Genome sequence and RNA-seq analysis reveal genetic basis of flower coloration in the giant water lily *Victoria cruziana*

Melina Sophie Nowak*[1], Benjamin Harder*[1], Samuel Nestor Meckoni*[1], Ronja Friedhoff[1], Katharina Wolff[1], Boas Pucker[1+]

[1] Plant Biotechnology and Bioinformatics, Institute of Plant Biology & BRICS, TU Braunschweig, Mendelssohnstrasse 4, 38106 Braunschweig, Germany
* contributed equally
+ corresponding author: b.pucker@tu-braunschweig.de

ORCIDs:
MSN: 0009-0000-9859-8992
BH: 0009-0001-5001-7164
SNM: 0000-0003-4877-8762
RF: 0009-0002-0270-4661
KW: 0000-0002-3616-0482
BP: 0000-0002-3321-7471

Keywords: anthocyanin biosynthesis, flavonoid biosynthesis, long read sequencing, direct RNA sequencing, Nymphaeceae, color change

Abstract

Victoria cruziana is well known for its huge floating leaves covered with sharp spines and its night blooming. Reports indicate that white flowers open during the first night and turn pinkish during the following day and the second night. Here, we set out to unravel the molecular basis and ecological function of the flower color change in *V. cruziana*.
A high quality genome sequence with a N50 of 14.3 Mbp and a total assembly size of 3.54 Gbp was generated as the genetic basis for this study. Comparative transcriptomics revealed the genes required for anthocyanin biosynthesis genes and their transcriptional regulators as differentially expressed between the white and the pinkish stage of a flower. Structural genes with expression differences between white and pinkish flower stages include *VcrF3'H*, *VcrF3'5'H*, *VcrDFR*, *VcrANS*, and *VcrarGST*. The expression pattern of the corresponding transcription factors *VcrMYBSG5*, *VcrMYBSG6*, *VcrTT8*, and *VcrTTG1* also showed differences that aligned with the flower color.

Introduction

Waterlilies of the genus *Victoria* are best known for their giant floating leaves that can endure a substantial weight if equally distributed across the leaf. The genus *Victoria* belongs to the *Nymphaeaceae* and harbors three robust species *V. amazonica*, *V. cruziana* and *V. boliviana* that were recently proposed (Smith *et al.*, 2022). These aquatic species are night bloomers, considered as short-lived perennial plants and naturally grow under tropical conditions. Distinctive shared characteristics are the sharp spines covering most plant parts, massive leaves, and a grand flower. The whole development from seeds to a mature giant water lily takes less than five months (Smith *et al.*, 2022). *Victoria spp.* also have an economic relevance, as the large seeds and rhizomes can be used as a food source (Bortolotto *et al.*, 2015).

The flower, a special feature of *Victoria spp.,* has been reported to undergo a color change from white to pinkish (**Figure 1**) (Wu *et al.*, 2018). Development of the *Victoria* flower bud starts underwater and the bud only opens after reaching the water surface. Interestingly, *Victoria* plants bloom at night. To the best of our knowledge, one individual plant exhibits only a single blooming flower at any given time. Each flower undergoes a flowering period of two consecutive nights, during which the flower color changes from white to pinkish. In addition, anthesis also occurs during these consecutive nights, starting with the female reception, followed by staminate maturation (male phase) (Seymour & Matthews, 2006; Jiang *et al.*, 2021). The ecological reason and molecular basis of the flower color change are currently unknown, but multiple hypotheses exist that could explain it: (1) flower color might change following a visitation/pollination event, (2) light could trigger the accumulation of pigments once flowers emerge from the water, or (3) a developmental program could lead to pigmentation in an age-dependent manner. In the following sections, we will provide context for each of these hypotheses.

Previous studies described flower color changes following pollination events for other plant species (Weiss, 1991; Ruxton & Schaefer, 2016). Although these events are rare, floral color change can serve as a signal between plant and insect to transmit the information that pollination, or at least visitation, already occurred (Weiss, 1991; Ruxton & Schaefer, 2016; Garcia *et al.*, 2022). Flowers of *V. cruziana* appear white in the first night, combined with a fruity smell, secreted by the carpellary appendages which could suggest a function in pollinator attraction (Zini *et al.*, 2019). A previously reported pollinator of *Victoria spp.* are *Cyclocephalini,* a tribe of scarab beetles that are pantropical distributed (Moore & Jameson, 2013). Willmer et al. described the floral color alterations from lilac to white in *Desmodium setigerum* after receiving visitors (Willmer *et al.*, 2009).

Light-induced accumulation of red anthocyanins has been described before in many plant species like *Arabidopsis thaliana* (Maier *et al.*, 2013; Li *et al.*, 2016), *Malus spec.* (Merzlyak & Chivkunova, 2000; Takos *et al.*, 2006), or *Pyrus spec.* (Dussi *et al.*, 1995). The molecular basis

of this light response has been extensively studied and involves an up-regulation of the transcription factors (TFs) controlling the anthocyanin biosynthesis genes (Takos *et al.*, 2006; Li *et al.*, 2016). Anthocyanins can serve as sunscreen thus protecting plants against excessive light irradiation (Gould, 2004) and are strong antioxidants that can scavenge reactive oxygen species (Wang *et al.*, 1997).

Anthocyanin formation might also take place as part of a developmental program, thus depending on the age of the flower and not directly connected to environmental conditions. For example, a study in *Fuchsia excorticata* demonstrated that the flower color change from green to red is age-dependent and not caused by pollination (Delph & Lively, 1989). An investigation in *Pulmonaria collina* also noted an age-dependent color change from red to blue and explains this as a mechanism to direct pollinators towards the young flowers thus increasing the pollination efficiency (Oberrath & Böhning-Gaese, 1999). Similar findings have been reported for *Pedicularis monbeigiana* which shows a color change from white to purple with increasing flower age (Sun *et al.*, 2005).

Figure 1: Flower of *Victoria cruziana*. Comparison of a flower between the first bloom (white) at night (A) and the consecutive night (pink) (B).

The pigments responsible for flower coloration are well studied and often belong to the anthocyanins, carotenoids, or betalains (Tanaka *et al.*, 2008). Anthocyanins and carotenoids are taxonomically widespread and responsible for most flower colors (Tanaka *et al.*, 2008), while betalains are restricted to the Caryophyllales (Timoneda *et al.*, 2019). Anthocyanins are derived from phenylalanine involving the general phenylpropanoid pathway and the flavonoid biosynthesis (**Figure 2**). Structural genes required for the anthocyanin biosynthesis include *phenylalanine ammonia lyase* (*PAL*), *cinnamate 4-hydroxylase* (*C4H*), and *4-coumarate-CoA ligase* (*4CL*) as part of the general phenylpropanoid pathway followed by *chalcone synthase* (*CHS*), *chalcone isomerase* (*CHI*), *flavanone 3-hydroxylase* (*F3H*), *dihydroflavonol 4-reductase* (*DFR*), *anthocyanidin synthase* (*ANS*), *anthocyanin-related Glutathione S-transferase* (*arGST*), and *UDP-glucose: flavonoid-3-O-glucosyltransferase* (*UFGT*) as one branch of the flavonoid

biosynthesis (Winkel-Shirley, 2001; Eichenberger *et al.*, 2023). Activity of these structural genes is orchestrated by an ensemble of transcription factors. The two largest transcription factor families in plants, MYBs and bHLHs, contribute substantially to the regulation of different branches of the flavonoid biosynthesis. MYB transcription factors are able to bind DNA through a highly conserved DNA-binding repeat domain (Prouse & Campbell, 2012), while bHLHs can bind DNA with a stretch of basic amino acids (Voronova & Baltimore, 1990). MYBs can be divided into different groups based on the number of characteristic repeats with the R2R3-MYBs playing a predominant role in the control of the flavonoid biosynthesis (Baudry *et al.*, 2004; Stracke *et al.*, 2007; Gonzalez *et al.*, 2008; Li, 2014; Marin & Pucker, 2023). Specific subgroups of these R2R3-MYBs have different and even opposing functions. Subgroup 4, including *AtMYB4*, *AtMYB32*, *AtMYB7*, *AtMYB3*, is known for encoding transcription factors with repressive functions (Jin *et al.*, 2000; LaFountain & Yuan, 2021). *AtMYB123* and *AtMYB5* belong to subgroup 5 and regulate proanthocyanidin biosynthesis in *Arabidopsis thaliana* (Baudry *et al.*, 2004; Xu *et al.*, 2014). MYBs assigned to subgroup 6 (*AtMYB113, AtMYB114, AtMYB90, AtMYB75*) are involved in controlling the anthocyanin biosynthesis in *A. thaliana* (Borevitz *et al.*, 2000; Marin & Pucker, 2023) whereas *AtMYB111, AtMYB12 and AtMYB11* (subgroup 7) are regulating the production of flavonols (Stracke *et al.*, 2007; Dubos *et al.*, 2010). Over the last years, various orthologous of the well characterized MYBs in *A. thaliana* have been identified across plant species which enables a broader understanding of regulatory mechanisms in the plant kingdom. For example, *At*MYB123 orthologs have been reported to control the anthocyanin biosynthesis in addition to the proanthocyanidin biosynthesis (Martínez-Rivas *et al.*, 2023) and orthologs of the *At*MYB5 have been reported to activate the anthocyanin biosynthesis in strawberry (Jiang *et al.*, 2023). This suggests that the generally well conserved regulation of the anthocyanin biosynthesis might be subject to lineage-specific differences. Given that *A. thaliana* does not have colorful fruits and flowers, it is likely that many secrets of the anthocyanin regulation can only be discovered by exploring non-model species.

Flavonoid Biosynthesis

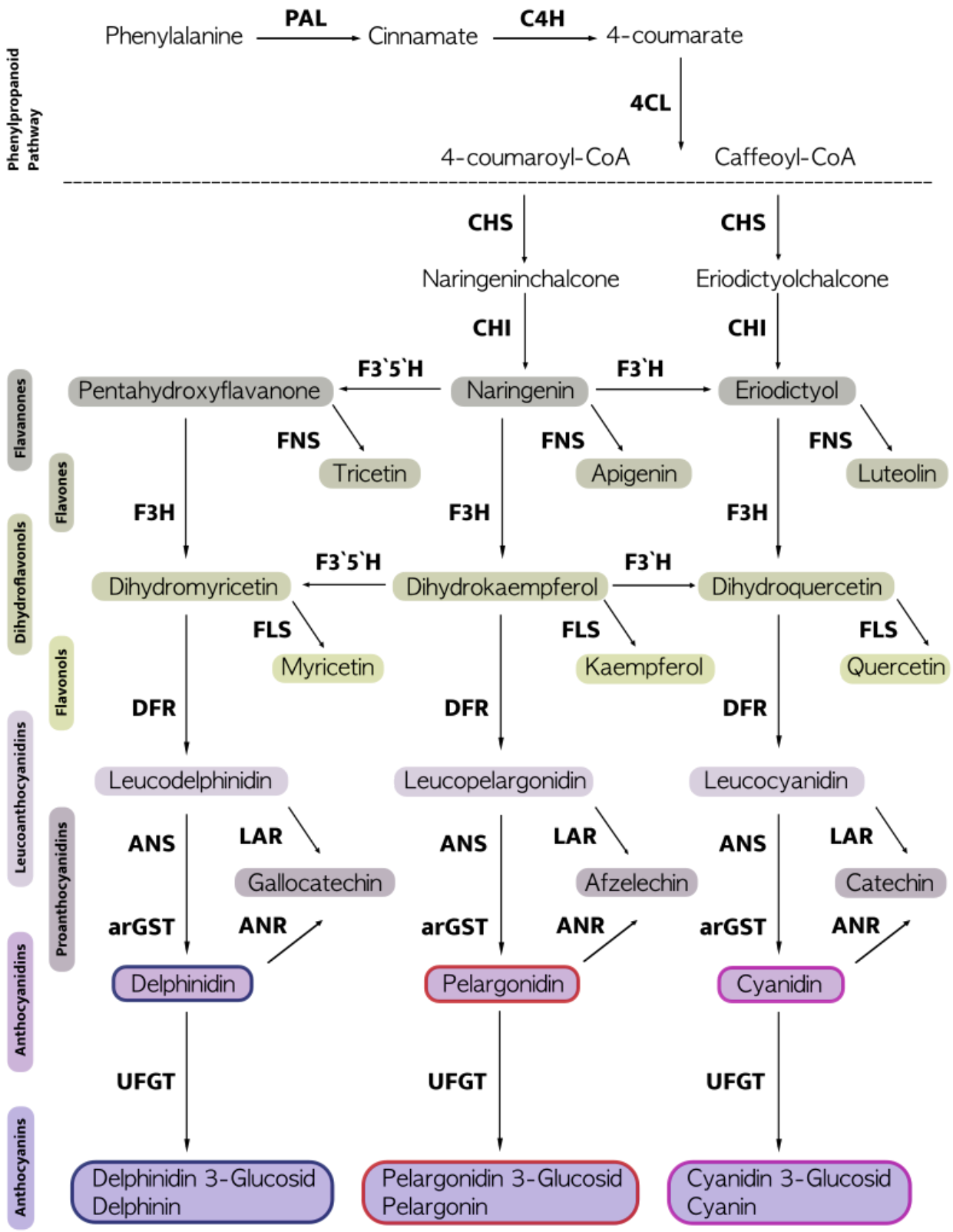

Figure 2: Schematic illustration of the phenylpropanoid and flavonoid biosynthesis pathway including the reactions leading to anthocyanins. PAL (Phenylalanine ammonia lyase), C4H (Cinnamate 4-hydroxylase), 4CL (4-coumarate-CoA ligase), CHS (Chalcone synthase), CHI (Chalcone isomerase), FNS (Flavone synthase), F3H (Flavanone 3-hydroxylase), F3'H (Flavonoid 3'-hydroxylase), F3'5'H (Flavonoid 3',5'-hydroxylase), FLS (Flavonol synthase), DFR (Dihydroflavonol 4-reductase), LAR (Leucoanthocyanidin reductase), ANR (Anthocyanidin reductase), ANS (Anthocyanidin synthase), arGST (anthocyanin-related Glutathione S-transferase), UFGT (UDP-glucose: flavonoid-3-O-glucosyltransferase).

Previous studies investigated the pigment composition of *Victoria cruziana* and identified four anthocyanins (two cyanidin and two delphinidin derivatives) in the flower (Wu *et al.*, 2018). An increase in anthocyanin content across the blooming days correlates with the intensifying pigmentation and supports the role of anthocyanins as the predominant flower colorants (Wu *et al.*, 2018).

To identify the genes involved in the anthocyanin biosynthesis and its regulation in *V. cruziana*, we performed long-read sequencing of the genome and transcriptome. Here, we provide a genome sequence of *V. cruziana* and a corresponding annotation supported by RNA-seq and direct RNA sequencing that serve as the basis for metabolic pathway explorations. Additionally, we conducted RNA-seq experiments to identify differentially expressed genes between white and pinkish flower stages of *V. cruziana*. Specifically, we aimed to unravel the genetic network underlying the flower color transition from white to pinkish.

Material and Methods

The genome of *V. cruziana* was sequenced with Oxford Nanopore Technologies (ONT) long reads to generate a high quality genome sequence that served as basis for the transcriptomics investigation based on direct RNA sequencing and RNA-seq (**Figure 3**).

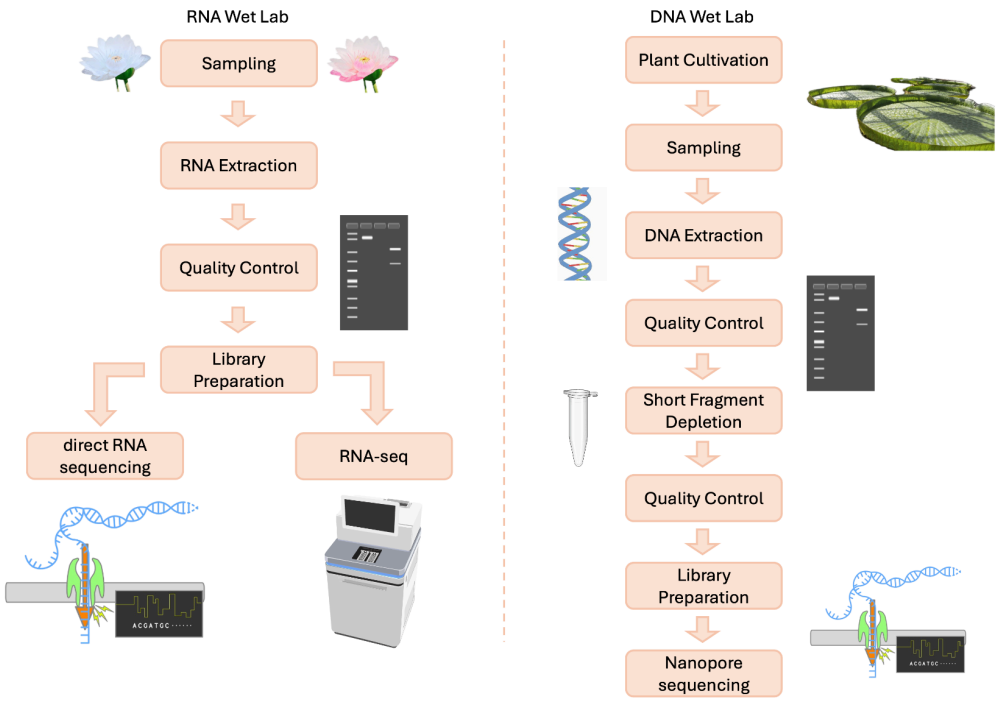

Figure 3: Workflow for the RNA-seq and long read genome sequencing project of *Victoria cruziana.* The left side of the figure illustrates the direct RNA sequencing and RNA-seq workflow targeting *V. cruziana* flowers, while the right side depicts the conducted long read genome sequencing workflow.

Plant cultivation, DNA extraction and RNA extraction

For genomic DNA extraction, a *V. cruziana* plant (XX-0-BRAUN-7477852) was cultivated at 28 °C in a 500 L container to facilitate repeated sampling of young leaves and immediate processing of material. Light was provided for 16 hours per day with two LEDs (Niello®, 300W) placed 1 m above the water surface. High molecular weight DNA extraction and quality assessment were performed based on a previously developed CTAB protocol optimized for high molecular weight DNA extraction from plants (Siadjeu *et al.*, 2020; Wolff *et al.*, 2023). In brief, fresh leaf material was ground in liquid nitrogen and the fine powder subjected to the CTAB-based DNA extraction procedure. RNase treatment in the CTAB-TE buffer was performed overnight at room temperature. After initial quality assessment via NanoDrop and agarose gel, short DNA fragments were depleted with the Short Read Eliminator kit. DNA quantification prior to library preparation for the nanopore sequencing was performed via Qubit.

For transcriptome analysis, young *V. cruziana* petals at the white and pink coloring stage were collected. Sampling was conducted in two consecutive nights during July 2023 in a dedicated *Victoria* glasshouse located in the Botanical Garden of TU Braunschweig (52.27043343004546, 10.5336707275785). Inner petals of the same flower at both color stages were carefully detached, immediately frozen in liquid nitrogen, and stored at -70 °C prior to RNA extraction. The initial sampling occurred on July 19th, 2023, at 10:50 PM. Four samples were taken from the not yet fully opened white flower. The weather on this day was cloudy. The subsequent day,

July 20th 2023, had cloudy and sunny conditions, and five samples were collected from the pink blooming flower. The ambient temperature in the greenhouse was 24 °C during the day and 22 °C at night. The ventilation temperature remained constant at 34 °C, while humidity fluctuated between 80 % and 100 %. The plants did not receive any artificial light exposure. For irrigation, tap water with a temperature of 27 °C was used. The plants were cultivated in a pool with a depth of 1.07 meters and a width of 8.8 m x 5.5 m, with a substrate ratio of 1:1 (compost to peat). Each liter of water received 6 grams of Osmocote Pro and 6 grams of lime as fertilizer.

Total RNA was extracted with the RNA plant and fungi kit (Macherey & Nagel), together with a DNase treatment according to the manufacturer's instructions. Cell homogenization was performed using the Bertin Ribolyser at 6000 RPM with three cycles of 30 seconds each, with a 30 second break between cycles, using 3 mm steel beads. The extracted RNA quality and quantity was determined by a NanoDrop measurement and an agarose gel. The RNA was stored at -70 °C until it was sent for RNA-seq. Paired-end RNA-seq (2 x 150 nt) was conducted using Illumina NovaSeq 6000 (BMKGene).

Genome sequencing

Per library, 1 µg of high molecular weight DNA was utilized following ONT's SQK-LSK109 protocol. Priming of R9.4.1 flow cells and washing steps between sequencing runs were performed according to ONT's protocols (EXP-FLP002, EXP-WSH004). Sequencing was performed on a MinION Mk1B. FAST5 files were collected and subjected to basecalling using Guppy v6.4.6+ae70e8f with the dna_r9.4.1_450bps_hac model and a minimum quality score of 9.

Direct RNA sequencing

RNA for direct RNA sequencing was extracted from one white petal with the Monarch Total RNA miniprep Kit (New England BioLabs) combined with the manufacturer specific DNase treatment. Homogenization was performed using the Bertin Ribolyser at 6500 RPM with two cycles of 30 seconds each, with a 10 second break between cycles. The extracted RNA quality and quantity was determined by NanoDrop measurement and additional Qubit measurement before conducting sequencing. Direct RNA sequencing was performed on a MinION Mk1B using an R9.4.1 flow cell and ONT's direct RNA sequencing kit SQK-RNA002 following the suppliers' protocol. Sequencing data were collected in FAST5 files and subjected to basecalling using Guppy v6.4.6+ae70e8f with the rna_r9.4.1_70bps_hac model and a minimum quality score of 7. Reads generated by direct RNA sequencing were aligned to the *V. cruziana* genome sequence (VB02) with minimap2 (v2.17 r941) using the following parameters: -ax splice -uf -k14 (Li, 2018). Sorting of the resulting SAM file and conversion into BAM was conducted with samtools v1.10 (Li *et al.*, 2009). Mappings to the genome sequence were visualized in the context of the structural annotation with Integrative Genomics Viewer (IGV) v2.17.4 (Robinson *et al.*, 2011).

Direct RNA sequencing data was utilized to validate the structure of genes of interest through manual inspection of the long read alignments.

Genome sequence assembly

An initial genome sequence assembly (VB01) was generated with all ONT reads using Shasta v0.10.0 (Shafin *et al.*, 2020) with the config set to 'Nanopore-May2022'. A second assembly (VB02) was generated with NextDenovo2 v2.5.2 (Hu *et al.*, 2024) using the config file that is presented in Additional File 1. Whitespaces in contig names were removed with "sed 's, ,_,g' *asm.fasta > asm.rm-wht-spc.fasta".* General assembly statistics like assembly size, number of contigs, and N50 were calculated with contig_stats3.py (Meckoni *et al.*, 2023). A completeness assessment was conducted with BUSCO v5.7.1 (Simão *et al.*, 2015; Manni *et al.*, 2021) using the genome or transcriptome mode according to the input data type and the reference dataset embryophyta_odb10.

Structural annotation

The prediction of gene models was conducted with AUGUSTUS v3.3 (Stanke *et al.*, 2006; Keller *et al.*, 2011) for the assembly VB01 with parameters previously optimized for the detection of non-canonical splice sites (Pucker *et al.*, 2017). For the assembly VB02, BRAKER3 v3.0.8 (Gabriel *et al.*, 2023) was applied with protein hints derived from Viridiplantae.fa (Kuznetsov *et al.*, 2023) and RNA-seq hints. The RNA-seq hints were derived from a mapping of paired-end RNA-seq data generated with HISAT2 v2.2.1 (Kim *et al.*, 2019) with default parameters against the assembly VB02. Samtools v1.10 (using htslib 1.10.2-3ubuntu0.1) (Li *et al.*, 2009) was applied to generate a sorted BAM file which was then passed to BRAKER3. The final structural annotation was largely based on the BRAKER3 results produced for VB02 only complemented with the coding sequence of an anthocyanin synthase (*ANS*) gene model lifted from the VB01 annotation. The derived polypeptide sequence of ANS was compared against the VB02 assembly via tBLASTn v2.15.0+ (Gertz *et al.*, 2006) to locate this gene.

Functional annotation and candidate gene identification

Genes associated with the anthocyanin metabolism were identified with three dedicated tools. To identify the structural genes of the flavonoid biosynthesis pathway, an analysis with KIPEs v3.2.4 (Rempel *et al.*, 2023) and the flavonoid baits data set v.3.1.7 was conducted (Additional File 2). Flavonoid biosynthesis controlling MYB transcription factors were annotated using the MYB_annotator v1.0.1 (Pucker, 2022) with parameters described in Additional File 3. The bHLH transcription factors were identified with the bHLH_annotator v1.04 (Thoben & Pucker, 2023) with parameters described in Additional File 4. A general annotation was produced using construct_anno.py and the functional annotation available for *A. thaliana* (Pucker & Iorizzo, 2023).

RNA-seq data analysis

Quantification of transcript abundances (referred to as ‘gene expression’ in the following) was performed using kallisto v0.48.0 (Bray *et al.*, 2016) on the predicted coding sequence. Count tables were merged using a customized python script as well as gene count normalization by transcript per million (TPM) (Pucker & Iorizzo, 2023). These results provided the basis for visualization of gene expression for the structural genes and transcription factors in heat maps. Principal component analysis (PCA) using ggplot2 v3.4.4 library was performed on transcripts with counts >3. For initial analysis of transcriptome data, DESeq2 v1.42.0 (Love *et al.*, 2014) using BiocManager v1.30.22 and matrixStats v1.2.0 was executed in R v2023.09.1+494. Gene candidates were determined through the described functional annotation and by sequence alignment using DIAMOND v2.1.9 (Buchfink *et al.*, 2015, 2021). Visualization of final gene candidates with predicted annotation and z-score normalization was conducted using customized python scripts (https://github.com/bpucker/vici) based on seaborn v0.13.2 (Waskom, 2021).

Phylogenetic analysis

For phylogenetic classifications, sequences of *V. cruziana* MYB candidates in subgroup 5 and subgroup 6 were aligned with the bait sequences of the MYB_annotator (Pucker, 2022) and KcMYB1 (Huang *et al.*, 2022) via MAFFT v7.526 (Katoh & Standley, 2013). IQ-TREE v2.0.7 (Nguyen *et al.*, 2015; Minh *et al.*, 2020) was applied to infer a maximum likelihood tree with the parameters -alrt 1000 -bb 1000. The final tree file was visualized in iTOL (Letunic & Bork, 2021).

Results

Assembly results

A completeness assessment revealed 92.4% and 93.6% complete BUSCO genes for the VB01 (Shasta) and VB02 (NextDenovo2) assembly, respectively. Due to a higher completeness, VB02 was selected as the representative assembly. VB02 has a size of 3.54 Gbp comprising 837 contigs with an N50 length of 14.3 Mbp. In total, 49,301 protein coding genes were annotated in VB02. The assembled genome sequence and the corresponding annotation are available via LeoPARD (https://doi.org/10.24355/dbbs.084-202406040628-0).

Structural genes of the anthocyanin biosynthesis are differentially expressed between white and pink petals

A principal component analysis (PCA) of the transcriptomic data revealed distinct clusters of the replicates derived from white flower and pink flower samples, respectively, supporting clear differences between the two groups and homogeneity within each group (Additional File 5). To understand the molecular basis of the flower color variations in *V. cruziana*, structural genes required for anthocyanin biosynthesis were identified. Expression profiles of the best candidates were analyzed with respect to the white and pink flower color (**Figure 4**). The majority of genes required for anthocyanin biosynthesis (*F3'H*, *F3'5'H*, *DFR*, *ANS*, *arGST*, and *UFGT*) exhibited higher expression levels in anthocyanin-pigmented pink petals compared to white petals (Additional File 6). Genes encoding enzymes of the general phenylpropanoid pathway (*PAL*, *C4H* and *4CL*) displayed increased expression levels in the pink petals. The expression pattern of *CHS* and *CHI* matches better to the flavonol biosynthesis gene *FLS* than to the anthocyanin biosynthesis gene *DFR*, i.e., *CHS* and *CHI* exhibited a higher expression within the white petals. To convert flavanones into dihydroflavonols, *F3H* is required and showed increased expression levels in the samples taken from pinkish petals. Branching reactions towards delphindin or cyanidin are catalyzed by F3'5'H and F3'H, respectively. Both encoding genes are present in *V. cruziana* and exhibited a higher activity in pinkish petals. Dihydroflavonols are converted to the precursors of anthocyanidins, the leucoanthocyanidins, by DFR. Higher expression levels of *DFR* were observed in pinkish petals compared to white petals. Afterwards, the production of the anthocyanidins requires ANS and arGST. These enzymes catalyze the conversion to the anthocyanidins delphinidin and cyanidin in *V. cruziana* and both corresponding genes were more active in the pink petals.

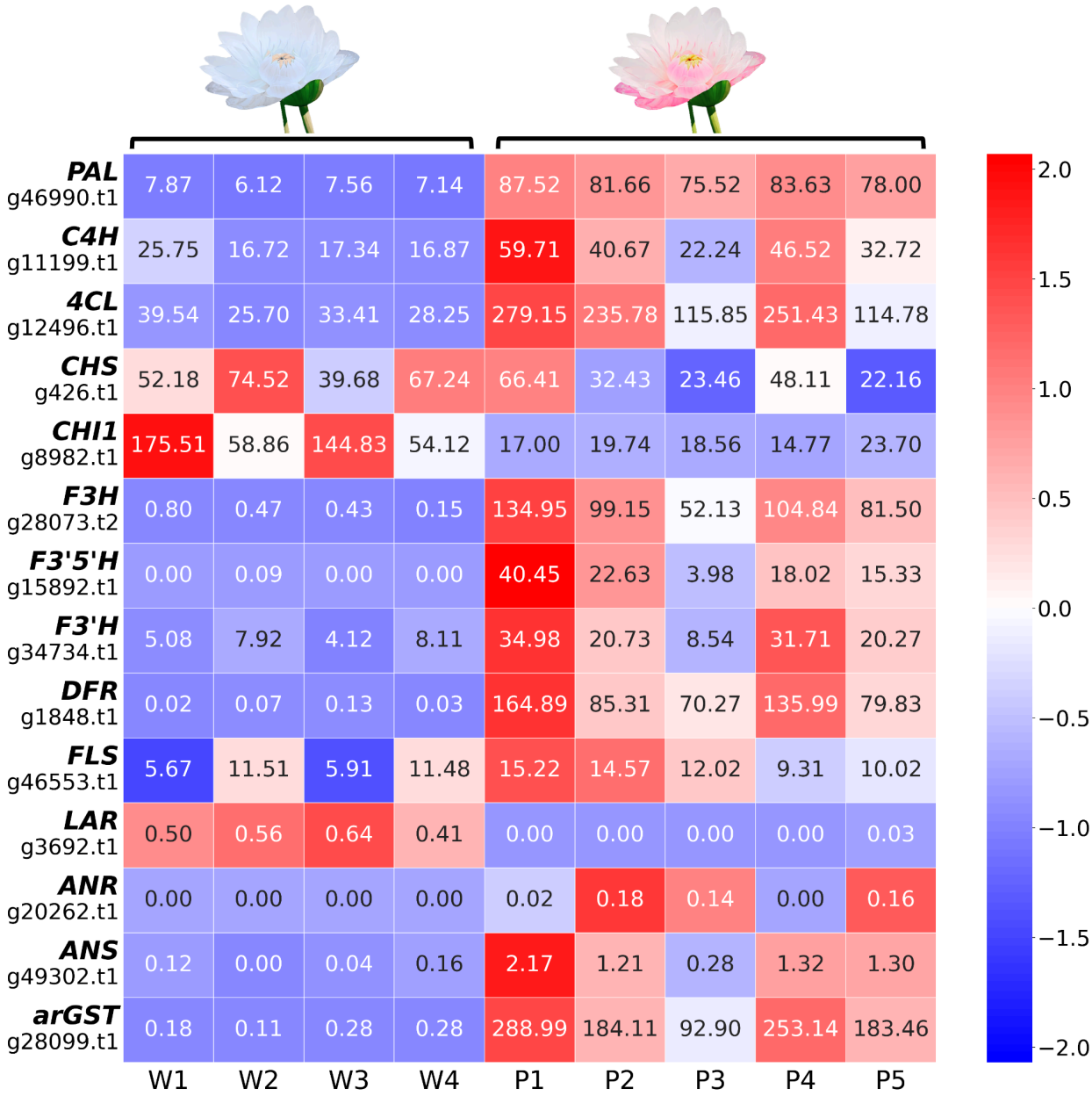

Figure 4: Expression pattern of structural genes associated with the flavonoid biosynthesis in *V. cruziana* flowers. *PAL* (*Phenylalanine ammonia lyase*), *C4H* (*Cinnamate 4-hydroxylase*), *4CL* (*4-coumarate-CoA ligase*), *CHS* (*Chalcone synthase*), *CHI* (*Chalcone isomerase*), *F3H* (*Flavanone 3-hydroxylase*), *F3'H* (*Flavonoid 3'-hydroxylase*), *F3'5'H* (*Flavonoid 3',5'-hydroxylase*), *FLS* (*Flavonol synthase*), *DFR* (*Dihydroflavonol 4-reductase*), *LAR* (*Leucoanthocyanidin reductase*), *ANR* (*Anthocyanidin reductase*), *ANS* (*Anthocyanidin synthase*), *arGST* (*anthocyanin-related Glutathione S-transferase*). Expression values are z-score normalized TPM values. Values given in the individual fields of the heatmap are TPM values. W1-W4 are samples taken from the white flower stage, P1-P5 are samples taken from the pinkish flower stage.

Expression of anthocyanin biosynthesis regulators differs between white and pinkish flowers

Anthocyanin biosynthesis is regulated by various transcription factors, specifically the MYB family, the bHLH family, and specific WD40 proteins. These TFs are capable of forming complexes, thus activating anthocyanin production in plants. After identification of the required structural genes, best candidates for potential transcriptional regulators were analyzed. As one transcription factor with activatory properties for proanthocyanidin production, we identified a subgroup 5 MYB (*VcrMYBSG5*) which exhibited higher expression levels in pink compared to white petals. Further, we annotated a candidate belonging to MYB subgroup 6 (*Vcr*MYBSG6), described for the positive regulation of anthocyanin biosynthesis. This *VcrMYBSG6* was only expressed in pink petals in *V. cruziana*. Additionally we identified *VcrMYBSG4*, which could be an orthologue of the anthocyanin biosynthesis repressing regulator *AtMYB004*. Expression levels increased within the white petals thus showing a pattern opposite to anthocyanin accumulation. Additionally, *VcrTT8* belonging to the bHLHs and *VcrTTG1* representing the WD40 proteins, were predicted in the functional annotation process. Both corresponding genes indicated higher expression levels in pink petals compared to white (**Figure 5**).

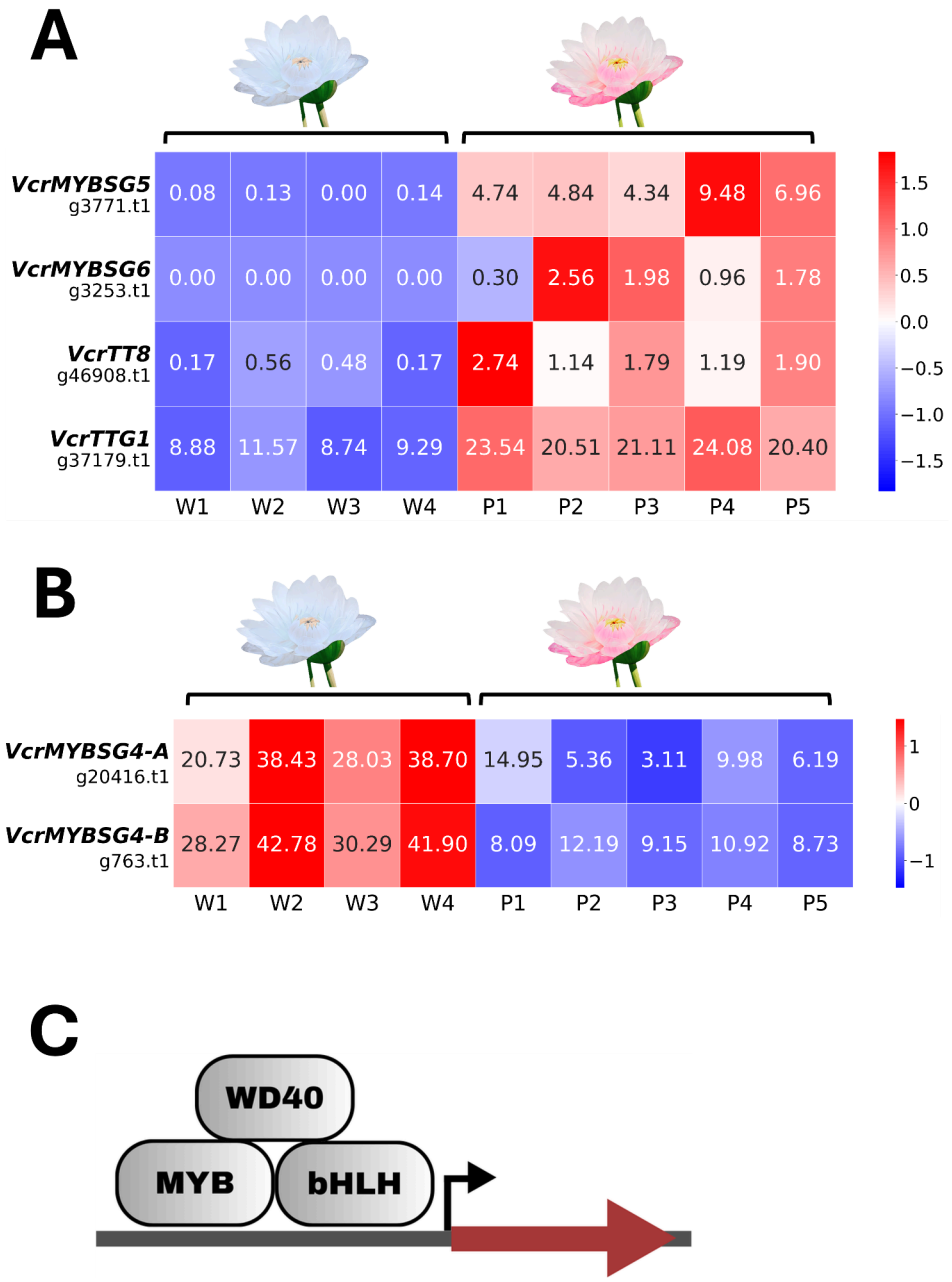

Figure 5: (A) Expression pattern of transcription factor genes putatively required for anthocyanin biosynthesis activation in *V. cruziana*. *Vcr*MYBSG5 and *Vcr*MYBSG6 expression patterns align with the anthocyanin accumulation. *VcrTT8* represents the associated bHLH transcription factor and *VcrTTG1* edcodes a WD40 protein required for anthocyanin biosynthesis activation. (B) *VcrMYBSG4-A* and *VcrMYBSG4-B* probably possess anthocyanin biosynthesis repressing activities and show an expression pattern opposite to anthocyanin accumulation. (C) MYB, associated bHLH transcription factor and WD40 protein forming a transcription factor complex. The WD40 protein is believed to serve as a scaffolding protein connecting MYB and bHLH. Expression values are z-score normalized TPM values. Values

given in the individual fields of the heatmap are TPM values. W1-W4 are samples taken from the white flower stage, P1-P5 are samples taken from the pinkish flower stage.

To explore other MYB transcription factors that have been reported in the context of anthocyanin biosynthesis in species closely related to *V.cruiziana*, a search for the ortholog of the recently published KcMYB1 from *Kadsura coccinea* was conducted. KcMYB1 belongs to the MYB subgroup 6 and overexpression in *Nicotiana benthamiana* resulted in anthocyanin accumulation (Huang *et al.*, 2022). Using functional annotation and a phylogenetic analysis, *g3771* was classified as a subgroup 5 R2R3-MYB, while KcMYB1 was located in the subgroup 6 clade close to the *V. cruziana* genes g3252-g3257 (Additional file 7). However, only g3253 (*VcrMYBSG6*) was substantially supported by aligned RNA-seq reads, indicating a promising SG6 MYB in *V. cruziana*. In contrast, genes without RNA-seq support might be incomplete gene models and have not been picked up as differentially expressed between white and pinkish petal samples (Additional file 7, Additional file 8).

Discussion

Differential expression of flavonoid biosynthesis genes explains anthocyanin production in pink petals of *V. cruziana*

The continuous genome sequence of *V. cruziana* and the corresponding annotation enabled the identification of genes encoding relevant proteins and transcription factors required for the anthocyanin biosynthesis (**Figure 6**). The differential expression patterns compared between white and pink petals align with the active anthocyanin production in the pink flower. Besides higher expression levels of genes belonging to the phenylpropanoid pathway within pink petals, the first committed genes of the flavonoid biosynthesis (*CHS*, *CHI*) were higher expressed in the white flower. Additionally, *FLS* and *LAR* expression patterns did not match the expression level patterns of the other genes. The *FLS* pattern can be due to its activity in flavonol biosynthesis which is active in the white petals, supported by Wu *et al.,* 2018 who identified flavonols in white *V. cruziana* flowers. The *ANR* expression pattern suggested some ongoing proanthocyanidin biosynthesis in white petals. Future metabolic studies are needed to confirm the presence of proanthocyanidins.
Furthermore, *F3'5'H* as the branching point towards delphinidin production exhibited increased expression levels in pink petals. This finding is surprising as delphinidin is one anthocyanidin required for blue flower colors which has not been observed in *Victoria spp*. flowers. However, delphinidin derivatives were identified besides cyanidin derivatives as the major anthocyanins in the *V. cruziana* and *V. amazonica* (Wu *et al.*, 2018). Perhaps blue colors are lacking due to overlay with cyanidin derivatives. As *ANS* and *arGST* encode major enzymes in anthocyanin biosynthesis, their expression pattern matches the anthocyanin accumulation in pink petals. Interestingly the *arGST* expression is significantly higher in pink petals compared to the

corresponding *ANS* expression. The substantially higher expression of *arGST* compared to *ANS* might ensure the channeling of metabolites towards anthocyanins.

*Vcr*MYBSG5 and *Vcr*MYBSG6 may interact with *Vcr*TT8 and *Vcr*TTG1 to activate the anthocyanin biosynthesis in pink flowers of *V. cruziana*

Transcription factors fulfill a crucial role when it comes to orchestrating gene expression. Essential for anthocyanin production in plants are MYB proteins, regulating various branches of the pathway. To understand why anthocyanin pigments are accumulating in flowers, the responsible transcription factor needs to be identified. *VcrMYBSG5*, most likely an *AtMYB123* orthologue, was identified in *V. cruziana. VcrMYBSG6* was identified as an potential orthologue of prominent anthocyanin-regulating MYBs like *MYB75/MYB90/MYB113/MYB114* in *A. thaliana.* Furthermore, four candidates were identified as potential members of the flavonol-regulating MYB subgroup 7. It appears likely that VcrMYBSG6 and potentially VcrMYBSG5 are the main anthocyanin-regulating MYBs in *V. cruziana*. Other gene candidates proposed for subgroup 6 are not supported by petal-derived RNA-seq data which could indicate that they might be active in other plant parts. Alternatively, no further active subgroup 6 MYBs might exist in *V. cruziana*. Previous studies reported a role of subgroup 5 MYB (MYB5, MYB123) in activating the anthocyanin biosynthesis (Martínez-Rivas *et al.*, 2023; Jiang *et al.*, 2023) thus corroborating the putative role of *VcrMYBSG5* in this context. The additional sequence alignment of *KcMYB1* TF (JHA, Additional file 7) from *K. coccinea* further supports the assumption that *VcrMYBSG6* encodes an anthocyanin activating MYB transcription factor as *K. coccinea* belongs to the *Schisandraceae* (Huang *et al.*, 2022), a basal angiosperm lineage like the *Nymphaceaee*. Further the *VcrMYBSG4*, an potential orthologue of the repressor *AtMYB004*, showed higher expression levels in the white compared to the pink flower, suggesting repressive function on anthocyanin production in the white outer petals. This can be further supported by a study, where they identified another AtMYB004 orthologue in *Musa* spp., *MaMYB4*, with repressive function on the anthocyanin biosynthesis genes (Deng *et al.*, 2021). Li et al. observed decreased transcript levels in *CsMYB4a* transgenic plants suggesting a repressive function on the flavonoid biosynthesis pathway (Li *et al.*, 2017). Furthermore, one of the two candidates for the *VcrMYBSG4* might represent another repressor as *AtMYB007* (Fornalé *et al.*, 2014) is homologue to *AtMYB004* (Jin *et al.*, 2000; Stracke *et al.*, 2001).

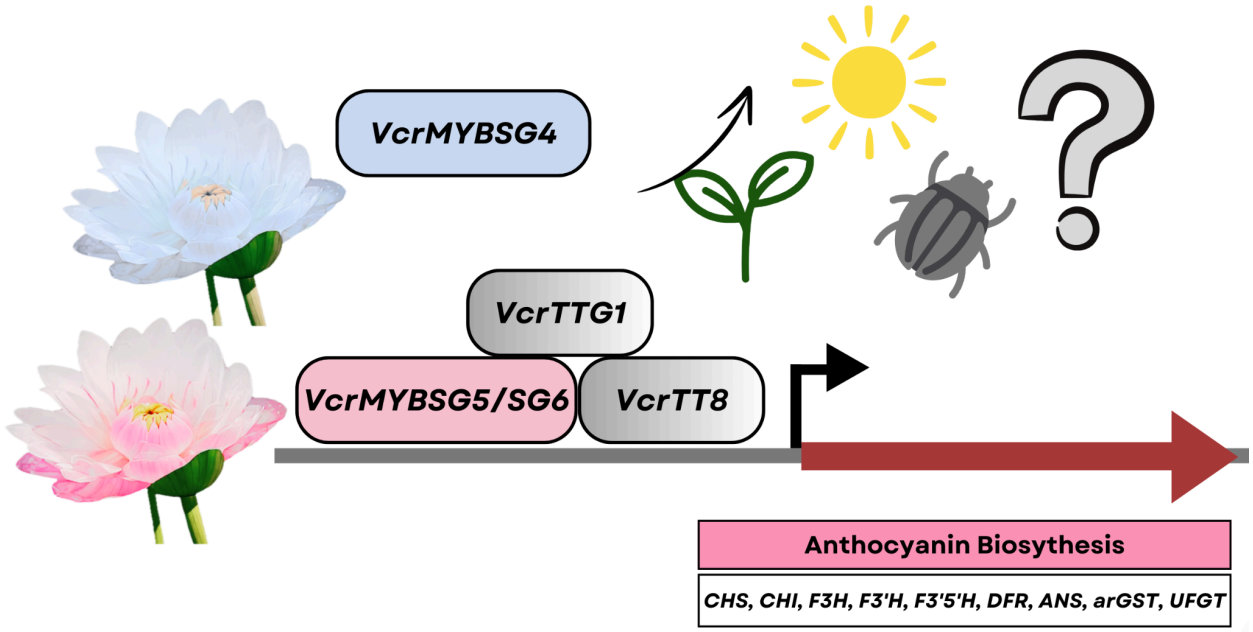

Figure 6: Graphical abstract summarizing the molecular mechanism involved in anthocyanin biosynthesis and potential factors triggering the flower color change of *V.cruziana. VcrMYBSG4* is active in white petals whereas *VcrMYBSG5* and *VcrMYBSG6* are expressed in pink petals. Factors that could trigger anthocyanin biosynthesis, thereby activating TFs and structural genes, might be visitation/pollination, light exposure, or a general developmental program of the plant.

Based on a continuous genome sequence, we successfully identified all structural genes required for anthocyanidin production in *Victoria cruziana*. However, there is potential to improve the structural annotation as 93.6% complete BUSCOs have been discovered. The inclusion of additional RNA-seq hints derived from different plant parts might contribute to further improvements of the structural gene prediction and subsequently result in an improved functional annotation.

Further, we identified *VcrMYBSG5* and *VcrMYBSG6* as potential transcriptional activators of anthocyanin and proanthocyanidin biosynthesis based on co-expression with anthocyanin accumulation. To confirm the anthocyanin biosynthesis activation properties, the genes could be introduced into *A. thaliana myb75* mutants to conduct a complementation experiment. Besides, expression levels of the structural genes as well as transcription factors of *V. cruziana* could be analyzed in leaves and other parts of the plant to further investigate potential TFs activating structural genes in the flavonoid biosynthesis.

To reveal developmental and environmental factors triggering the coloration of *V. cruziana* flowers, future experiments need to control light exposure and pollinator access to the flowers. Since *V. cruziana* is only opening its flowers at night, light triggering the anthocyanin formation inside the flower would have to pass through the outer petals. However, if the signal to close the flower after the first night is the first sunlight of the next day, this could be enough light to induce anthocyanin biosynthesis. Further experiments are needed to fully understand the ecological relevance and environmental factors triggering anthocyanin formation in *V. cruziana*.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Data availability
Genomic sequencing data (PRJEB63973), direct RNA sequencing data (PRJEB63453), and RNA-seq data (PRJEB63973) are available via the European Nucleotide Archive (ENA). The assembled *Victoria cruziana* genome sequence VB02 and the corresponding annotation are available via LeoPARD (https://doi.org/10.24355/dbbs.084-202406040628-0).

Competing interests
MSN and BH received a material donation from Oxford Nanopore Technologies to conduct direct RNA sequencing. Expenses connected to RF's presentation at LodonCalling2024 about her research on medicinal plant genomics were covered by ONT. The other authors declare that they have no competing interests.

Funding
We acknowledge support by the Open Access Publication Funds of Technische Universität Braunschweig. The funders had no influence in study design or the decision to publish this work.

Authors' contributions
MSN, BH, SNM, BP planned the study. MSN, BH, RF, KW, and BP performed sampling, nucleic acid extraction, and ONT sequencing. MSN, SNM, BH, KW and BP conducted bioinformatic analyses. MSN, BH, SNM, and BP interpreted the results and wrote the manuscript. All authors have read the final version of the manuscript and approved its submission.

Acknowledgements
We thank Thorsten Marschall, Michael Kraft, and the entire team operating the Botanical Garden of TU Braunschweig for their excellent technical support. This work was supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A532B, 031A533A, 031A533B, 031A534A, 031A535A, 031A537A, 031A537B, 031A537C, 031A537D, 031A538A). For Figure 3: Microtube-closed, Nanopore and sequencer icon by DBCLS https://togotv.dbcls.jp/en/pics.html is licensed under CC-BY 4.0. We also thank all students participating in our 'Data Literacy in Genome Research' course and members of the research group Plant Biotechnology and Bioinformatics for discussion and support.

References

Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. **2004**. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. *The Plant Journal: For Cell and Molecular Biology* **39**: 366–380. doi: 10.1111/j.1365-313X.2004.02138.x.

Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C. **2000**. Activation Tagging Identifies a Conserved MYB Regulator of Phenylpropanoid Biosynthesis. *The Plant Cell* **12**: 2383–2393.

Bortolotto IM, Amorozo MC de M, Neto GG, Oldeland J, Damasceno-Junior GA. **2015**. Knowledge and use of wild edible plants in rural communities along Paraguay River, Pantanal, Brazil. *Journal of Ethnobiology and Ethnomedicine* **11**: 46. doi: 10.1186/s13002-015-0026-2.

Bray NL, Pimentel H, Melsted P, Pachter L. **2016**. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**: 525–527. doi: 10.1038/nbt.3519.

Buchfink B, Reuter K, Drost H-G. **2021**. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* **18**: 366–368. doi: 10.1038/s41592-021-01101-x.

Buchfink B, Xie C, Huson DH. **2015**. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12**: 59–60. doi: 10.1038/nmeth.3176.

Delph LF, Lively CM. **1989**. THE EVOLUTION OF FLORAL COLOR CHANGE: POLLINATOR ATTRACTION VERSUS PHYSIOLOGICAL CONSTRAINTS IN FUCHSIA EXCORTICATA. *Evolution; International Journal of Organic Evolution* **43**: 1252–1262. doi: 10.1111/j.1558-5646.1989.tb02572.x.

Deng G-M, Zhang S, Yang Q-S, Gao H-J, Sheng O, Bi F-C, Li C-Y, Dong T, Yi G-J, He W-D, *et al.*2021. MaMYB4, an R2R3-MYB Repressor Transcription Factor, Negatively Regulates the Biosynthesis of Anthocyanin in Banana. *Frontiers in Plant Science* **11**.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. **2010**. MYB transcription factors in Arabidopsis. *Trends in Plant Science* **15**: 573–581. doi: 10.1016/j.tplants.2010.06.005.

Dussi MC, Sugar D, Wrolstad RE. **1995**. Characterizing and Quantifying Anthocyanins in Red Pears and the Effect of Light Quality on Fruit Color. *Journal of the American Society for Horticultural Science* **120**: 785–789. doi: 10.21273/JASHS.120.5.785.

Eichenberger M, Schwander T, Hüppi S, Kreuzer J, Mittl PRE, Peccati F, Jiménez-Osés G, Naesby M, Buller RM. **2023**. The catalytic role of glutathione transferases in heterologous anthocyanin biosynthesis. *Nature Catalysis* **6**: 927–938. doi: 10.1038/s41929-023-01018-y.

Fornalé S, Lopez E, Salazar-Henao JE, Fernández-Nohales P, Rigau J, Caparros-Ruiz D. **2014**. AtMYB7, a New Player in the Regulation of UV-Sunscreens in Arabidopsis thaliana. *Plant and Cell Physiology* **55**: 507–516. doi: 10.1093/pcp/pct187.

Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. **2023**. BRAKER3: Fully Automated Genome Annotation Using RNA-Seq and Protein Evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. : 2023.06.10.544449. doi: 10.1101/2023.06.10.544449.

Garcia JE, Hannah L, Shrestha M, Burd M, Dyer AG. **2022**. Fly pollination drives convergence of flower coloration. *New Phytologist* **233**: 52–61. doi: 10.1111/nph.17696.

Gertz EM, Yu Y-K, Agarwala R, Schäffer AA, Altschul SF. **2006**. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biology* **4**: 41. doi: 10.1186/1741-7007-4-41.

Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. **2008**. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *The Plant Journal* **53**: 814–827. doi: 10.1111/j.1365-313X.2007.03373.x.
Gould KS. **2004**. Nature's Swiss Army Knife: The Diverse Protective Roles of Anthocyanins in Leaves. *Journal of Biomedicine and Biotechnology* **2004**: 314–320. doi: 10.1155/S1110724304406147.
Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, Li J, Sandoval JR, Cooper DN, Ye K, *et al.*2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biology* **25**: 107. doi: 10.1186/s13059-024-03252-4.
Huang D, Xu S, Qin Y, Li Y, Ming R, Huang R, Wang J, Tan Y. **2022**. Comparative transcriptomic analysis identifies KcMYB1 as a R2R3-MYB anthocyanin activator in *Kadsura coccinea*. *Plant Science* **324**: 111458. doi: 10.1016/j.plantsci.2022.111458.
Jiang Y, Liu G, Zhang W, Zhang C, Chen X, Chen Y, Yu C, Yu D, Fu J, Chen F. **2021**. Biosynthesis and emission of methyl hexanoate, the major constituent of floral scent of a night-blooming water lily *Victoria cruziana*. *Phytochemistry* **191**: 112899. doi: 10.1016/j.phytochem.2021.112899.
Jiang L, Yue M, Liu Y, Zhang N, Lin Y, Zhang Y, Wang Y, Li M, Luo Y, Zhang Y, *et al.*2023. A novel R2R3-MYB transcription factor FaMYB5 positively regulates anthocyanin and proanthocyanidin biosynthesis in cultivated strawberries (Fragaria × ananassa). *Plant Biotechnology Journal* **21**: 1140–1158. doi: 10.1111/pbi.14024.
Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C. **2000**. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *The EMBO Journal* **19**: 6150–6161. doi: 10.1093/emboj/19.22.6150.
Katoh K, Standley DM. **2013**. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**: 772–780. doi: 10.1093/molbev/mst010.
Keller O, Kollmar M, Stanke M, Waack S. **2011**. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics (Oxford, England)* **27**: 757–763. doi: 10.1093/bioinformatics/btr010.
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. doi: 10.1038/s41587-019-0201-4.
Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. **2023**. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research* **51**: D445–D451. doi: 10.1093/nar/gkac998.
LaFountain AM, Yuan Y-W. **2021**. Repressors of anthocyanin biosynthesis. *New Phytologist* **231**: 933–949. doi: 10.1111/nph.17397.
Letunic I, Bork P. **2021**. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49**: W293–W296. doi: 10.1093/nar/gkab301.
Li S. **2014**. Transcriptional control of flavonoid biosynthesis. *Plant Signaling & Behavior* **9**: e27522. doi: 10.4161/psb.27522.
Li H. **2018**. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100. doi: 10.1093/bioinformatics/bty191.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. **2009**. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079. doi: 10.1093/bioinformatics/btp352.
Li M, Li Y, Guo L, Gong N, Pang Y, Jiang W, Liu Y, Jiang X, Zhao L, Wang Y, *et al.*2017. Functional Characterization of Tea (Camellia sinensis) MYB4a Transcription Factor Using an Integrative Approach. *Frontiers in Plant Science* **8**. doi: 10.3389/fpls.2017.00943.
Li S, Wang W, Gao J, Yin K, Wang R, Wang C, Petersen M, Mundy J, Qiu J-L. **2016**. MYB75 Phosphorylation by MPK4 Is Required for Light-Induced Anthocyanin Accumulation in

Arabidopsis[OPEN]. *The Plant Cell* **28**: 2866–2883. doi: 10.1105/tpc.16.00130.
Love MI, Huber W, Anders S. **2014**. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550. doi: 10.1186/s13059-014-0550-8.
Maier A, Schrader A, Kokkelink L, Falke C, Welter B, Iniesto E, Rubio V, Uhrig JF, Hülskamp M, Hoecker U. **2013**. Light and the E3 ubiquitin ligase COP1/SPA control the protein stability of the MYB transcription factors PAP1 and PAP2 involved in anthocyanin accumulation in Arabidopsis. *The Plant Journal: For Cell and Molecular Biology* **74**: 638–651. doi: 10.1111/tpj.12153.
Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. **2021**. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* **38**: 4647–4654. doi: 10.1093/molbev/msab199.
Marin FMR, Pucker B. **2023**. Genetic factors explaining anthocyanin pigmentation differences. : 2023.06.05.543820. doi: 10.1101/2023.06.05.543820.
Martínez-Rivas FJ, Blanco-Portales R, Serratosa MP, Ric-Varas P, Guerrero-Sánchez V, Medina-Puche L, Moyano L, Mercado JA, Alseekh S, Caballero JL, *et al.*2023. FaMYB123 interacts with FabHLH3 to regulate the late steps of anthocyanin and flavonol biosynthesis during ripening. *The Plant Journal* **114**: 683–698. doi: 10.1111/tpj.16166.
Meckoni SN, Nass B, Pucker B. **2023**. Phylogenetic placement of Ceratophyllum submersum based on a complete plastome sequence derived from nanopore long read sequencing data. *BMC Research Notes* **16**: 187. doi: 10.1186/s13104-023-06459-z.
Merzlyak MN, Chivkunova OB. **2000**. Light-stress-induced pigment changes and evidence for anthocyanin photoprotection in apples. *Journal of Photochemistry and Photobiology. B, Biology* **55**: 155–163. doi: 10.1016/s1011-1344(00)00042-7.
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. doi: 10.1093/molbev/msaa015.
Moore MR, Jameson ML. **2013**. Floral associations of cyclocephaline scarab beetles. *Journal of Insect Science* **13**: 100. doi: 10.1673/031.013.10001.
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. **2015**. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**: 268–274. doi: 10.1093/molbev/msu300.
Oberrath R, Böhning-Gaese K. **1999**. Floral color change and the attraction of insect pollinators in lungwort (Pulmonaria collina). *Oecologia* **121**: 383–391. doi: 10.1007/s004420050943.
Prouse MB, Campbell MM. **2012**. The interaction between MYB proteins and their target DNA binding sites. *Biochimica Et Biophysica Acta* **1819**: 67–77. doi: 10.1016/j.bbagrm.2011.10.010.
Pucker B. **2022**. Automatic identification and annotation of MYB gene family members in plants. *BMC Genomics* **23**: 220. doi: 10.1186/s12864-022-08452-5.
Pucker B, Holtgräwe D, Weisshaar B. **2017**. Consideration of non-canonical splice sites improves gene prediction on the Arabidopsis thaliana Niederzenz-1 genome sequence. *BMC Research Notes* **10**: 667. doi: 10.1186/s13104-017-2985-y.
Pucker B, Iorizzo M. **2023**. Apiaceae FNS I originated from F3H through tandem gene duplication. *PLOS ONE* **18**: e0280155. doi: 10.1371/journal.pone.0280155.
Rempel A, Choudhary N, Pucker B. **2023**. KIPEs3: Automatic annotation of biosynthesis pathways. *PLOS ONE* **18**: e0294342. doi: 10.1371/journal.pone.0294342.
Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. **2011**. Integrative Genomics Viewer. *Nature biotechnology* **29**: 24–26. doi: 10.1038/nbt.1754.
Ruxton GD, Schaefer HM. **2016**. Floral colour change as a potential signal to pollinators. *Current Opinion in Plant Biology* **32**: 96–100. doi: 10.1016/j.pbi.2016.06.021.

Seymour RS, Matthews PGD. **2006**. The Role of Thermogenesis in the Pollination Biology of the Amazon Waterlily Victoria amazonica. *Annals of Botany* **98**: 1129–1135. doi: 10.1093/aob/mcl201.
Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, Armstrong J, Tigyi K, Maurer N, Koren S, *et al.*2020. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nature Biotechnology* **38**: 1044–1053. doi: 10.1038/s41587-020-0503-6.
Siadjeu C, Pucker B, Viehöver P, Albach DC, Weisshaar B. **2020**. High Contiguity de novo Genome Sequence Assembly of Trifoliate Yam (Dioscorea dumetorum) Using Long Read Sequencing. *Genes* **11**: 274. doi: 10.3390/genes11030274.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. **2015**. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212. doi: 10.1093/bioinformatics/btv351.
Smith LT, Magdalena C, Przelomska NAS, Pérez-Escobar OA, Melgar-Gómez DG, Leitch IJ, Dodsworth S, Maurin O, Antonelli A, Monro AK. **2022**. Revised Species Delimitation in the Giant Water Lily Genus Victoria (Nymphaeaceae) Confirms a New Species and Has Implications for Its Conservation. *Frontiers in Plant Science* **13**. doi: 10.3389/fpls.2022.883151.
Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. **2006**. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* **34**: W435–W439. doi: 10.1093/nar/gkl200.
Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B. **2007**. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabidopsis thaliana seedling. *The Plant Journal: For Cell and Molecular Biology* **50**: 660–677. doi: 10.1111/j.1365-313X.2007.03078.x.
Stracke R, Werber M, Weisshaar B. **2001**. The R2R3-MYB gene family in Arabidopsis thaliana. *Current Opinion in Plant Biology* **4**: 447–456. doi: 10.1016/S1369-5266(00)00199-0.
Sun SG, Liao K, Xia J, Guo YH. **2005**. Floral colour change in Pedicularis monbeigiana (Orobanchaceae). *Plant Systematics and Evolution* **255**: 77–85. doi: 10.1007/s00606-005-0348-y.
Takos AM, Jaffé FW, Jacob SR, Bogs J, Robinson SP, Walker AR. **2006**. Light-Induced Expression of a MYB Gene Regulates Anthocyanin Biosynthesis in Red Apples. *Plant Physiology* **142**: 1216–1232. doi: 10.1104/pp.106.088104.
Tanaka Y, Sasaki N, Ohmiya A. **2008**. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *The Plant Journal* **54**: 733–749. doi: https://doi.org/10.1111/j.1365-313X.2008.03447.x.
Thoben C, Pucker B. **2023**. Automatic annotation of the bHLH gene family in plants. *BMC Genomics* **24**: 780. doi: 10.1186/s12864-023-09877-2.
Timoneda A, Feng T, Sheehan H, Walker-Hale N, Pucker B, Lopez-Nieves S, Guo R, Brockington S. **2019**. The evolution of betalain biosynthesis in Caryophyllales. *New Phytologist* **224**: 71–85. doi: 10.1111/nph.15980.
Voronova A, Baltimore D. **1990**. Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 4722–4726. doi: 10.1073/pnas.87.12.4722.
Wang H, Cao G, Prior RL. **1997**. Oxygen Radical Absorbing Capacity of Anthocyanins. *Journal of Agricultural and Food Chemistry* **45**: 304–309. doi: 10.1021/jf960421t.
Waskom M. **2021**. seaborn: statistical data visualization. *Journal of Open Source Software* **6**: 3021. doi: 10.21105/joss.03021.
Weiss MR. **1991**. Floral colour changes as cues for pollinators. *Nature* **354**: 227–229. doi: 10.1038/354227a0.
Willmer P, Stanley DA, Steijven K, Matthews IM, Nuttman CV. **2009**. Bidirectional flower color and shape changes allow a second opportunity for pollination. *Current biology: CB* **19**:

919–923. doi: 10.1016/j.cub.2009.03.070.
Winkel-Shirley B. **2001**. Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. *Plant Physiology* **126**: 485–493. doi: 10.1104/pp.126.2.485.
Wolff K, Friedhoff R, Schwarzer F, Pucker B. **2023**. Data literacy in genome research. *Journal of Integrative Bioinformatics* **20**. doi: 10.1515/jib-2023-0033.
Wu Q, Li P-C, Zhang H-J, Feng C-Y, Li S-S, Yin D-D, Tian J, Xu W-Z, Wang L-S. **2018**. Relationship between the flavonoid composition and flower colour variation in Victoria. *Plant Biology (Stuttgart, Germany)* **20**: 674–681. doi: 10.1111/plb.12835.
Xu W, Grain D, Bobet S, Le Gourrierec J, Thévenin J, Kelemen Z, Lepiniec L, Dubos C. **2014**. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in Arabidopsis seed. *The New Phytologist* **202**: 132–144. doi: 10.1111/nph.12620.
Zini LM, Galati BG, Gotelli M, Zarlavsky G, Ferrucci MS. **2019**. Carpellary appendages in Nymphaea and Victoria (Nymphaeaceae): evidence of their role as osmophores based on morphology, anatomy and ultrastructure. *Botanical Journal of the Linnean Society* **191**: 421–439. doi: 10.1093/botlinnean/boz078.

Supplementary

Additional File 1: Configuration file for NextDenovo2 as it was used to assemble *Victoria cruziana* VB02.

Additional File 2: Documentation of the KIPEs3 analysis.

Additional File 3: Documentation of the MYB_annotator analysis.

Additional File 4: Documentation of the bHLH_annotator analysis.

Additional File 5: Principal Component Analysis of the RNA-seq samples confirming the clustering of biological replicates in their respective groups.

Additional File 6: Table containing the DESeq2 results of the white vs. pinkish flower stage comparison.

Additional File 7: Phylogenetic classification of VcrMYB candidates

Additional File 8: Table containing TPM data