https://doi.org/10.1093/plphys/kiad363 PLANT PHYSIOLOGY 2023: 193: 1213–1226

Plant Physiology®

Research Article

Transcription factor NnMYB5 controls petal color by regulating *GLUTATHIONE S-TRANSFERASE2* in *Nelumbo nucifera*

Juan Liu,[1,2,†] Yuxin Wang,[1,3,†] Xianbao Deng,[1,2] Minghua Zhang,[1,3] Heng Sun,[1,2] Lei Gao,[1] Heyun Song,[1,3] Jia Xin,[1,3] Ray Ming,[4] Dong Yang[1,2,*] and Mei Yang[1,2,*]

1 Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China
2 Aquatic Plant Research Center, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China
3 College of Life Science, University of Chinese Academy of Sciences, 19A Yuquanlu, Beijing 100049, China
4 Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China

*Authors for correspondence: yangmei815815@wbgcas.cn (M.Y.), yangdong@wbgcas.cn (D.Y.)
†These authors contributed equally.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is Mei Yang.

Abstract

Lotus (*Nelumbo* spp.) is an important aquatic ornamental genus in the family Nelumbonaceae comprising only 2 species: *Nelumbo lutea* with yellow flowers and *Nelumbo nucifera* with red or white flowers. The petal color variations between these 2 species have previously been associated with the potential activities of FLAVONOL SYNTHASE (FLS) and MYB5. However, the underlying genetic mechanisms of flower color divergence within the *N. nucifera* species remain unclear. Here, quantitative trait locus mapping led to the identification of *MYB5*, a candidate gene controlling petal color in *N. nucifera*. Genotyping of 213 natural lotus accessions revealed an 80 kb presence/absence variant (PAV) of the *NnMYB5* gene that is associated with petal color variation. Transcriptome analysis, dual-luciferase, and yeast 1-hybrid assays showed that NnMYB5 could directly activate the anthocyanin transporter gene *GLUTATHIONE S-TRANSFERASE2* (*NnGST2*). Heterologous expression of *NnGST2* in Arabidopsis (*Arabidopsis thaliana*) and its overexpression in lotus petals induced anthocyanin accumulation. Deletion of the 80 kb PAV within *NnMYB5* inactivated *NnGST2* expression and blocked anthocyanin accumulation in white *N. nucifera* petals. In contrast, the anthocyanin deficiency of *N. lutea* occurred due to pseudogenized *NlMYB5* alleles. Our results establish a regulatory link between NnMYB5 and *NnGST2* in petal anthocyanin accumulation and demonstrate the independent mechanisms controlling flower coloration in *Nelumbo*.

Introduction

Petal color is typically generated by variously accumulated pigments, which predominantly include anthocyanins. Anthocyanins are a group of ubiquitously occurring polyphenolic plant pigments that produce diverse color phenotypes and antioxidant properties against various stresses (Tanaka et al. 2008). The anthocyanin biosynthesis is a dynamic branch of the flavonoid biosynthetic pathway that is catalyzed by a series of highly conserved enzymes, including PHENYLALANINE AMMONIA-LYASE (PAL), CHALCONE SYNTHASE (CHS), CHALCONE ISOMERASE (CHI), FLAVONOID 3′-MONOOXYGENASE (F3′H), FLAVONOID 3′, 5′-HYDROXYLASE (F3′5H), FLAVANONE

Received March 28, 2023. Accepted May 15, 2023. Advance access publication June 22, 2023
© The Author(s) 2023. Published by Oxford University Press on behalf of American Society of Plant Biologists.
Open Access
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

3-HYDROXYLASE (F3H), DIHYDROFLACONOL 4-REDUCTASE (DFR), ANTHOCYANIDIN SYNTHASE (ANS), and UDP-GLUCOSE:FLAVONOID 3-*O*-GLUCOSYLTRANSFERASE (UFGT), and has extensively been characterized in plants. Moreover, its biosynthesis is regulated by various transcription factors (TFs), such as MYBs and MYB/BASIC HELIX-LOOP-HELIX (bHLH)/WD40 (MBW) protein complexes (Xu et al. 2015).

The cytosolic surface of the endoplasmic reticulum (ER) acts as the site of anthocyanin biosynthesis, which subsequently is transported and stored in the vacuole. Several anthocyanin transporters have been isolated, such as GLUTATHIONE S-TRANSFERASE (GST) (Zhao 2015). Plant GSTs is a gene superfamily encoding multifunctional enzymes that recruit the tripeptide glutathione as coenzymes or cosubstrate for cellular activity (Vaish et al. 2020). Plant GSTs are divided into 9 subfamilies, with the plant specific phi (F) subfamily members being predominantly involved in anthocyanin transport (Edwards and Dixon 2005). The first anthocyanin-associated GST was reported in maize (*Zea mays*) (Marrs et al. 1995). The maize bronze-2 (*bz2*) mutants accumulate anthocyanin in the cytoplasm, where it is rapidly oxidized and polymerized, leading to brown rather than red kernels. Functional characterization of this mutant gene revealed that *ZmBZ2* encodes a phi GST transporter. Subsequent studies have characterized anthocyanin GST transporters in other plants, such as *ANTHOCYANIN9* (*AN9*), *TRANSPARENT TESTA 19* (*TT19*), and *REDUCED ANTHOCYANINS IN PETIOLES* (*RAP*) in petunia (*Petunia hybrida*), *Arabidopsis* (*Arabidopsis thaliana*), and common strawberry (*Fragaria ananassa*), respectively (Alfenito et al. 1998; Kitamura et al. 2004; Luo et al. 2018).

Lotus (*Nelumbo* spp.) is an aquatic primitive eudicot plant that belongs to the family Nelumbonaceae (Shen-Miller 2007). Due to its high ornamental values, such as unique fragrance, bright petal color, and variable flower shape, lotus is considered as one of the most popular traditional flowers in China. The genus *Nelumbo* contains only 2 species, American lotus (*Nelumbo lutea* Willd.) and Asian lotus (*Nelumbo nucifera* Gaertn.) (Lin et al. 2019), and flower color is the most distinctive phenotypic variation between the 2 species, with the former only having yellow flowers, while the latter has white or red flowers. Numerous lotus cultivars with dual color petals have been developed by interspecific and intraspecific hybridization (Lin et al. 2019).

Similar to other ornamentals, anthocyanins are the major pigments determining the red lotus flower color (Yang et al. 2009). Anthocyanin content is positively associated with petal redness, and the red lotus cultivars predominantly accumulate anthocyanins compared with the pink varieties, while no anthocyanin is accumulated in the white and yellow cultivars (Deng et al. 2013). Deciphering the mechanisms of inter- and intraspecies pigment divergence in *Nelumbo* is currently a trending research topic. A recent integrated metabolite and transcriptome study between the yellow and white flower lotus cultivars identified 18 candidate genes potentially associated with the yellow color formation, such as *FLAVONOL SYNTHASE* (*FLS*) gene, which encodes a crucial flux enzyme in the flavonol biosynthesis pathway (Wu et al. 2022). The competition for substrate between FLSs and other biosynthetic enzymes contributes to the high quercetin derivatives in the yellow flowers of *N. lutea* (Liu, Wang, et al. 2022). Allelic variation in *NnMYB5* has previously been reported to likely contribute to flower color difference between *N. lutea* and *N. nucifera* (Sun et al. 2016). A comparative genomic analysis revealed structural variations (SVs) adjacent to *NnMYB5* between the 2 *Nelumbo* species, and transient expression of this gene induced anthocyanin accumulation in lotus petal (Zheng et al. 2022).

The genetic basis of color differentiation within the *N. nucifera* species has previously been explored. Comparative proteomic analysis of red and white *N. nucifera* cultivars identified 4 enzymes involved in anthocyanin biosynthesis and its trafficking pathway, including F3H, ANS, UFGT, and GST (Deng et al. 2015). Moreover, the study observed different methylation intensities on the promoter of *ANS* gene, which potentially resulted in different flower coloration between the red and white *N. nucifera* cultivars. The flower color of a single lotus cultivar is stable under different external environmental conditions and can stably be inherited by offspring, which suggest that flower color differentiation within *N. nucifera* is predominantly regulated by genetic factors. However, the genetic basis underlying the regulation of color diversity in *N. nucifera* is yet to be reported.

In this study, we conducted quantitative trait locus (QTL) mapping using an F_2 population derived from crosses between the red and white petal cultivars to identify the genetic loci controlling petal color divergence in *N. nucifera*. Utilizing the recently reported high-density genetic linkage maps (Liu, Zhang, et al. 2022), a QTL with high logarithm of odds (LOD) value on linkage group 3 (LG3) was detected. After bulk sequencing and gene expression analysis, a putative petal color regulator gene, designated as *NnMYB5*, was identified. Comparative genomic analysis revealed that an 80 kb presence/absence causal variant harboring the *NnMYB5* gene was responsible for color diversity between the red and white cultivars. Further analyses identified an anthocyanin transporter *NnGST2* as the downstream target gene of *NnMYB5*. Transgenic assays in *Arabidopsis* and lotus confirmed the role of *NnGST2* in anthocyanin transport and accumulation. Our finding sheds light on the molecular mechanism of flower color diversity in *N. nucifera* and provides molecular markers for flower color improvement during lotus breeding.

Results

NnMYB5 is a candidate regulator of petal color variation in *N. nucifera*

A QTL mapping of an F_2 population derived from a cross between the white flower cultivar 'BG' and red flower cultivar 'WR1' was conducted to identify putative genes controlling the flower color phenotype in *N. nucifera* (Fig. 1A). High

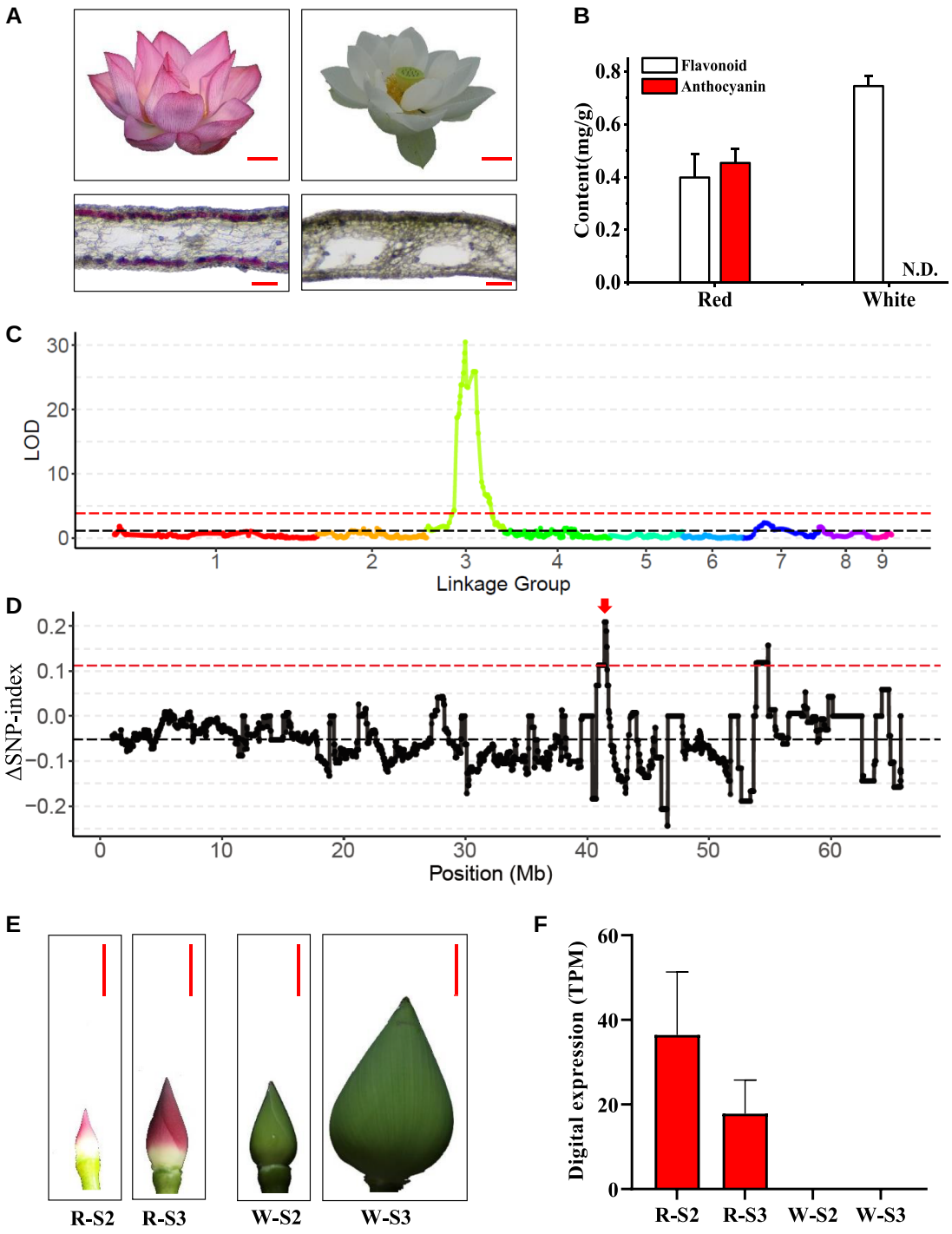

Figure 1. QTL mapping of petal color trait in *N. nucifera*. **A)** Flowers of the parental genotype in the mapping population. The upper images show the paternal flower 'WR1' (left) and maternal flower 'BG' (right). Scale bar = 3 cm. The lower images are petal sections showing pigment accumulation in the abaxial and adaxial epidermis of 'WR1' petals, but not in the 'BG' petals. Scale bar = 300 μm. **B)** The anthocyanin and other flavonoid contents in the red and white petals. N.D., not detected. **C)** Mapping of major QTL for petal color phenotype on LG3. The black and red dashed line represent median of LOD and significant limit of LOD threshold of 3.895 at $P = 0.05$ (1,000 permutation), respectively. **D)** ΔSNP index across mega-scaffold 5 between the white and red lotus accessions. The black dots show the mean ΔSNP in a 1 Mb sliding window with 20 kb increments. The black dashed line represents the mean ΔSNP across the region, while the red dashed lines represent 95% CI. Red arrow shows the target interval related to petal color. **E)** Petal developmental stages used for high-throughput RNA-seq. R-S2 and R-S3 (W-S2 and W-S3) represent petals from *N. nucifera* cv. 'QX' ('ZGWS') flowers at developmental stages 2 and 3, respectively. Scale bar = 1 cm. The images in panels **A)** and **E)** were digitally extracted for comparison. **F)** The digital expression profiling of candidate *NnMYB5* (*NNU_03972*) gene between the 'QX' and 'ZGWS' petals at 2 stages. TPM, transcripts per kilobase of exon model per million mapped reads. Data are mean ± SD ($n = 3$).

anthocyanin content was detected in red lotus petals but with no accumulation in white petals (Fig. 1B). Of the 125 F_2 individuals, 95 exhibited red flower phenotype, while the remaining (30 individuals) had white flowers, showing a segregation ratio of 3:1 (χ^2 test $P = 0.796$) for a single-gene controlled trait. Using the recently reported high-density genetic linkage map constructed with this F_2 population (Liu, Song, et al., 2022), a QTL on LG3 at the 22.87 to 60.57 cM region was found to be significantly associated with flower color (Fig. 1C). The phenotypic variation explained by the QTL was 85% (30.41 LOD). When anchored on the lotus reference genome, the QTL interval covered 27.75 Mb region on megascaffold 5 from 24.15 to 51.90 Mb (Fig. 1C). To narrow down the interval, 25 red and 23 white lotus accessions were grouped into 2 pools for bulked segregant analysis (BSA). After calculating allele frequencies for the 2 pools in all polymorphic sites across megascaffold 5, the QTL interval was narrowed to a 1 Mb region within the range of 40.75 to 41.75 Mb and a 2 Mb region within the range of 52.89 to 54.89 Mb (Fig. 1D).

Forty-one and 39 annotated genes were identified in the candidate regions, and their expression profiles were determined at the R-S2 and R-S3 or W-S2 and W-S3 developmental stages in the red or white flower petals, respectively, using the RNA-seq TPM data (Fig. 1E). Interestingly, all the 80 genes showed no differential expression patterns between the 2 different petal phenotypes. Consequently, the candidate interval was expanded by 250 kb at both ends, and the expression profiles of genes in the 4.0 Mb region (40.50 to 42.00 Mb and 52.64 to 55.14 Mb) were analyzed. As a result, 62 and 48 genes were detected in the target regions, of which, *NNU_03972* in the first interval, annotated as *NnMYB5*, showed high expression levels in red petal, but with no expression in the white petals (Fig. 1F and Supplemental Fig. S2). *NnMYB5* has been demonstrated to be involved in lotus petal color variation, whose natural variation coincided with petal color, and the function of *NnMYB5* in anthocyanin accumulation was validated by overexpression in Arabidopsis and transient expression in lotus (Sun et al. 2016; Gao et al. 2022). Based on its unique profile in the red petals and previous reports, *NnMYB5* was deemed as a key candidate gene controlling petal color diversity in *N. nucifera*.

An 80 kb presence/absence variant (PAV) harboring *NnMYB5* is associated with the petal color phenotype in *N. nucifera*

The candidate *NnMYB5* gene located at 41.91 Mb on megascaffold 5, corresponding to the 33.77 Mb region on Chr 4 in the reference genome of the red flower *N. nucifera*, occurred within a 158 kb region containing a 6-member *MYB* gene cluster, including *NnMYB2* (*NNU_03970*), *NnMYB3* (*NNU_03969*), *NnMYB4* (*NNU_03973*), *NnMYB5* (*NNU_03972*), *NnMYB6* (*NNU_03978*), and *NnMYB7* (*NNU_03981*) (Fig. 2A). Subsequently, the resequencing data of 213 natural lotus accessions was analyzed to determine sequence variation of *NnMYB5* between the red and white flower cultivars. The BLAST analysis against the resequenced lotus genomes revealed 3 *NnMYB5* sequence genotypes, designated Nl-G, Nn-G1, and Nn-G2. The Nl-G genotype was a pseudogene caused by single-nucleotide polymorphism (SNPs) and small insertions and deletions (InDels) specific in *N. lutea*, and 92.31% of accessions with this genotype had yellow flowers. The Nn-G1 genotype was due to the absence of *NnMYB5* caused by the 80 kb deletion, and this genotype occurred in *N. nucifera* with white (56.36%) or yellow (43.64%) flowers. The Nn-G2 genotype contained a functional *NnMYB5* gene without the 80 kb deletion and was present in the red, white, and yellow *N. nucifera* flowers in proportions of 71.43%, 18.49%, and 10.08%, respectively (Fig. 2B). These results strongly suggested that the absence of *NnMYB5* could result in the lack of anthocyanin accumulation in flower petals. Genotyping showed that the yellow flower lotus accessions were hybrids of *N. nucifera* and *N. lutea*. Further sequence variation analysis between the red and white *N. nucifera* revealed the presence of an 80 kb PAV on Chr 4 of the lotus genome, which harbored 3 MYB genes, including *NnMYB5*, *NnMYB6*, and partial *NnMYB7* (Fig. 2A). Gene expression analysis showed that *NnMYB5* was specifically expressed in the red lotus petals (Supplemental Fig. S3).

To determine the association between PAV and anthocyanin deficiency in lotus petals, a pair of primers covering this PAV was designed in its flanking regions for PCR amplification, followed by Sanger sequencing for PAV sequence confirmation. In addition, another primer pair was designed to amplify the *NnMYB5* genomic sequence in 14 red and 12 white lotus cultivars. PCR results showed that all 12 white flower accessions contained this 80 kb deletion of the PAV region, while the red cultivars contained an 80 kb homozygous or heterozygous PAV segment (Fig. 2C and Supplemental Table S2). Moreover, PCR screening of 50 segregating F_2 population offspring derived from 'BG' and 'WR1' revealed 3 genotypes, designated AA, Aa, and aa with a segregation ratio of 1:2:1 (Fig. 2D). The white individuals had aa genotype, while red individuals had Aa or AA genotypes (Supplemental Fig. S4). These results supported the hypothesis that the flower color variation in *N. nucifera* was affected by the 80 kb PAV harboring the *NnMYB5* gene.

NnMYB5 targets and interacts with the anthocyanin transporter *NnGST2* gene in *N. nucifera*

Comparative transcriptome analysis between the red and white petals was conducted to explore the regulatory mechanism of the 80 kb PAV in *N. nucifera* petal color variation. As a result, 283,294,832 clean reads were obtained from 12 petal libraries. The mapping ratio of each sample against the reference genome ranged from 90.46% to 92.82% (Supplemental Table S3). A total of 313 DEGs were identified between the R-S2 and W-S2 stages, including 137 up- and 176 downregulated genes. Similarly, 283 DEGs were identified between the R-S4 and W-S4, including 124 up- and 159 downregulated

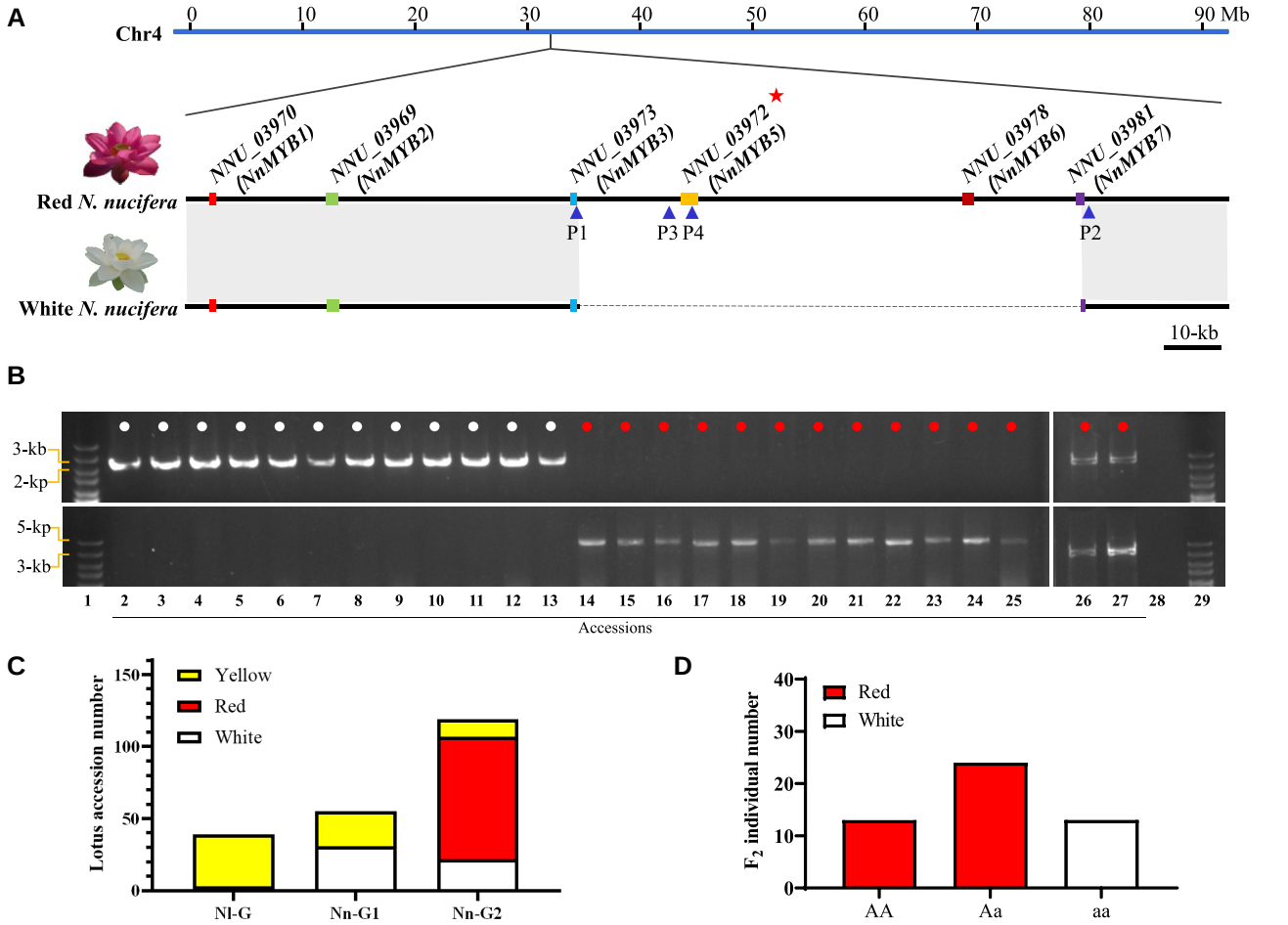

Figure 2. An 80 kb PAV on Chr 4 controlling petal color divergence in *N. nucifera*. **A)** Physical maps of the 80 kb PAV between lotus cultivars with white or red flowers on Chr 4 showing physical locations of members of the *MYB* gene cluster marked in different colored boxes. The dashed line shows the 80 kb deletion in lotus cultivars with white flower. *NnMYB5* is highlighted with red asterisk. Blue triangles at the bottom represent location where the 2 pairs of primers were developed. **B)** Genotyping of *N. nucifera* accessions to verify the association between PAV and petal color. Red dots indicate *N. nucifera* accessions with red flower, and white dots indicate accessions with white flower. DNA fragments on the upper and lower were amplified with the P1/P2 and P3/P4 primer pairs, respectively. **C)** Distribution of 213 lotus accessions with different petal colors based on the 3 genetic variations at the PAV locus. Nl-G represents genotype from *N. lutea*, while Nn-G1 or Nn-G2 represents *N. nucifera* genotypes harboring or lacking the 80 kb PAV, respectively. **D)** Genotype distribution of 50 F_2 individuals derived from a cross between 'BG' and 'WR1.'

genes. Among the DEGs between the red and white petals at the 2 stages, 170 genes were commonly shared, with 87 and 83 exhibiting up- and downregulated expression, respectively (Fig. 3A and Supplemental Table S4). Gene Ontology (GO) enrichment showed that the 170 commonly shared genes were enriched in biological processes, such as anthocyanin containing compound biosynthesis, as well as functional molecular categories related to secondary metabolite transport, such as drug transmembrane transporter activity (Supplemental Fig. S5).

Most anthocyanin biosynthesis genes showed no significant differential expression between the red and white petals (Fig. 3, B and C). The *CHS* gene (*NNU_16744*) that catalyzes the first committed step in the flavonoid biosynthetic pathway displayed significantly higher expression levels in the red lotus petals. Conversely, the *FLS* gene (*NNU_05540*), which controls the flux of flavonol rather than the anthocyanin biosynthetic pathway, showed significantly lower expression levels in the red lotus petals. One *UFGT* gene (*NNU_01692*) was considerably upregulated in the red lotus petals relative to white petals. Three anthocyanin transporter genes, including 2 *GSTs*, *NNU_15680* and *NNU_03253* as well as a MATE transporter *TT12* gene (*NNU_09866*), were differentially expressed between the red and white lotus petals. The expression level of an anthocyanin transporter gene *NNU_03253*, named *NnGST2*, corresponded with that of *NnMYB5*, with both being highly expressed in the red petals, but absent in the white petals. Previous experiments showed that the expression of *NnGST2* was upregulated by overexpression of *NnMYB5* in Arabidopsis (Sun et al. 2016). In addition, *NnGST2* was likely to be involved in petal color variation based on allele-specific expression in F1 hybrids (Gao et al. 2022). Based on these results, *NnGST2* was designated as a candidate downstream target gene of *NnMYB5* (Fig. 3D).

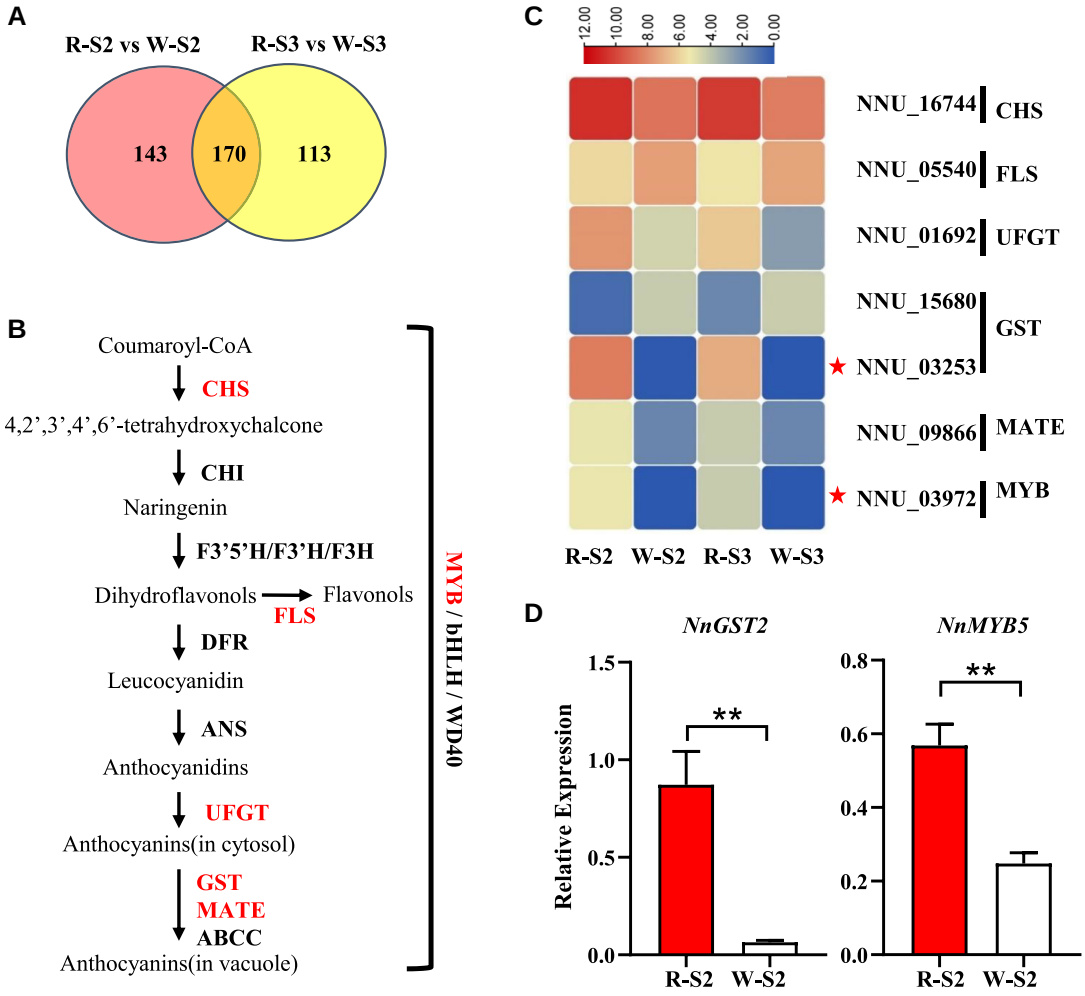

Figure 3. Transcriptome profiles of genes putatively involved in the anthocyanin biosynthetic pathway. **A)** Venn diagram showing the number of overlapping DEGs between the 'QX' and 'ZGWS' lotus petals in 2 developmental stages. R-S2 and R-S3 (W-S2 and W-S3) represent petals from *N. nucifera* cv. 'QX' ('ZGWS') flowers at developmental stages 2 and 3, respectively. **B)** Schematic diagram of anthocyanin biosynthesis pathway. Genes highlighted in red were differentially expressed between the 2 developmental stages. **C)** Expression patterns of anthocyanin-related DEGs. The red asterisks indicate candidate genes analyzed in this study. The color scale indicates the $\log_2$FC values of DEGs. **D)** RT-qPCR expression patterns of candidate genes in petal at the S2 developmental stage. Data are mean ± SD ($n = 3$). Double asterisks indicate statistical significance at $P < 0.01$ (*t* test).

The candidate *NnGST2* gene is an anthocyanin-related GST putative ortholog in *N. nucifera*

GSTs are ubiquitous multifunctional plant enzymes that are encoded by a large gene family. In this study, 67 GSTs were identified in the *N. nucifera* genome, which were phylogenetically grouped in 9 classes, including phi (F), tau (U), lambda (L), zeta (Z), dehydroascorbate reductase (DHAR), glutathionyl hydroquinone reductase (GHR), theta (T), tetrachlorohydroquinone dehalogenase (TCHQD), and elongation factor 1Bγ (EF1Bγ) (Supplemental Figure S6A). Of these 9 GST categories, members of the phi class have previously been associated with the transport and metabolism of secondary compounds, such as anthocyanins, and 10 phi class *NnGSTs* were detected in this study (Supplemental Fig. S6A).

Expression profiling of these 10 *NnGSTs* in the petiole, leaf, petal, pistil, stamen, and rhizome tissues of the red lotus using a previously reported RNA-seq FPKM data revealed that *NnGST1*, *NnGST4*, *NnGST7*, and *NnGST8* were predominantly expressed in root, pistil, stamen, and petiole tissues, respectively (Supplemental Figure S6B). *NnGST3* was mainly expressed in pistil, petiole, and leaf tissues, while *NnGST6*, *NnGST9*, and *NnGST10* were highly expressed in the lotus petiole, petal, and root organs. *NnGST5* showed preferential expression in the lotus petal, stamen, and root tissues. Notably, *NnGST2* was highly expressed specifically in the lotus petals, suggesting its potential regulatory function in the lotus petal color formation.

RT-qPCR analysis to validate the tissue expression profiles of *NnGST2* and determine its biological function in petal color formation revealed that the candidate gene was highly expressed in the lotus petals at early developmental stages before decreasing in later stages, which was consistent with

anthocyanin accumulation patterns (Fig. 4A). Subcellular localization of NnGST2 showed that green GFP signal of 35S::*NnGST2*-GFP fusion protein overlapped with the red mCherry signal, which indicated ER localization of NnGST2 (Fig. 4B).

NnMYB5 is a transcriptional activator of *NnGST2*

Transcriptional regulators, such as MYBs, bHLH, WRKY, and WD40 proteins, play pivotal roles in the biosynthesis of anthocyanins. Pearson correlation analysis revealed a highest coefficient of $r = 0.95$ between the expression levels of *NnGST2* and *NnMYB5* (Supplemental Fig. S7). Moreover, similar trends in the expression patterns of *NnMYB5* and *NnGST2* were observed (Fig. 4, A and C), with the former showing predominant expression in lotus petals at the S2 stage, which coincided with rapid anthocyanin accumulation.

Yeast 1-hybrid (Y1H) assay to verify the regulatory role of NnMYB5 revealed that it could bind to the *NnGST2* promoter (Fig. 4D). Dual-luciferase assays in *Nicotiana benthamiana* leaf showed that NnMYB5 could significantly induce the expression of firefly luciferase driven by the *NnGST2* promoter, which demonstrated its capacity to induce the expression of *NnGST2* (Fig. 4E). Moreover, transient expression of *NnMYB5* in lotus petals could upregulate the transcript levels of *NnGST2*, which was consistent with the transactivation capacity of NnMYB5 on the *NnGST2* promoter (Figs. 4F and Supplemental Fig. S8). Further Y1H screening assay to determine the potential role of NnMYB5 in the regulation of other anthocyanin biosynthesis genes revealed the *NnFLS* gene as its likely target, and LUC assay indicated that NnMYB5 could inhibit *NnFLS* transcription (Supplemental Fig. S9).

Ectopic overexpression of *NnGST2* promotes anthocyanin accumulation in the *Arabidopsis tt19* mutants

Complementation assay to validate the functional role of NnGST2 in anthocyanin transport was carried out in *Arabidopsis tt19* harboring a mutant anthocyanin transporter *GST-TT19*, and 3 distinct 35S::*NnGST2* transgenic lines with similar phenotypes were obtained. Notably, a green hypocotyl of stem in the *tt19* mutant was observed in the 5-d-old *Arabidopsis* seedlings, while the 35S::*NnGST2*/*tt19* transgenic line could not only rescue the red pigment phenotype but also produce normal seeds color in *tt19* (Fig. 5A). HPLC results showed significantly higher anthocyanin content in the 35S::*NnGST2*/*tt19* line than in the wild-type (WT) or *tt19* (Fig. 5, B and C). Moreover, expression profiling of structural anthocyanin biosynthetic genes in the WT, *tt19* mutant, and transgenic lines indicated that the early biosynthesis genes, such as *AtCHS*, *AtCHI*, and *AtF3H*, had similar expression profiles, while late biosynthesis genes leading to other metabolic branches, such as *AtF3H* and *AtFLS*, had decreased expression profiles in the 35S::*NnGST2*/*tt19* lines (Fig. 5D). In contrast, the expression of key anthocyanin biosynthesis genes, such as *AtDFR*, *AtANS*, and *AtUF3GT*, was upregulated in the transgenic lines.

Transient expression of *NnGST2* alters lotus petal coloration

Transient expression of *NnGST2* in the white lotus petals revealed that petals turned red after infiltration with 35S::*NnGST2*, while no pigment accumulation was observed in the control petals (Fig. 6A). Similarly, anthocyanin content and *NnGST2* transcript abundance in the infiltrated sites were significantly higher than those of control petals (Fig. 6, C and D). Moreover, the transcript levels of *NnCHS* were upregulated in the 35S::*NnGST2* infiltrated petals, whereas the expression levels of *NnFLS*, *NnANS*, and *NnF3H* displayed no significant differences (Fig. 6D). These results demonstrated the critical role of *NnGST2* in anthocyanin accumulation and lotus petal coloration, potentially by altering both anthocyanin transport and biosynthesis processes.

Discussion

A PAV of *NnMYB5* is associated with petal color in *N. nucifera*

Anthocyanins are pigments responsible for the red flower color in *N. nucifera*. The anthocyanin biosynthesis is genetically controlled by structural and regulatory genes, with MYB TFs as the core transcriptional regulators. Natural variations in *MYB* genes have been previously demonstrated to alter the color phenotypes of organs in numerous plants. For example, a 487 bp deletion in the promoter of *PpMYB10.1* was shown to be associated with the flesh color around the peach stone (Guo et al. 2020). An insertional long terminal repeat (LTR) retrotransposon upstream of the *MdMYB1* gene could be attributed to the red apple skin color (Zhang et al. 2019). Natural SNP variations in the MYB recognition site at the *DFR* promoter were reported to cause differential *DFR* gene expression and distinct anthocyanin accumulation in fruits of *Solanum* species (Wang, Lu, et al. 2022). In this study, a petal color QTL harboring an *MYB* gene cluster was identified, among which only *NnMYB5* gene was differentially expressed between the white and red lotus petals, suggesting its potential role in the *N. nucifera* petal color variation (Fig. 1).

SVs are DNA sequence rearrangements across or within genomes that range from SNPs to large-scale PAVs and inversions or translocations. Large SVs are not easily detectable, and they have greater influence on the expression and function gene compared with SNPs or small InDels (Chiang et al. 2017). Here, comparative genomic analysis identified an 80 kb PAV containing a *NnMYB5* gene, which was highly associated with *N. nucifera* petal color variation (Fig. 2). This result is consistent with several recent studies demonstrating the effects of large SVs on the phenotypic variations within and between plant species. For example, a 1.7 Mb inversion was reported to cause the flat peach fruit trait by activating

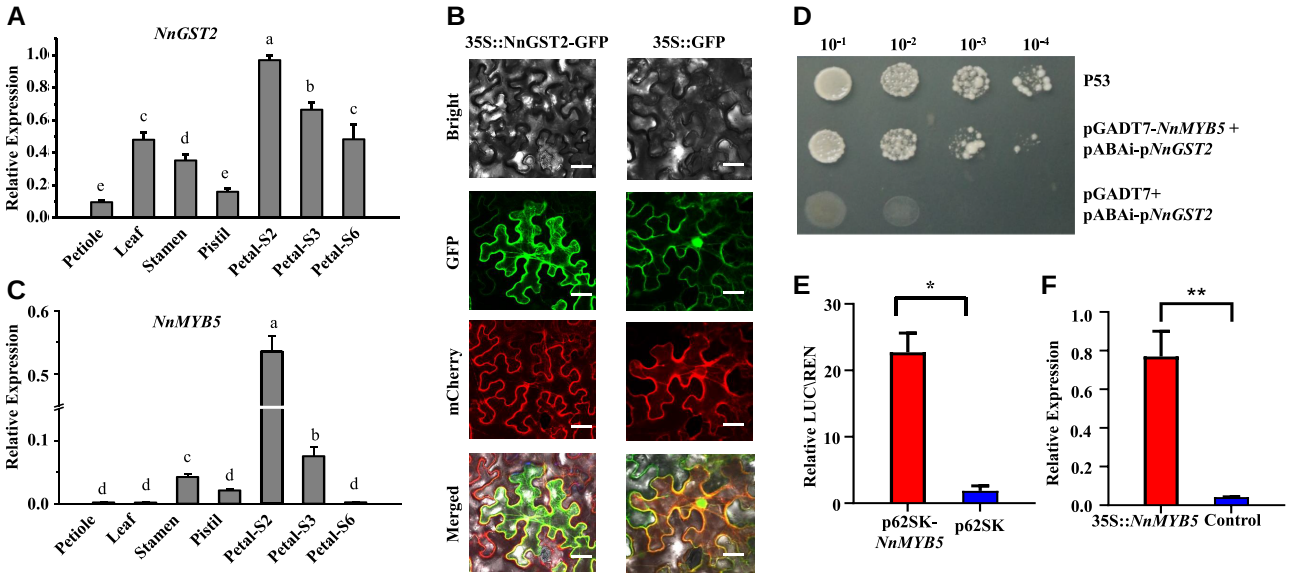

Figure 4. Functional analysis of *NnGST2*. **A)** RT-qPCR expression profiles of *NnGST2* in different lotus flower tissues. Different alphabetical letters indicate statistical significance at $P < 0.05$ (1-way ANOVA with Tukey's test). Data are mean ± SD ($n = 3$). S2 and S3 were *N. nucifera* cv. 'QX' petals at stages 2 and 3 as shown in Fig. 1E, while S6 was petal from first day opening flower. **B)** Subcellular localization of NnGST2 in *N. benthamiana* leaf. Leaves expressing 35S::*GFP* were used as the control. The green fluorescent protein (GFP) and ER located mCherry proteins are stained with green and red hues, respectively. **C)** RT-qPCR expression profiles of *NnMYB5* in different lotus flower tissues. Different alphabetical letters indicate statistical significance at $P < 0.05$ (1-way ANOVA with Tukey's test). Data are mean ± SD ($n = 3$). **D)** The Y1H assay showing the interaction between NnMYB5 and the *NnGST2* promoter. From left to right, the yeast concentration at $OD_{600} = 0.1$, 0.01, 0.001, and 0.0001, respectively. P53 was used as positive control and pGADT7 + pABAi-p*NnGST2* as negative control. **E)** Dual-luciferase assay showing NnMYB5 activation capacity on the *NnGST2* promoter. LUC\REN: the ratio of firefly luciferase activities to renilla luciferase activities. p62SK was negative control. Single asterisk (*) indicates statistical significance at $P < 0.05$ (*t* test). Data are mean ± SD ($n = 3$). **F)** Transcript levels of *NnGST2* in transgenic petals expressing 35S::*NnMYB5*. Double asterisk (**) indicates statistical significance at $P < 0.01$ (*t* test). Data are mean ± SD ($n = 3$).

the *OVATE FAMILY PROTEIN1* (*PpOFP1*) gene located around its breakpoints (Zhou et al. 2021). Similarly, a large inversion SV at the *I* locus associated with the regulation of seed coat color by silencing the expression of anthocyanin biosynthetic *CHS* gene during soybean domestication has recently been reported (Xie et al. 2019). Moreover, a pan-genome analysis of 29 soybean accessions verified the association between the inversion SV and yellow seed coat color, and several additional SV events accompanying the inversion and *CHS* gene duplication in different haplotypes were detected (Liu et al. 2020). This study reports the occurrence and effects of a large PAV on the regulation of lotus petal color.

Two independent allelic variants of *MYB5* results in intra- and interspecies petal color diversity in *Nelumbo*

The lack of anthocyanin accumulation in the white petals of *N. nucifera* and yellow petals of *N. lutea* is potentially associated with distinct genetic mechanisms between the 2 species. The allelic variants of *MYB5* have been implicated in the color differentiation between the 2 *Nelumbo* species, and multiple SNPs and InDels have been detected in the coding and promoter regions of *MYB5* gene in *N. nucifera* and *N. lutea* (Sun et al. 2016; Zheng et al. 2022). A premature translational termination of *MYB5* due to SNPs in its exons caused nonfunctional alleles with no capacity to regulate anthocyanin biosynthesis pathway in *N. lutea* (Sun et al. 2016). A recent transient expression assay demonstrated the role of *NnMYB5* in the regulation of anthocyanin accumulation in lotus (Zheng et al. 2022). In this study, pigment deficiency in the white *N. nucifera* cultivars was shown to be caused by the deletion of a large genomic fragment harboring the crucial *NnMYB5*. The petal color difference between the white *N. nucifera* and yellow *N. lutea* was attributed to a deletion of *NnMYB5* and SNP/InDel mutations in *NlMYB5* that affected the downstream regulatory network. The 80 kb deletion was detected in some heterozygous hybrid red flower cultivars. Hybrids of white *N. nucifera* and *N. lutea* exhibited yellow flowers, which supported the conclusion that *MYB5* mutations in *N. lutea* and deletion in the white *N. nucifera* cultivars were both functional and not complementary. Consistently, 2 independent mutations in the strawberry *FaMYB10-2* gene were also shown to cause natural variation in fruit skin and flesh color (Castillejo et al. 2020). These results provide vital information for future lotus genetic improvement.

NnMYB5-mediated activation of *NnGST2* regulates anthocyanin transport in *N. nucifera*

GSTs are crucial regulators of anthocyanin translocation from the ER into vacuoles (Zhao 2015). The subcellular

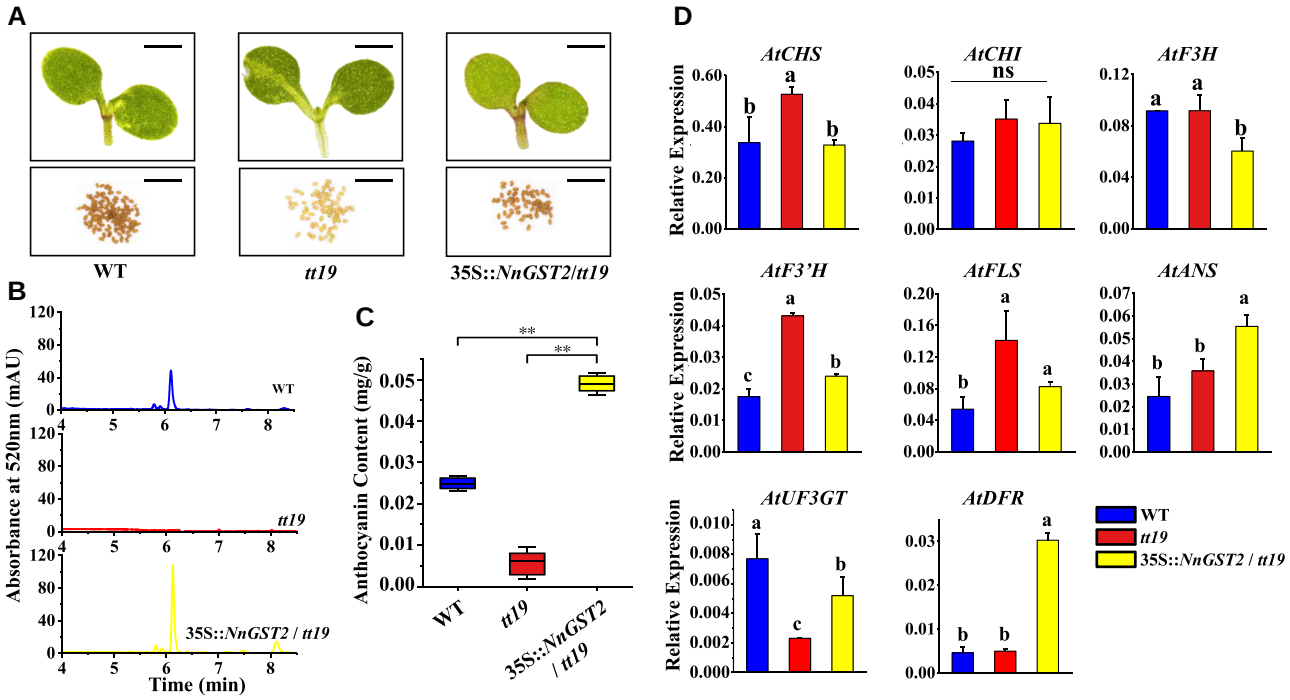

Figure 5. NnGST2 rescues anthocyanin deficiency in *Arabidopsis tt19* mutants. **A)** Phenotypes of 7-d-old WT, *tt19*, and 35S::*NnGST2*/*tt19* *Arabidopsis* seedlings and their corresponding mature seeds. Images were digitally extracted for comparison. Scale bar = 500 μm. **B)** HPLC chromatograms of anthocyanin profile in the WT, *tt19*, and 35S::*NnGST2*/*tt19* seedlings. **C)** Anthocyanin contents in the WT, *tt19*, and 35S::*NnGST2*/*tt19* seedlings calculated based on the peak area method. Double asterisks (**) indicate statistical significance at $P < 0.01$ (*t* test). Data are mean ± SD ($n = 3$). The boxplot elements are defined as follows: centerline, median; box limits, upper and lower quartiles; and whiskers, 1.5× interquartile range. **D)** The expression of anthocyanin biosynthesis genes in the WT, *tt19*, and 35S::*NnGST2*/*tt19* transgenic lines. Different alphabetical letters show significant differences at $P < 0.05$ (1-way ANOVA with Tukey's test). Data are mean ± SD ($n = 3$).

localization assay of NnGST2 showed GFP signal in the ER (Fig. 4). Transgenic expression in the *Arabidopsis tt19* mutants and lotus petals for functional validation of *NnGST2* further demonstrated its involvement in anthocyanin transport and accumulation, with predominant expression in the red cultivars indicating its role in *N. nucifera* petal color divergence (Figs. 5 and 6). A recent study on the allele-specific expression analysis in F_1 *N. nucifera* and *N. lutea* hybrids also predicted the *GST2* gene, designated as *GSTF11*, to be the key hub gene controlling color divergence between the 2 *Nelumbo* species (Gao et al. 2022). Overall, these findings indicated that anthocyanin translocation play vital roles in lotus color formation.

A previous gene coexpression analysis showed that both *cis-* and *trans*-acting regulatory capacity of *GSTF11* could contribute to color divergence in *N. nucifera* and *N. lutea* (Gao et al. 2022). In contrast to interspecies sequence variations, no intraspecies SNPs or InDels were detected on the genomic region of *GST2* in *N. nucifera*, suggesting the *trans* regulatory effects, especially due to TFs that regulate the intraspecies expression levels of *NnGST2*. MYBs have extensively been reported to regulate anthocyanin accumulation through transcriptional activation or suppression of pathway structural genes (Xu et al. 2015). However, increasing evidence shows that the MYB TF family is also involved in the anthocyanin transport. For example, a key strawberry fruit pigmentation *RAP* (*reduced anthocyanin in petioles*) gene that encodes a GST transporter was shown to act downstream of the fruit-specific *FvMYB10* gene (Luo et al. 2018). In apple, an MdMYB1 protein could directly bind to the promoter of *MdGSTF6* and activate its expression (Jiang et al. 2019). In addition, an anthocyanin biosynthetic MYB activator, LEGUME ANTHOCYANIN PRODUCTION 1 (LAP1), could bind to the *MtGSTF7* promoter and activate its expression in *Medicago* (Wang, Chen, et al. 2022). In this study, dual-luciferase and Y1H assays revealed that *NnGST2* was directly targeted and interacted with NnMYB5, which is a key of anthocyanin biosynthetic pathway regulator in lotus petal (Fig. 4). These findings revealed that the NnMYB5-activated *NnGST2* gene could regulate anthocyanin transport in the red *N. nucifera* petals, which comprehensively demonstrated the molecular regulatory mechanism of anthocyanin accumulation in lotus.

In summary, this study reports a PAV harboring *NnMYB5* that controls petal color variation between the white and red *N. nucifera*. Missing of *NnMYB5* that failed to activate *NnGST2* is responsible for the observed anthocyanin deficiency in the white lotus petals. Ectopic overexpression of *NnGST2* in *Arabidopsis* and transient expression in lotus validated its role in anthocyanin translocation. This study

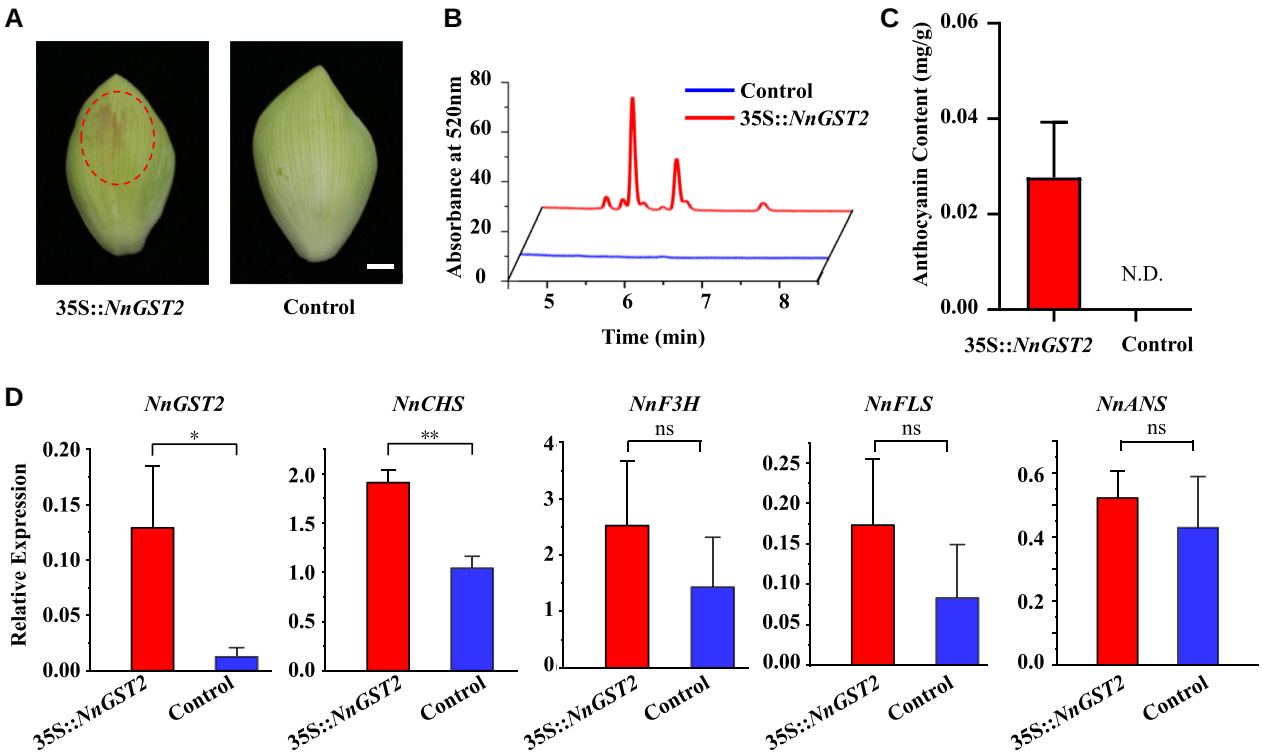

Figure 6. Transient expression of *NnGST2* induces anthocyanin accumulation in the white lotus petals. **A)** The phenotype of lotus petal transiently overexpressing *NnGST2*. Scale bar = 1 cm. **B)** HPLC chromatogram of anthocyanin profile in the control and petals transiently overexpressing 35S::*NnGST2*. **C)** Anthocyanin content in the control and transgenic petals overexpressing 35S::*NnGST2*. N.D., not detected. **D)** Transcript levels of anthocyanin biosynthesis genes in the control and petals overexpressing 35S::*NnGST2*. Data are mean ± SD ($n = 3$). Asterisk (*) indicates statistical significance at $P < 0.05$, and double asterisk (**) indicates statistical significance at $P < 0.01$ (t test) in **C)** and **D)**. ns, no significant difference.

uncovered the molecular mechanism underlying petal color differentiation within the *N. nucifera* species and demonstrated the role of 2 independent *NnMYB5* mutations in the control of petal color variation in the genus *Nelumbo* (Fig. 7).

Materials and methods

Plant materials

N. nucifera cultivars (cv.) 'Qiuxing' ('QX') with red flower and 'Zhigao Wushang' ('ZGWS') with white flower were cultivated at the Wuhan Botanical Garden of the Chinese Academy of Sciences (Wuhan, China). Fresh petals were used to prepare tissue sections for observation with a Phenix PH100 microscope equipped with a Phenix MC-D500U camera. Flower development was divided into 4 stages (S1 to S4) based on color changes in the red flower bud prior to blooming (Supplemental Fig. S1). The S2 flower developmental stage was characterized with rapid petal anthocyanin accumulation, which peaked at S3 in the red lotus, while no anthocyanin accumulation was detected throughout flower development in white lotus. Petal samples from flower buds at S2 and S3 stages were collected in 3 independent biological replicates and then immediately frozen in liquid nitrogen and stored at −80 °C for flavonoid extraction, genomic DNA, and total mRNA isolation. Arabidopsis (*A. thaliana*) and *N. benthamiana* plants were cultivated in an incubator at 23 °C with 16 h light and 8 h dark cycle.

Linkage mapping and BSA

A total of 125 individuals in an F_2 population derived from a cross between *N. nucifera* cultivars (cv.) 'Baige' ('BG,' female parent, white flower) and 'Winter Red 1' ('WR1,' male parent, red flower) were used to identify the QTL for lotus petal color by composite interval method as previously described using the WinQTL Cartographer 2.5 software (Liu et al. 2022). The significant limit of detection (LOD) threshold was determined with 1,000 permutations.

Two DNA pools for BSA were constructed using a mixture of 25 red and 23 white individuals from the F_2 population. High-quality SNPs and InDels were used here as previously reported (Liu et al. 2022). Difference in allele frequencies between the 2 bulks was computed using a 1 Mb sliding window with 20 kb increments in a customized Perl script.

Evaluation of *NnMYB5* gene SVs among lotus accessions

To explore the genetic variations in *NnMYB5* gene between the red and white lotus flowers, small SVs (SNPs and InDels) in the genic and upstream sequence regions of *NnMYB5* in 240 lotus accessions were analyzed using the previously reported genome resequencing data retrieved from

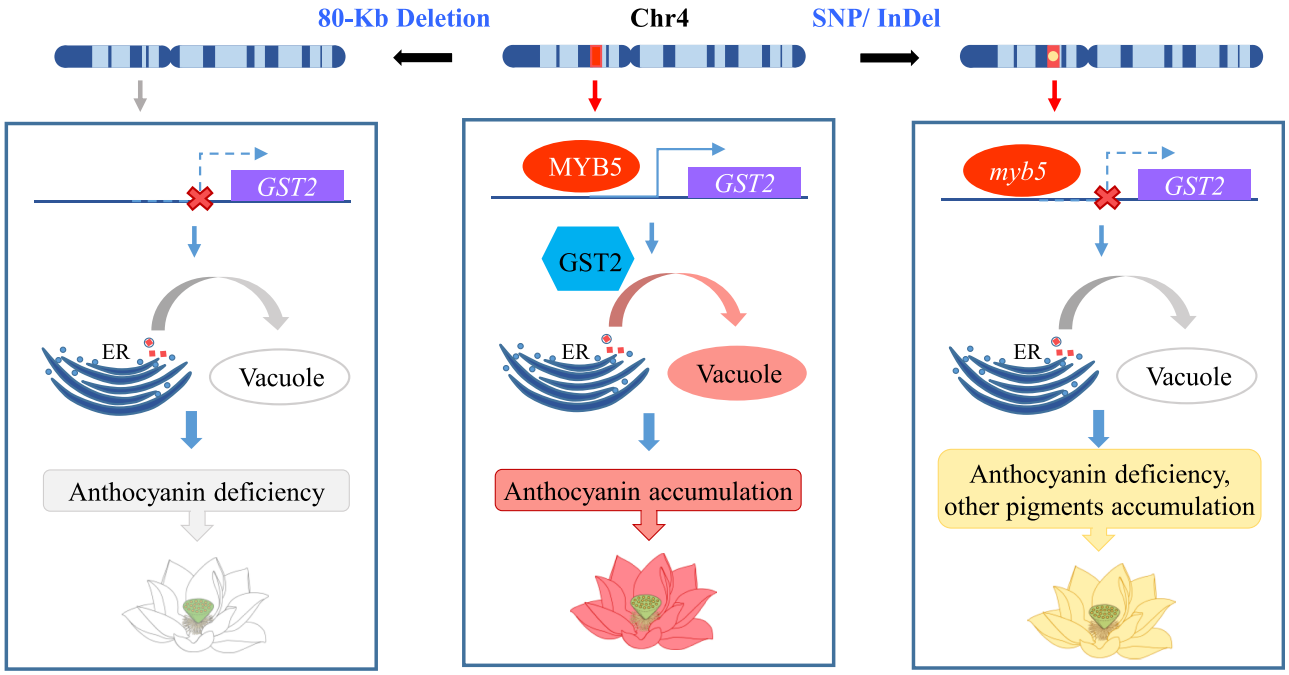

Figure 7. Hypothetical model illustrating the genetic mechanism underlying petal color variation in *Nelumbo*. In *N. nucifera* cultivars with red flower, MYB5 activates the transcription of *GST2*, which transports anthocyanin from the ER to the vacuole. In *N. nucifera* cultivars with white flower, deletion of a large 80 kb PAV region harboring the functional *MYB5* gene causes failure of *GST2* activation, consequently resulting in pigment deficiency. In *N. lutea*, SNPs in the *MYB5* exon cause a truncated protein, thereby blocking anthocyanin accumulation. The red boxes in Chr 4 represent the presence of *MYB5*, while the yellow circle shows SNPs in *MYB5*. The gray arrow under Chr 4 shows the absence of MYB5. The solid blue arrow represents success transcription of GST2, and the dashed ones together with the red crosses represent failing to activate the transcription of GST2. The gray curved arrows showed no anthocyanin transport from the ER to the vacuole.

NCBI under the BioProject number PRJNA749672, and highly-quality SNPs were obtained (Zheng et al. 2022). For analysis of large genetic variations, the genomes of 213 lotus accessions having pronounced flower color trait were assembled into contigs individually. Then, the *NnMYB5* gene sequences including their 2 kb promoters were searched in the lotus genome with the BLAST criteria of *e*-value $\leq 1e^{-6}$ and sequence identity $\geq 99\%$. Large SVs were manually detected based on sequence alignments in the BLAST results.

PCR validation of the 80 kb PAV between the white and red flower lotus cultivars

The genomic DNA was isolated from fresh leaves of 14 red and 12 white *N. nucifera* cultivar samples using DNAsecure Plant Kit (TIANGEN Biotech Co., Ltd., Beijing, China) following the manufacturer's instructions. PCR was performed with PrimeSTAR Max DNA Polymerase (Takara Bio., Shiga, Japan) to detect PAVs of *NnMYB5*. A pair of primers, P1/P2 (Supplemental Table S1), located at each end of the PAV, was designed to assay the white flower lotus accessions lacking the *NnMYB5* gene, while a pair of primers, P3/P4 (Supplemental Table S1), located in the promoter and second exon regions of *NnMYB5*, was designed to assay the red flower lotus accessions harboring the *NnMYB5* gene. As a result, 2 white flower lotus accessions with positive P1/P2 primer amplifications were detected and subsequently sequenced to verify the 80 kb deletion.

HPLC quantitation of anthocyanins and total flavonoids

Total flavonoids were extracted according to previously reported methods with minor modification (Chen et al. 2013). Briefly, ~1 g fresh petal sample was ground in liquid nitrogen and then extracted twice with 10 mL of extracting solution (methanol/water/formic acid = 70:28:2, *v*:*v*:*v*). After centrifugation at 10,000 rpm for 10 min, the supernatant was collected and filtered through a 0.22 μm membrane prior to HPLC analysis.

Chromatographic separation was performed on a liquid chromatography (Accela 1250) system with a SunFire C18 column (150 mm × 4.6 mm, 3.5 μm, Waters, MA, United States) following the protocol described by Yang et al. (2009). Anthocyanin and other flavonoids were detected at wavelengths of 520 and 350 nm, respectively. Petunidin-3-*O*-glucoside and isorhamnetin-3-*O*-rutinoside purchased from Yuanye Bio-Technology Co., Ltd. (Shanghai, China) were used to generate standard curves for quantification of anthocyanin and other flavonoids.

RNA extraction, library preparation, and sequencing

Petals of white and red flower lotus cultivars were harvested in 3 biological replicates at S2 and S3 developmental stages for transcriptome analysis. Total RNA was extracted from each sample using the fast plant RNA kit for polysaccharide and polyphenolic rich (Zoman Biotechnoligy, Beijing, China).

The RNA quality was determined by agarose gel, while RNA concentration was detected using a NanoDrop 2000 spectrophotometer (Thermo Scientific, MA, United States). cDNA libraries were constructed and sequenced by Nextomics Biosciences Co., Ltd. (Wuhan, China).

Gene expression analysis

Clean reads obtained after discarding low-quality reads were mapped to the lotus reference genome by HISAT2 (Ming et al. 2013; Pertea et al. 2016). Gene expression levels were quantified by transcripts per million reads (TPM) using StringTie. Differentially expressed genes (DEGs) were identified using the DESeq2 R package based on $|\log2FC| \geq 1$ and false discovery rate (FDR) < 0.05 filter criteria. GO and Kyoto Encyclopedia of Genes and Genome (KEGG) enrichment analysis were performed with the "topGO" R package and KEGG database, respectively.

RT-qPCR analysis was used to validate the expression levels of selected genes from the transcriptome data. The first-strand cDNA was synthesized using the TransScript II One-Step gDNA Removal and cDNA Synthesis SuperMix Kit (TransGen, Beijing, China). The gene-specific primers used for RT-qPCR were designed using Premier 6 (Supplemental Table S1), and the RT-qPCR reactions in 3 biological and 3 technical replicates were performed on a StepOnePlus Real-Time PCR System (Applied Biosystems, MA, United States) with *NnACTIN* as a reference gene. The reaction condition was as follows: 1 cycle of 30 s at 94 °C, followed by 40 cycles of 5 s at 94 °C, 15 s at 58 °C, and 10 s at 72 °C. The relative expression levels were calculated using the $2^{-\Delta Ct}$ method and normalized using *NnACTIN* (gene ID *NNU_24864*).

Phylogenetic analysis

The BLAST module in the TBtools software was used to retrieve *GST* genes in the *N. nucifera* genome (Chen et al. 2020; Shi et al. 2020). The *A. thaliana* and maize (*Z. mays*) GST proteins were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) and used as query sequences. The conserved domains of candidate genes were analyzed in the online conserved domain database (CDD) and protein classification server (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) to validate a total of 67 identified *NnGSTs*. The amino acid sequences of 62 AtGSTs, 20 ZmGSTS, and 67 NnGSTs were aligned with ClustalW, and the phylogenetic tree was constructed using MEGA7 software with the Neighbor-Joining (NJ) method (Thompson et al. 2002; Kumar et al. 2016).

Subcellular localization assay

The *NnGST2* coding region was cloned into pMDC43 vector to construct a 35S::*NnGST2*-GFP recombinant vector. The pMDC43 vector was used as a negative control and P53 as a positive control. An ER marker fused with a red fluorescent protein mCherry was used to determine the cellular ER location. The vectors were then transformed into *Agrobacterium tumefaciens* GV3101 strain and then infiltrated in *N. benthamiana* leaves. GFP was excited by the 488 nm laser line and was detected at 495 to 545 nm and mCherry 552 nm laser line and was detected at 600 to 650 nm using a confocal laser scanning microscopy (Leica TCS SP8, Wetzlar, Germany).

Overexpression of *NnGST2* in *Arabidopsis tt19* mutant

The *A. tumefaciens* GV3101 strain containing 35S::*NnGST2*-GFP was introduced into *A. thaliana tt19* mutant line. T1 seeds were screened on 1/2 MS media containing 50 mg/L kanamycin. The *NnGST2* gene primer was used to PCR amplify the genomic DNA (gDNA) isolated from seedlings to validate positive transgenic lines. The resulting positive T2 generation lines were collected and used for phenotypic evaluation, anthocyanin quantification, and gene expression analysis.

Transient expression of *NnGST2* in lotus petals

A. tumefaciens GV3101 strain harboring 35S::*NnGST2*-GFP was incubated and resuspended to a final optical density of OD_{600} of 0.8 in infiltration buffer (10 mM MES, 100 μM AS, and 10 mM $MgCl_2$) and then kept at room temperature for 2 h before injection in the petals of flower buds at developmental stages (S3 to S4). Empty pMDC43 vector was used as a control experiment. The injection experiment was conducted in the evening (5 PM to 7 PM) under good weather conditions and petals collected after 3 d for anthocyanin quantification and gene expression analysis.

Y1H assay

The 2 kb promoter regions of *NnGST2*, *NnFLS*, *NnF3H*, *NnANS*, and *NnCHS* were cloned into the pAbAi vector to generate bait constructs (Clontech, United States), while *NnMYB5* gene was cloned into the pGADT7 vector (Clontech, United States). The bait vector was digested with *BstB*I and transformed into yeast strain Y1HGold to generate the bait reporter construct. Then, the recombinant pGADT7-NnMYB5 was transformed into the positive bait reporter strain and grown on SD/-Leu/AbA (200 ng/mL) plates for 3 d. The empty pGADT7 and p53-AbAi + pGADT7-53 vectors were used as negative and positive controls, respectively.

Dual-luciferase reporter assay

The coding region of *NnMYB5* was cloned into pGreenII 62-SK, while the promoter sequence of anthocyanin-related structural genes, including *NnGST2*, *NnFLS*, *NnF3H*, *NnANS*, and *NnCHS*, was inserted into pGreenII 0800-LUC. The empty p62SK vector was used as a negative control. Both constructs were individually transformed into *A. tumefaciens* GV3101 using the electro-transformation method. The *A. tumefaciens* solution with plasmid was injected into *N. benthamiana* leaves with a needleless syringe. The Firefly luciferase (LUC) and Renilla luciferase (REN) activities were analyzed 3 d after infiltration using the Dual-Luciferase Reporter Assay System

(E1910 and E1960, Promega, Madison, WI, United States). Three biological replicates were performed for each treatment.

Accession numbers

The raw RNA-seq sequencing reads are submitted to NCBI Sequence Read Archive (SRA) database under the accession number PRJNA758182. The accession numbers of NnMYB5 and NnGST2 are ALU11263.1 and XP_010270745.1

Acknowledgments

We thank Dr. Kang Chunying from the College of Horticulture and Forestry Sciences, Huazhong Agricultural University (Wuhan, China), for kindly donating the *Arabidopsis tt19* mutant seeds.

Author contributions

M.Y. and D.Y. conceptualized the study and designed the experiments. D.Y., X.D., M.Z., Hen.S., L.G., Hey.S., and J.X. performed the experiments. M.Y., J.L., and Y.W. wrote the manuscript. L.G. and R.M. provide constructive advice for the manuscript. All authors discussed the results and commented on the manuscript.

Supplemental data

The following materials are available in the online version of this article.

Supplemental Figure S1. Flower development and anthocyanin accumulation in the red and white lotus flowers.

Supplemental Figure S2. The expression patterns of 110 genes in the 2 target intervals between the red and white petals at developmental stages 2 (S2) and stage 3 (S3).

Supplemental Figure S3. Digital expression levels of 3 genes in the 80 kb PAV in red petals at different developmental stages.

Supplemental Figure S4. PCR genotype verification of the F_2 individuals used in this study.

Supplemental Figure S5. GO enrichment of the 170 DEGs.

Supplemental Figure S6. Phylogenetic analysis and expression patterns of GSTs in *N. nucifera*.

Supplemental Figure S7. Pearson correlation analysis between candidate TFs and *NnGST2*.

Supplemental Figure S8. Transient expression of *NnMYB5* induces anthocyanin accumulation in the white lotus petals.

Supplemental Figure S9. The Y1H and dual-luciferase reporter assay (LUC) validation of NnMYB5 regulatory activity on 4 anthocyanin biosynthesis genes.

Supplemental Table S1. Primers used in this study.

Supplemental Table S2. List of natural lotus accessions tested in Fig. 2B.

Supplemental Table S3. Summary statistics of clean reads and RNA-seq mapping results.

Supplemental Table S4. List of the 170 overlapping DEGs between the white and red petal at developmental stages 2 and 3.

Supplemental Table S5. Gene ID and accession numbers of *GSTs* used for phylogenetic analysis.

Funding

This research was financially supported by the Biological Resources Programme, Chinese Academy of Sciences (KFJ-BRP-007-009) and the National Natural Science Foundation of China (Grant No.32070336).

Conflict of interest statement. None declared.

References

Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R, Walbot V. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. Plant Cell 1998:**10**(7):1135–1149. https://doi.org/10.1105/tpc.10.7.1135

Castillejo C, Waurich V, Wagner H, Ramos R, Oiza N, Muñoz P, Triviño JC, Caruana J, Liu Z, Cobo N, et al. Allelic variation of MYB10 is the major force controlling natural variation in skin and flesh color in strawberry (*Fragaria* spp.) fruit. Plant Cell 2020:**32**(12):3723–3749. https://doi.org/10.1105/tpc.20.00474

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020:**13**(8):1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Chen S, Xiang Y, Deng J, Liu Y, Li S. Simultaneous analysis of anthocyanin and non-anthocyanin flavonoid in various tissues of different lotus (*Nelumbo*) cultivars by HPLC-DAD-ESI-MS(n). PLoS One 2013:**8**(4):e62291. https://doi.org/10.1371/journal.pone.0062291

Chiang C, Scott AJ, Davis JR, Tsang EK, Li X, Kim Y, Hadzic T, Damani FN, Ganel L, Montgomery SB, et al. The impact of structural variation on human gene expression. Nat Genet. 2017:**49**(5):692–699. https://doi.org/10.1038/ng.3834

Deng J, Chen S, Yin X, Wang K, Liu Y, Li S, Yang P. Systematic qualitative and quantitative assessment of anthocyanins, flavones and flavonols in the petals of 108 lotus (*Nelumbo nucifera*) cultivars. Food Chem. 2013:**139**(1–4):307–312. https://doi.org/10.1016/j.foodchem.2013.02.010

Deng J, Fu Z, Chen S, Damaris RN, Wang K, Li T, Yang P. Proteomic and epigenetic analyses of lotus (*Nelumbo nucifera*) petals between red and white cultivars. Plant Cell Physiol. 2015:**56**(8):1546–1555. https://doi.org/10.1093/pcp/pcv077

Edwards R, Dixon DP. Plant glutathione transferases. In: Sies H, Packer L, editors. Methods in enzymology. San Diego (CA): Academic Press; 2005. p. 169–186

Gao Z, Yang X, Chen J, Rausher MD, Shi T. Expression inheritance and constraints on cis- and trans-regulatory mutations underlying lotus color variation. Plant Physiol. 2022:**191**(3):1662–1683. https://doi.org/10.1093/plphys/kiac522

Guo J, Deng C, Li Y, Zhu G, Fang W, Chen C, Wang X, Wu J, Guan L, Wu S, et al. An integrated peach genome structural variation map uncovers genes associated with fruit traits. Genome Biol. 2020:**21**(1):58. https://doi.org/10.1186/s13059-020-02169-y

Jiang S, Chen M, He N, Chen X, Wang N, Sun Q, Zhang T, Xu H, Fang H, Wang Y, et al. MdGSTF6, activated by MdMYB1, plays an essential role in anthocyanin accumulation in apple. Hortic Res. 2019:**6**(1):40. https://doi.org/10.1038/s41438-019-0118-6

Kitamura S, Shikazono N, Tanaka A. TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and

proanthocyanidins in Arabidopsis. Plant J. 2004:**37**(1):104–114. https://doi.org/10.1046/j.1365-313X.2003.01943.x

Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016:**33**(7): 1870–1874. https://doi.org/10.1093/molbev/msw054

Lin Z, Zhang C, Cao D, Damaris RN, Yang P. The latest studies on Lotus (*Nelumbo nucifera*)-an emerging horticultural model plant. Int J Mol Sci. 2019:**20**(15):3680. https://doi.org/10.3390/ijms20153680

Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou GA, Zhang H, Liu Z, Shi M, et al. Pan-genome of wild and cultivated soybeans. Cell 2020:**182**(1):162–176.e113. https://doi.org/10.1016/j.cell.2020.05.023

Liu Y, Song H, Zhang M, Yang D, Deng X, Sun H, Liu J, Yang M. Identification of QTLs and a putative candidate gene involved in rhizome enlargement of Asian lotus (*Nelumbo nucifera*). Plant Mol Biol. 2022:**110**(1–2):23–36. https://doi.org/10.1007/s11103-022-01281-w

Liu J, Wang Y, Zhang M, Wang Y, Deng X, Sun H, Yang D, Xu L, Song H, Yang M. Color fading in lotus (*Nelumbo nucifera*) petals is manipulated both by anthocyanin biosynthesis reduction and active degradation. Plant Physiol Biochem. 2022:**179**(1):100–107. https://doi.org/10.1016/j.plaphy.2022.03.021

Liu Q, Zhang D, Liu F-L, Liu Z, Wang X, Yang Y, Li S-S, Li H, Tian D, Wang L. Quercetin-derivatives paint the yellow petals of American lotus (*Nelumbo lutea*) and enzymatic basis for their accumulation. Horticultural Plant J. 2022:**9**(1):169–182. https://doi.org/10.1016/j.hpj.2022.02.001

Luo HF, Dai C, Li YP, Feng J, Liu ZC, Kang CY. Reduced anthocyanins in petioles codes for a GST anthocyanin transporter that is essential for the foliage and fruit coloration in strawberry. J Exp Bot. 2018:**69**(10):2595–2608. https://doi.org/10.1093/jxb/ery096

Marrs KA, Alfenito MR, Lloyd AM, Walbot V. A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. Nature 1995:**375**(6530):397–400. https://doi.org/10.1038/375397a0

Ming R, VanBuren R, Liu Y, Yang M, Han Y, Li LT, Zhang Q, Kim MJ, Schatz MC, Campbell M, et al. Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn. Genome Biol. 2013:**14**(5):R41. https://doi.org/10.1186/gb-2013-14-5-r41

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016:**11**(9):1650–1667. https://doi.org/10.1038/nprot.2016.095

Shen-Miller J. Sacred lotus, the long-living fruits of China antique. Seed Sci Res. 2007:**12**(3):131–143. https://doi.org/10.1079/SSR2002112

Shi T, Rahmani RS, Gugger PF, Wang M, Li H, Zhang Y, Li Z, Wang Q, Van de Peer Y, Marchal K, et al. Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. Mol Biol Evol. 2020:**37**(8): 2394–2413. https://doi.org/10.1093/molbev/msaa105

Sun S-S, Gugger PF, Wang Q-F, Chen J-M. Identification of a R2R3-MYB gene regulating anthocyanin biosynthesis and relationships between its variation and flower color difference in lotus (Nelumbo Adans.). Peerj. 2016:**4**(1):e2369. https://doi.org/10.7717/peerj.2369

Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant J. 2008:**54**(4):733–749. https://doi.org/10.1111/j.1365-313X.2008.03447.x

Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics. 2002: **Chapter 2**:Unit 2.3. https://doi: 10.1002/0471250953.bi0203s00

Vaish S, Gupta D, Mehrotra R, Mehrotra S, Basantani MK. Glutathione S-transferase: a versatile protein family. 3 Biotech. 2020:**10**(7):321. https://doi.org/10.1007/s13205-020-02312-3

Wang X, Chen X, Luo S, Ma W, Li N, Zhang W, Tikunov Y, Xuan S, Zhao J, Wang Y, et al. Discovery of a DFR gene that controls anthocyanin accumulation in the spiny Solanum group: roles of a natural promoter variant and alternative splicing. Plant J. 2022:**111**(4): 1096–1109. https://doi.org/10.1111/tpj.15877

Wang R, Lu N, Liu C, Dixon RA, Wu Q, Mao Y, Yang Y, Zheng X, He L, Zhao B, et al. MtGSTF7, a TT19-like GST gene, is essential for accumulation of anthocyanins, but not proanthocyanins in *Medicago truncatula*. J Exp Bot. 2022:**73**(12):4129–4146. https://doi.org/10.1093/jxb/erac112

Wu Y, Wu S, Shi Y, Jiang L, Yang J, Wang X, Zhu K, Zhang H, Zhang J. Integrated metabolite profiling and transcriptome analysis reveal candidate genes involved in the formation of yellow Nelumbo nucifera. Genomics 2022:**114**(6):110513. https://doi.org/10.1016/j.ygeno.2022.110513

Xie M, Chung CY, Li MW, Wong FL, Wang X, Liu A, Wang Z, Leung AK, Wong TH, Tong SW, et al. A reference-grade wild soybean genome. Nat Commun. 2019:**10**(1):1216. https://doi.org/10.1038/s41467-019-09142-9

Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. Trends Plant Sci. 2015:**20**(3):176–185. https://doi.org/10.1016/j.tplants.2014.12.001

Yang R-Z, Wei X-L, Gao F-F, Wang L-S, Zhang H-J, Xu Y-J, Li C-H, Ge Y-X, Zhang J-J, Zhang J. Simultaneous analysis of anthocyanins and flavonols in petals of lotus (*Nelumbo*) cultivars by high-performance liquid chromatography-photodiode array detection/electrospray ionization mass spectrometry. J Chromatogr A. 2009:**1216**(1): 106–112. https://doi.org/10.1016/j.chroma.2008.11.046

Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H, et al. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat Commun. 2019:**10**(1):1494. https://doi.org/10.1038/s41467-019-09518-x

Zhao J. Flavonoid transport mechanisms: how to go, and with whom. Trends Plant Sci. 2015:**20**(9):576–585. https://doi.org/10.1016/j.tplants.2015.06.007

Zheng P, Sun H, Liu J, Lin J, Zhang X, Qin Y, Zhang W, Xu X, Deng X, Yang D, et al. Comparative analyses of American and Asian lotus genomes reveal insights into petal color, carpel thermogenesis and domestication. Plant J. 2022:**110**(5):1498–1515. https://doi.org/10.1111/tpj.15753

Zhou H, Ma R, Gao L, Zhang J, Zhang A, Zhang X, Ren F, Zhang W, Liao L, Yang Q, et al. A 1.7-Mb chromosomal inversion downstream of a PpOFP1 gene is responsible for flat fruit shape in peach. Plant Biotechnol J. 2021:**19**(1):192–205. https://doi.org/10.1111/pbi.13455