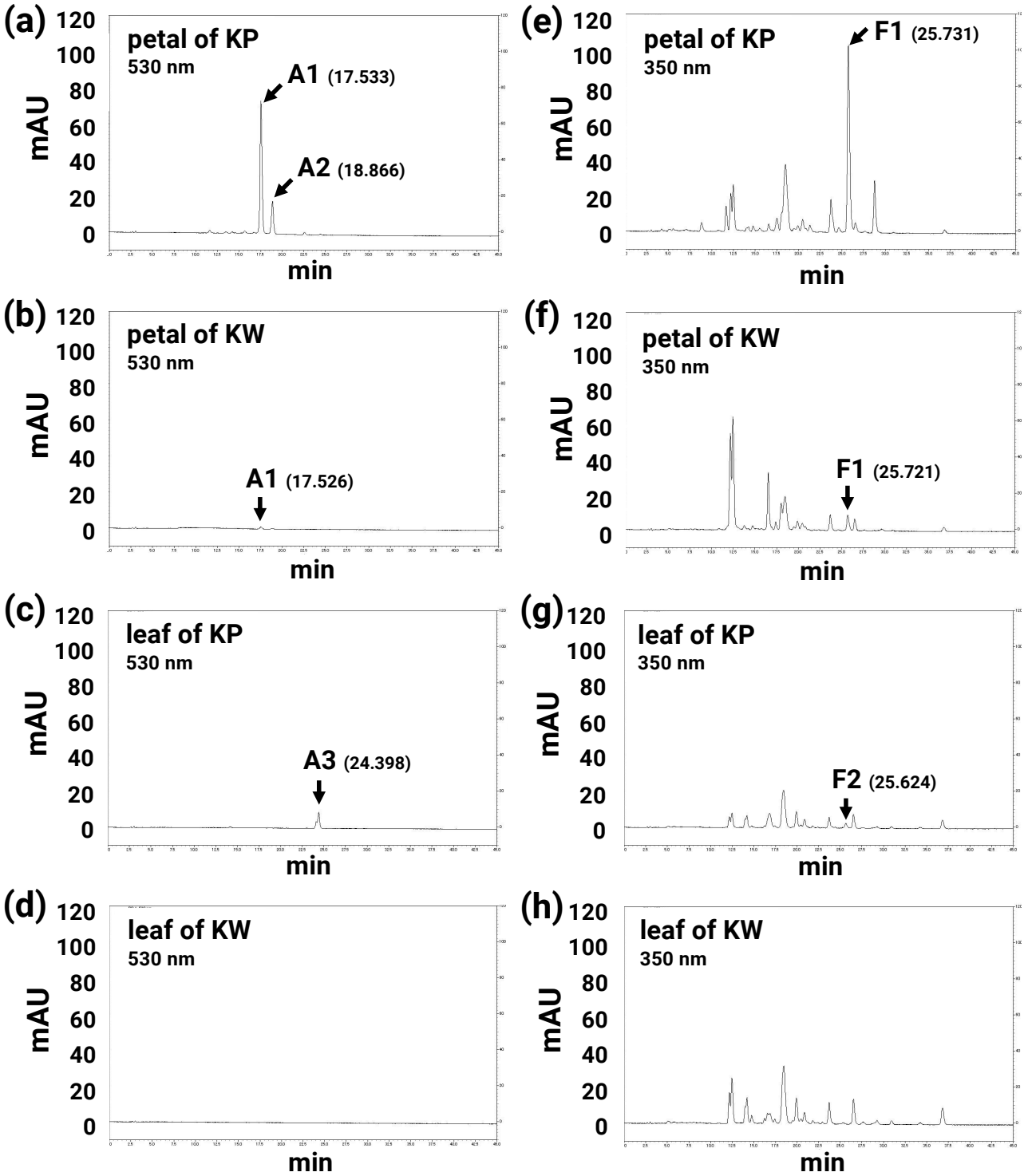

Fig. S2 HPLC analysis of KP and KW petals. (a,b,e,f) Petal. (c,d,g,h) Leaf. (a–d) 530 nm for anthocyanin. (e–h) 350 nm for flavone. A1: Pelargonidin-3-acetyl-rutinoside-5-glucoside, A2: Peonidin-3-acetyl-rutinoside-5-glucoside, A3: Cyanidin-3-acetyl-sambubioside, F1: Apigenin-4'-glucronide, F2: Luteolin-4'-glucronide.

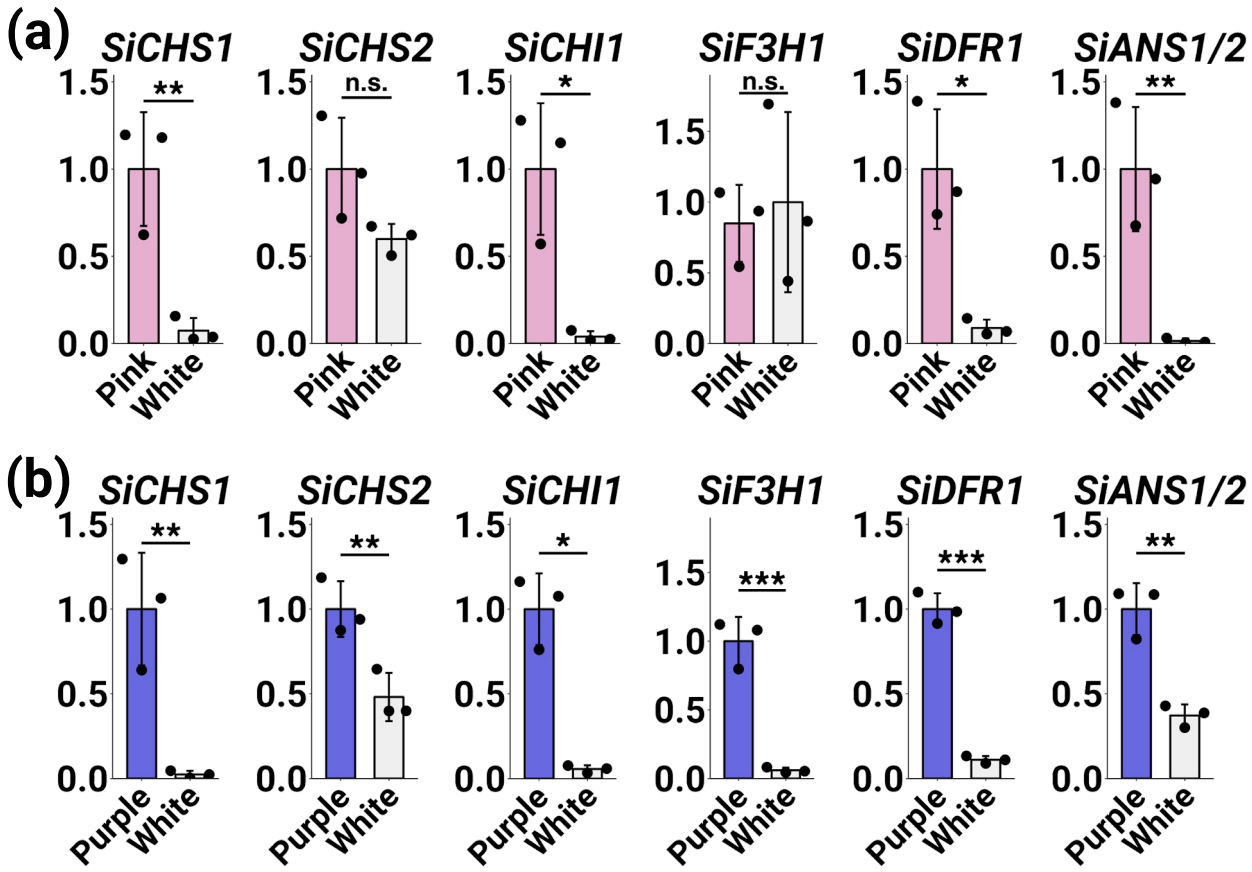

Fig. S3 Expression levels of ABGs in white-striped petals. Asterisks indicate significant differences according to Student's *t*-test (**P* < 0.05, ***P* < 0.01, ****P* < 0.001). n.s. indicates no significant difference. Error bars represent means ± SD (n = 3). (a) Expression levels in KWS petals. Pink indicates pink regions in KWS petals. White indicates white regions. (b) Expression levels in SWS petals. Purple indicates purple regions in SWS petals. White indicates white regions.

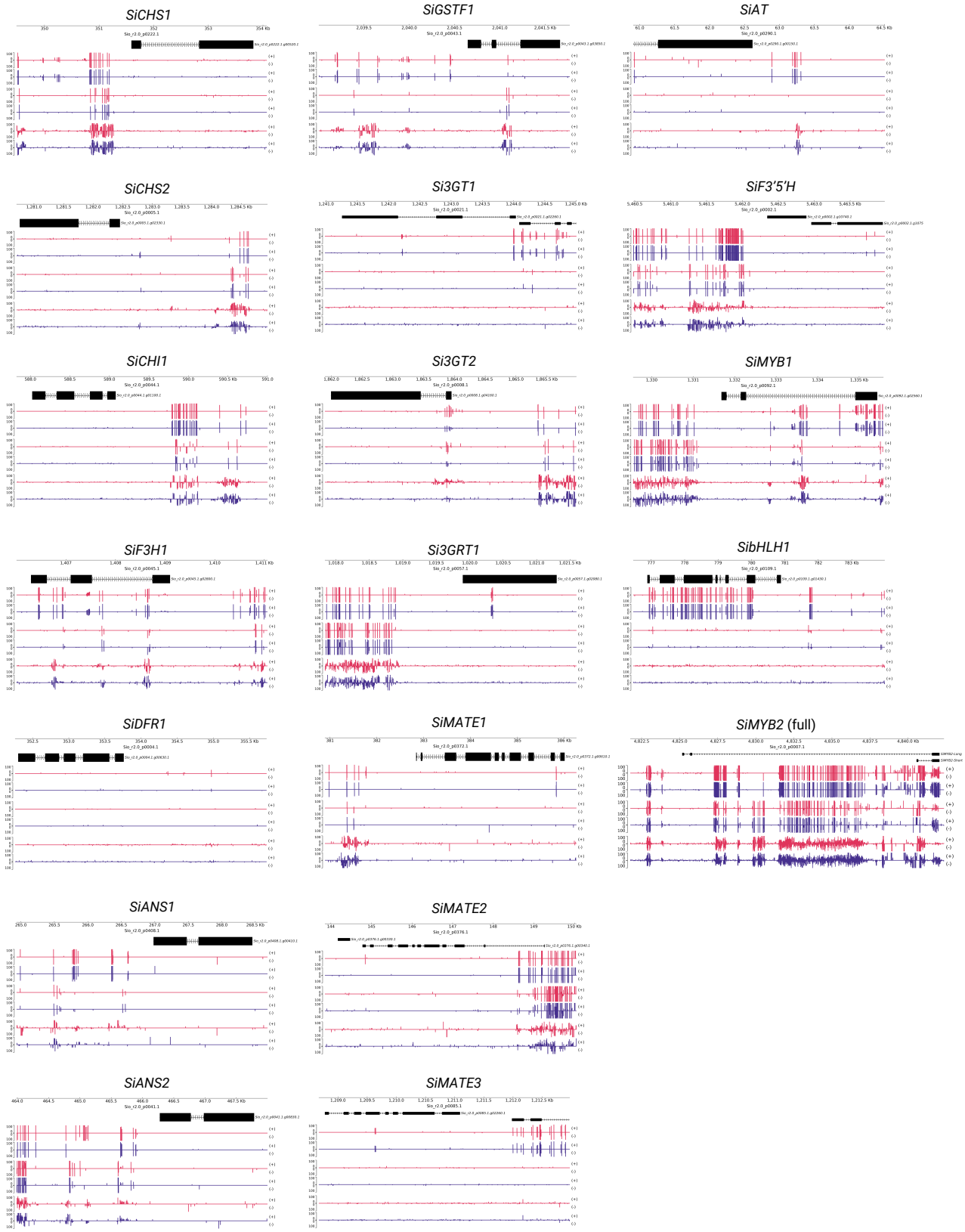

Fig. S4 Methylation levels of anthocyanin biosynthesis related genes in KP and KW leaves.

TTTTTATTTGATATATTAACTTTACTTATTGAGCTATTGTATTCTCAGCCGCCTAAAAATTTA
TTTTCCAACCAAAAAATATGATTAGCAGAAAATTTATATTATCAAATACCTGGATGCGTAA
AATCTGTGGCATTTCAATATTTTTTATTTTTAATGTAATTGATTATTGATTTTAAAATACTTT
TTACTTTTCTATTTTATTATTTAATATATATATATATATTAATATATCAATACACGGGTTTTT
TGCAAAAATAATGTATTTTTTGTAGGTTTTTGCAAAAATAATGTGTATTTAAAATCTTTGCA
AAAGTAATGTAGGCCCGCCCTTGTCAGGGGCGGGCCTATTTTTTTTTTATTTAAATTTTTAA
GGACCGCCCCCGTGATGGGCGGTCCTTAAATTATTTTTTTTTTTTTTAAAAAAAACATCATT
GGCCCGCCCTTGACAGGGGCGGGCCTTTTTTTTAAAAAAAAAATTCTGTAAGGACCGCCC
CTGCCAGGGGCGGTCCTTGGGCTGTGTGCCCTTCACGTGAATCTTAAATGATTCACGTGAA
CTTGTCTACGCTTTCAGAAATCGCCTCTATATATTTAGTTCATTGTTCGTCAAATGAATTCA
GAATCCTCAATTCCTTCGTCTTTTAAATTGTTCGTTGTTTGACATAATTTTCTTTTTTTGGCT
TTGTATCATATATCGTGCCAGAAATTGATCGTATAAAGGTTAGTTTATACCCATTTTTTTTT
GTTTTACTTAAAATTTTTATTTGTTAAACTTTAATTATTATTAGTATTGTTTAAATTACGGTA
GTTAAATTTTATTTTATTCTTTCATTTGTGCAGGTTTGAAGAATTTATACATTTGGTTACAAC
TACATATATGTATATTATTGTAGAAATTAATATGATTATTTGAATGTTGTATGGTTTTTTAC
AAAATTATATTCAATTTAGGTGTTAATTTTAAAAACTATGTATATATTTTAAATATTATCTT
TTTATAATATTATATATTTTTTTCGAAA**TATAAATA**TATTTTTTTGTAAAATTATACGTGTG**TA**
TATATATATATATAAATATGAA**TATATATATATATATA**CGTGTATATATGTATATATGTACACT
AAATA**TATAAATA**TATATGTATGTACGTGTGTATATG**TATATATA**TATGAA**TATAAATA**TATG
TATT**TATATATA**TGTACGTGTATATATTTATATATGTACAGTGAATATATGTATTATATTATA
TATGTATGTACGTGTGTGTATATG**TATATATA**TATGAA**TATAAATA**TATGTATT**TATATATA**TG
TACGCGTATATATTTATATATGTACAGTGAATATATGTATTATATTATATATGTATGTACGT
G**TATATATATATATATA**TATATACATATGAGTGTAAATATATATATATATAAG**TATATATA**TA
TGTACAGTAAATA**TATAAATA**TATATGTATGTACGTACGTGTGTGTATATATC**TATATATATA**
TATATAAATATAAAAATATATATATATGTACGTGTATACATGTATATATG**TATATATA**TATAT
TATTATTTAGTTAAAGCTTAAATATTTTTAAAATAAGTATATTACGTTAGAAAAATATTT**TAT**
ATATATAATTAATTTCTGTATTTTTTAAAAAAAATATATGAAACTTTTACTCGCTATAATTGT
ACACATATATGTTTAATTAATTGCTATGATTGTAGTAAACGTATATAGTATTACGAATTAT
ATGTAAATATAATTAAAAAGTGACTTTTACTAAAATATTGATATTATTTCGGA**TATATATA**A
TAATTGTAGGGCATAATAGTAACTATATAATAATTAATTTTTAATATCTATAAATTATTCTT
ATATAATCTATATTCTCAAAAGTGG**TATAAATA**CGAGAGATTATCTTAAAAATATATTCTAT
AAATTACTATACGCAATTATCAATTACTTGAATGTTATTTTTTTTTTTACAAAATAATTTTTC
ATTTTTTCGAAATATAAACATATATTTATTTTAATTATTTAGA**TATATATA**TATGTATCTAAA
TATATATATATATATATTTACGTGTATTTATGTACG**TATATATA**GTAAAATTGTTACAAATTAA
TTTTGAACAAATATTATCTTATTTAGTTAAAACTTAAATATTTTTAAAATATGTATATTACTT
TAGAAAAATATTTATATATATTTTTAGAATATTTGTATATATAATTAAGTAGTTCGAACAAGT
AATATATTTTAGTAAATAATTAGTAAATCTGTATTTTTTTAAAACAAAAATTATATTTTTAA
GAAATATATATTGTATTTTTCGAAAATATGTATATACATATAAAAATAATTATACGTATTC
AAATTCTTTCTCTTTAATTATTAACATTAGATTTTGAAAAAAATAAAATGAATAATTTTATA
CTAAACATTTGTCAAGTTAAAGTTAGTTTCTGGTAACTTATCATTCTGTTCTTAATTTTTTAG
ATGTCTCAGAATATTACTCCCGAAGATCCGTCAGTCCTATACTTACAGGCGACCCATGTTT
CTAGTGATGTCTCTCCTCAGACTATTGATAAAATCATTAAAGTGAAGCGGTGCGATAATCT

Fig. S5 Predicted promoter sequences of *SiMYB2-Short*. Two types of TATA-boxes reported by Mukumoto et al. (1993) are highlighted. Start codon is indicated in red letters. Predicted 5' UTR sequences is indicated in blue letters.

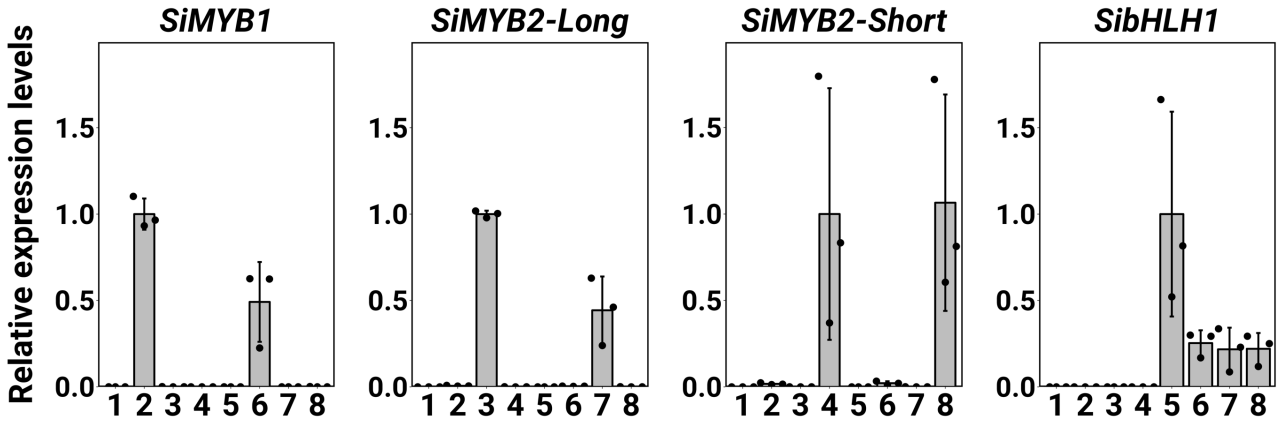

Fig. S6 qRT-PCR analysis of the agroinfiltrated genes in tobacco leaves. 1: GFP, 2: *SibHLH1*, 3: *SiMYB1*, 4: *SiMYB2-Long*, 5: *SiMYB2-Short*, 6: *SiMYB1* + *SibHLH1*, 7: *SiMYB2-Long* + *SibHLH1*, 8: *SiMYB2-Short* + *SibHLH1*. Error bars represent means ± SD (n = 3).

Fig. S7 Confirmation of *SiMYB2-Short* repression activity as R3-MYB. The ratio indicates *SiMYB1*: *SibHLH1*: *GFP* (left) or *SiMYB1* (right). Tobacco leaves were photographed four days post infiltration. Infiltration was performed using two leaves from each of the three tobacco plants.

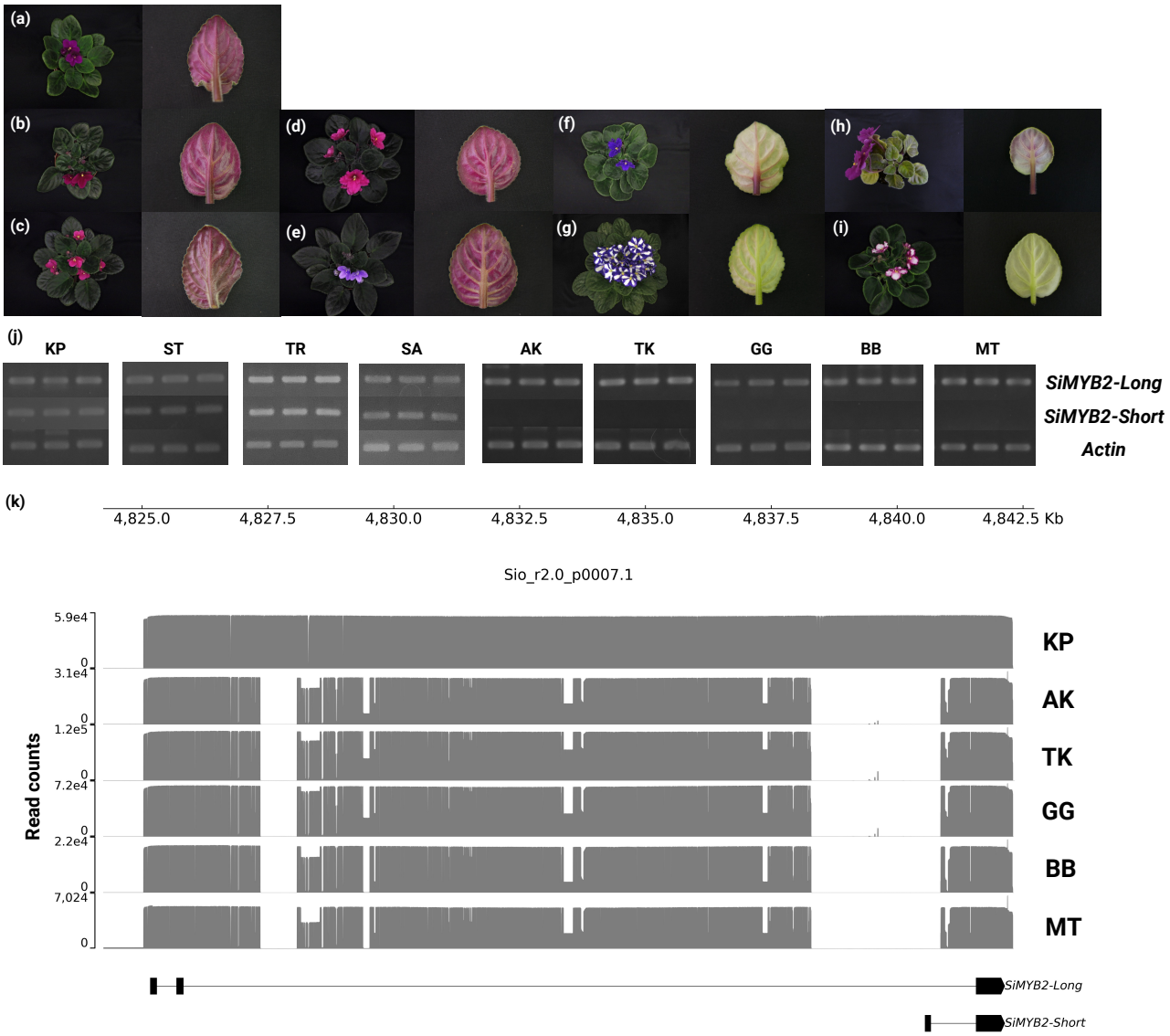

Fig. S8 Transcript variants expression and genome structure of the *SiMYB2* gene in several cultivars. (a-i) Rosette and leaf photos of several cultivars used in this study. (a) 'Akira' AK, (b) 'Tomoko' TK, (c) 'Georgia' GG, (d) 'Barbara' BB, (e) 'Manitoba' MT, (f) 'Taro' TR, (g) TR with white-striped petals, (h) 'Saturn' SA with red petals, (i) SA. (j) RT-PCR results of *SiMYB2* variants in several cultivars. The amplification reaction was carried out for 40 cycles (plateau). Actin was used as control. (k) Amplicon-seq results in several cultivars. The vertical axis indicates read counts and the horizontal axis indicates the position in the Sio_r2.0_p0007.1 contig.