The Genetic Basis of Flower Color Differences in *Nicotiana tabacum*

11

Elizabeth W. McCarthy, Jacob B. Landis, Amelda Kurti, Amber J. Lawhorn and Amy Litt

Abstract

Nicotiana tabacum L accessions vary in flower color from light pink to magenta. The differences in flower color are attributable to differences in anthocyanin content. To determine the genetic basis of flower color differences, we generated transcriptomes and quantified transcript levels of flavonoid biosynthetic genes in four *N. tabacum* accessions and their diploid progenitors. High expression ratios of the flavonol synthase (*FLS*) gene to dihydroflavonol 4-reductase (*DFR*) gene are found in light-pink flowers, suggesting that competition between the FLS and DFR enzymes for the same substrates may drive the flux of the flavonoid biosynthetic pathway toward producing flavonols at the expense of anthocyanins, resulting in light-pink flowers. The high *FLS*:*DFR* expression ratio appears to be attributable to *DFR* activation later in development in light-pink flowers.

11.1 Introduction

The genus *Nicotiana* (Solanaceae) consists of 76 species, which have flowers with tubular corollas that display diverse colors, including white, pink, magenta, purple, red, green, yellow, and white with ultraviolet light reflectance (Goodspeed 1954; McCarthy et al. 2015). Approximately half of *Nicotiana* species are allopolyploids that arose from both whole-genome duplication and interspecific hybridization (Chase et al. 2003; Knapp et al. 2004; Clarkson et al. 2004, 2010). These hybrid species may be predicted to display intermediate phenotypes between their diploid progenitors; however, *Nicotiana* allotetraploids can show transgressive floral colors (which fall outside the range of the diploid progenitors), and related polyploid accessions from the same origin can display floral color variation (McCarthy et al. 2015). We are interested in the genetic basis of these floral color differences.

Nicotiana tabacum L originated approximately 0.6 million years ago from maternal *N.*

E. W. McCarthy · J. B. Landis · A. Kurti · A. J. Lawhorn · A. Litt (✉)
Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA 92521, USA
e-mail: amy.litt@ucr.edu

E. W. McCarthy
Department of Biological Sciences, SUNY Cortland, Cortland, NY 13045, USA

J. B. Landis
School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca, NY 14853, USA

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_11

sylvestris and paternal *N. tomentosiformis* progenitors (Clarkson et al. 2004, 2010, 2017). First-generation allotetraploid lines that share the same parentage as natural *N. tabacum* were created synthetically by K. Y. Lim, Queen Mary University of London (London, UK). Floral color in accessions of natural *N. tabacum* ranges from light pink to magenta, and independent synthetic lines also display floral color differences, although the range is limited from light pink to dark pink (McCarthy et al. 2015, 2017). To determine the gene expression differences underlying this variation in floral color, we focused on two natural *N. tabacum* accessions (*N. tabacum* 095-55, magenta flowers, termed 095-55, and *N. tabacum* "Chulumani", light-pink flowers, termed Chulumani), two first-generation synthetic lines (QM24, light-pink flowers, and QM25, dark-pink flowers), and their diploid progenitors, *N. sylvestris* (white flowers) and *N. tomentosiformis* (dark-pink flowers) (Fig. 11.1).

In many plant species, anthocyanin pigments are responsible for the red, pink, purple, and blue floral colors. Equally important for pollination are flavonols, which are not visible to humans, but absorb ultraviolet light and thereby provide visual cues to birds and bees. In a previous study, we quantified anthocyanidins (anthocyanins without attached sugars) and flavonols in *N. sylvestris*, *N. tomentosiformis*, 095-55, Chulumani, QM24, and QM25 to determine the biochemical basis of floral color differences. The predominant anthocyanidin in all samples was cyanidin, which provides a pink color. The amount of cyanidin was consistent with flower

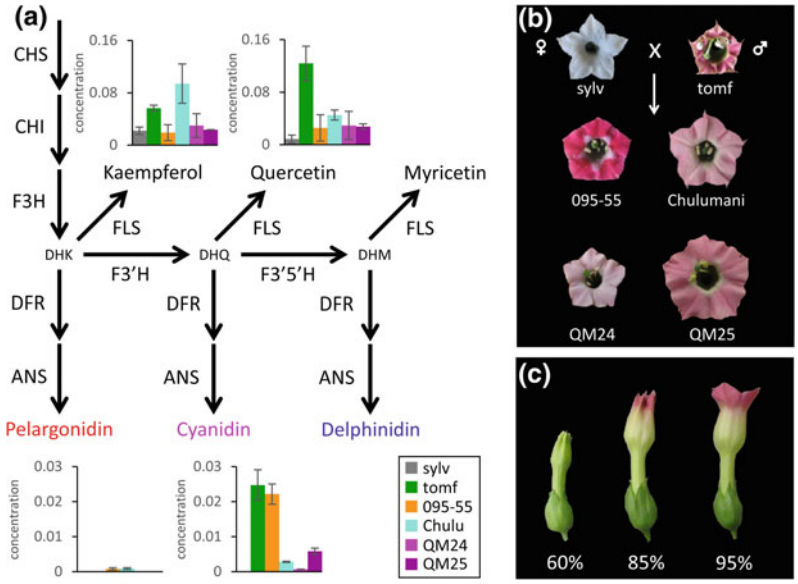

Fig. 11.1 Flavonoid biosynthetic pathway and *Nicotiana* flower colors and pigments. **a** Flavonoid biosynthetic pathway with structural enzymes in large font next to arrows; reproduced from McCarthy et al. (2017). CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase. Shared substrates appear in small font (DHK, dihydrokaempferol; DHQ, dihydroquercetin; DHM, dihydromyricetin), and resulting flavonol (black text) and anthocyanin (colored text) pigments. Bar graphs represent the pigment concentration (mg/g fresh tissue) of kaempferol, quercetin, pelargonidin, and cyanidin in the *Nicotiana* accessions examined in this chapter. **b** Floral limb photographs of the *Nicotiana* accessions used. *Nicotiana sylvestris* (white) and *N. tomentosiformis* (dark pink) are the maternal and paternal progenitors, respectively, of *N. tabacum*. Natural tobacco accessions: *N. tabacum* 095-55 (dark pink) and *N. tabacum* "Chulumani" (light pink). Synthetic lines QM24 (light pink) and QM25 (dark pink). **c** Floral bud photographs of *N. tabacum* 095-55 at 60, 85, and 95% of anthesis length, the developmental time points used in this study. 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

color: low levels in the light-pink flowers of Chulumani and QM24, none in the white flowers of *N. sylvestris*, and high levels in the dark-pink flowers of *N. tomentosiformis* and QM25 and the magenta flowers of 095-55 (McCarthy et al. 2017; Fig. 11.1). In addition to cyanidin, the natural *N. tabacum* accessions contain a small amount of transgressive pelargonidin, which is not present in either progenitor or in the synthetic lines. Floral tissue in all accessions contains the flavonols kaempferol and quercetin. Kaempferol levels are low in *N. sylvestris*, 095-55, and QM25, high in *N. tomentosiformis* and Chulumani, and intermediate in QM24. Quercetin levels are low in *N. sylvestris*, high in *N. tomentosiformis*, and intermediate in natural and synthetic *N. tabacum* accessions (McCarthy et al. 2017; Fig. 11.1).

Anthocyanin and flavonol pigments are produced by the flavonoid biosynthetic pathway, a branched pathway that yields multiple pigment types (Grotewold 2006; Fig. 11.1). The enzymes chalcone synthase (CHS), chalcone isomerase (CHI), and flavanone 3-hydroxylase (F3H) act early in the pathway and are necessary for the production of both flavonols and anthocyanins. The enzymes flavonoid 3′-hydroxylase (F3′H) and flavonoid 3′,5′-hydroxylase (F3′5′H) shift the pathway along different branches toward different pigments. Flavonol synthase (FLS) produces flavonols from the intermediate substrates dihydrokaempferol, dihydroquercetin, and dihydromyricetin, whereas dihydroflavonol 4-reductase (DFR) and anthocyanin synthase (ANS) convert these substrates into anthocyanidins, which can then be converted to anthocyanins (Grotewold 2006). The transcription factors anthocyanin 1 (AN1, basic helix-loop-helix), anthocyanin 2 (AN2, R2R3 MYB), and WD40 act together as a complex to induce anthocyanin production (Grotewold 2006).

Many evolutionary changes in floral color have been attributed to mutations in genes of the flavonoid biosynthetic pathway. Floral color shifts from blue to red are the result of inactivation, down-regulation, or deletion of *F3′5′H* or *F3′H* genes, resulting in production of red pelargonidin pigmentation instead of blue cyanidin or delphinidin (Zufall and Rausher 2004; Streisfeld and Rausher 2009; Des Marais and Rausher 2010; Smith and Rausher 2011; Wessinger and Rausher 2015). Several studies of the genetic basis of shifts from pigmented to white flowers have identified the cause as inactivation of the R2R3 MYB gene (orthologous to *AN2* in *Nicotiana*) that regulates the pathway (Quattrocchio et al. 1999; Schwinn et al. 2006; Hoballah et al. 2007). Some studies, however, have found that partial reduction or complete abolition of anthocyanin pigment is driven by repression of the pathway by an R3 MYB transcription factor (Yuan et al. 2013; Gates et al. 2017). Our questions focus on the genetic basis of the variety of floral colors observed in *N. tabacum* accessions and are as follows: (1) Do the expression levels of genes involved in anthocyanin synthesis (*DFR*, *ANS*, *AN1*, and *AN2*) predict cyanidin concentration, i.e., is the expression of these genes up-regulated in dark-pink flowers? (2) Are light-pink flowers lighter than dark-pink flowers because *DFR* and *ANS* are activated later in development, resulting in lower cyanidin accumulation? (3) In pairwise comparisons of dark-pink and light-pink flowers, are the same genes involved in the flavonoid biosynthetic pathway differentially expressed?

11.2 Materials and Methods

11.2.1 Plant Growth and Material

We grew plant material in greenhouses with natural sunlight and temperatures between 10 and 30 °C. We used the following plant material: *N. sylvestris* A04750326 (Radboud University, Nijmegen, The Netherlands), *N. tomentosiformis* BRNO 4103 (acquired from A. Kovařík, Brno, Czech Republic), *N. tabacum* 095-55 (IPK Gatersleben, Gatersleben, Germany), *N. tabacum* "Chulumani" (collected in the field in Bolivia by S. Knapp), and two first-generation synthetic lines, QM24 and QM25 (created by K. Y. Lim at Queen Mary University of London by crossing 4× autotetraploid *N. sylvestris* and 4× autotetraploid *N. tomentosiformis*). Because we used multiple accessions of *N. tabacum*, we will refer to all six plant lines as accessions.

Table 11.1 Primer sequences used for digital droplet PCR

Gene	Sequence (5′→3′)
DFR F	CGTCACTGGAGCAGCTGGAT
DFR R	TCTCAGGATCACGAACAGTAGCG
FLS F	TGAAGGGTATGGTACTTCTTTGCAGA
FLS R	CGATAATTGATGGCAGAAGGAGGC
GAPDH F	TGACAGATTTGGAATTGTTGAGGGTCT
GAPDH R	CTTCCACCCCTCCAGTCCTTG
EF1α F	CTTGGTGGTATTGACAAGCGTGT
EF1α R	TGTCAAGCACCCAGGCATACT
L25 F	GGACAAAAGTTACATTCCACCGACC
L25 R	AGTTTGTTCCTTCCAGGTGCACT

11.2.2 Developmental Series, RNA Extraction, and cDNA Synthesis

We examined floral development in each accession to identify stages for use in transcriptome analyses by timing of the appearance of floral pigment. We measured corolla tube length, including the floral limb (the portion of the flower that opens at anthesis), at anthesis in 5–10 flowers per accession by dissecting the corolla tube, pinning it flat, photographing it, and measuring length using ImageJ (version 1.51 k; Rasband 1997). We used the mean tube length at anthesis to calculate the percentage of anthesis length at which pigment appeared for each accession. At 60% of anthesis length, none of the accessions displayed floral pigmentation, but different accessions began to show pigmentation by 85% or 95% of anthesis length. We therefore selected 60, 85, and 95% of anthesis corolla length as stages for transcriptome analysis. Floral buds were collected using the corolla length values ±1 mm. We dissected the corolla tissue, cut the stamens from the corolla, and placed the tissue in liquid nitrogen. We collected three biological replicates from different plants, where available, for each accession at each developmental time point. Each independent synthetic accession consisted of a single first-generation individual, so three replicate flowers were collected from each plant for these lines. Because of the scarcity of material, we collected only two biological replicates for *N. tomentosiformis* at 85% of anthesis length and two flowers for the synthetic line QM24 at 95% of anthesis length.

We extracted RNA using the RNeasy Mini Plant Kit (Qiagen, Hilden, Germany), DNase-treated using the Turbo-DNase Kit (Ambion, Thermo Fisher Scientific, Waltham, MA, USA), and synthesized cDNA from 1 μg of RNA using the SuperScript III Reverse Transcriptase Kit (Invitrogen, Carlsbad, CA, USA) with the provided oligo-dT primer.

11.2.3 Digital Droplet PCR (ddPCR)

To quantify *FLS* and *DFR* transcript accumulation, we performed ddPCR on a QX200 ddPCR system (Bio-Rad Laboratories, Hercules, CA, USA). We used QX200 ddPCR EvaGreen SuperMix and 2 pmol of each primer with the following program: 95 °C for 5 min, followed by 40 cycles of 95 °C for 30 s and 60 °C for 60 s, followed by 4 °C for 5 min and 90 °C for 5 min, using a ramp rate of 2 °C per second. We normalized transcript accumulation using the geometric mean of three reference genes: elongation factor 1 alpha (*EF1α*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), and ribosomal protein gene *L25*. Table 11.1 lists the primer sequences. We analyzed the ddPCR data using QuantaSoft Analysis Pro, version 1.0 (Bio-Rad Laboratories).

11.2.4 Illumina Library Preparation and Sequencing

To prepare cDNA libraries for 60, 85, and 95% of anthesis length samples from all six accessions, strand-specific libraries were constructed from mRNA as previously described (Zhong et al. 2011) and sequenced on a HiSeq 2500 system (Illumina, San Diego, CA) with 1 × 85 bp reads. Sequencing data from a smaller, preliminary experiment was also included to create our reference assembly (see below in the "Transcriptome analyses" section for details), using corolla tissue at anthesis from three biological replicates each of *N. tabacum* 095-55 and *N. tabacum* 51789 (this accession was not used in the larger analysis). Libraries were prepared with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's protocols. Library quality was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and libraries were sequenced on an Illumina NextSeq 500 platform at the University of California, Riverside Institute for Integrated Genome Biology Genomics Core Facility with 1 × 150 bp reads.

11.2.5 Transcriptome Analyses

We processed raw sequencing reads on the University of California, Riverside Biocluster by trimming adaptors using Cutadapt (version 1.15; Martin 2011), followed by filtering with a quality score of 30 and a minimum length of 60 bp using Sickle (version 1.33; Joshi and Fass 2011). We initially performed analyses of cleaned reads using a genome-guided approach in TopHat (version 2.1.1), Cufflinks (version 2.2.1; Trapnell et al. 2012), and cummerbund (version 2.20; Goff et al. 2013) with the available genomes of *N. sylvestris* (Sierro et al. 2013), *N. tabacum* (Edwards et al. 2017), and *N. tomentosiformis* (Sierro et al. 2013). Although *N. sylvestris* and *N. tomentosiformis* are the progenitors of the allotetraploid *N. tabacum*, the percentage of reads mapped varied widely (20–70%) both within and among accessions and depending on which genomes were used as the reference. We therefore adopted a de novo approach using the Trinity pipeline (version 2.4; Grabherr et al. 2011; Haas et al. 2013).

We generated a single reference assembly using all sequencing reads (60, 85, and 95% of anthesis length from the four *N. tabacum* accessions, *N. sylvestris*, and *N. tomentosiformis* and at anthesis from *N. tabacum* 095-55 and *N. tabacum* 51789), and normalized them in silico to 50× to improve the efficiency of the assembly by decreasing the number of reads included (Haas et al. 2013). We mapped sequencing reads back to the reference assembly using RNA-Seq by Expectation Maximization implemented with Bowtie2 (version 2.2.9; Langmead and Salzberg 2012). We annotated the reference assembly using the Transdecoder (version 3.0) plugin in Trinity by performing BLASTx and BLASTp searches against the Swiss-Prot (The UniProt Consortium 2017) and Pfam (Finn et al. 2016) databases. Because of short read lengths, it was difficult to assemble full gene transcripts; therefore, any given gene locus may have several corresponding contigs in the de novo assembly. Additionally, likely due to short read lengths and nonuniformity of polymorphisms between the *N. sylvestris* and *N. tomentosiformis* genomes, there were difficulties isolating homeologous sequences in the de novo assembly; therefore, some contigs correspond to *N. sylvestris* copies, some contigs correspond to *N. tomentosiformis* copies, and some correspond to both homeologs.

We performed differential expression analyses with the bioconductor package limma/voom (version 3.6; Ritchie et al. 2015). We ran two sets of comparisons, across accessions within developmental time points and across developmental time points within accessions. To ensure robust differences, we classified differentially expressed contigs as those with p-value < 0.05, false discovery rate < 0.05, and log2 fold change > 2 (absolute value).

11.2.6 Identifying Flavonoid Biosynthesis Transcripts in the de novo Assembly

To identify contigs in the de novo assembly that correspond to genes in the flavonoid biosynthetic pathway, we performed BLAST searches using previously identified flavonoid biosynthetic genes from *Iochroma* and *Petunia* species (both in family Solanaceae) to find putative orthologs from *N. sylvestris* and *N. tomentosiformis* in GenBank. A single hit was found for each gene in the flavonoid biosynthetic pathway for *N. sylvestris* and *N. tomentosiformis*, apart from *CHS* for which two genes were found. We then executed custom BLAST searches in Geneious (version 10, Auckland, New Zealand) with the detected *N. sylvestris* and *N. tomentosiformis* orthologs against the de novo reference assembly to identify the contigs of genes likely involved in flavonoid biosynthesis in *Nicotiana*. We extracted these flavonoid biosynthesis contigs from read count data for each transcriptome and from the differential expression analyses for use in additional investigations.

11.2.7 Comparative Quantitative Expression of Flavonoid Biosynthetic Genes

To determine whether increase in *DFR* and *ANS* expression accounts for the higher cyanidin concentration observed in dark-pink-flowered species, we generated strip charts in R (version 3.4.2) using the ggplot2 package (version 2.2.1; Wickham 2009). We plotted the log2 fold change values of differentially expressed flavonoid biosynthetic genes for pairwise comparisons between accessions within each developmental time point. To establish whether the magnitude and direction of expression differences in these transcripts reflected those observed in the transcriptome overall, we overlaid these strip charts onto violin plots (which display the distribution of the data in a manner similar to that of a histogram) of the differentially expressed genes (DEGs) from the whole transcriptome using ggplot2. As our findings suggested that some floral color differences involved more complex dynamics than simply increased levels of *DFR* and *ANS*, we also examined relative expression dynamics across the pathway to identify patterns that involved more than a single gene. To do this, we summed the transcript levels for each gene (e.g., *CHS*) and produced heatmaps in R using the mixOmics package (version 6.3.1; Cao et al. 2016). To further investigate whether competition for the same substrates may affect pigment concentrations, we calculated the *FLS*:*DFR* ratio using both transcriptome and ddPCR data and compared it to floral phenotype and total anthocyanidin concentration.

11.2.8 Comparative Quantitative Expression of Flavonoid Biosynthetic Genes in Light-Pink and Dark-Pink Flowers

To determine whether accessions with the same floral color display similar relative transcript levels of flavonoid biosynthetic genes, we plotted Venn diagrams using the R packages Venn diagram (version 1.6.18; Chen and Boutros 2011) and gplots (version 3.0.1; Warnes et al. 2009). We compared DEGs between *N. tabacum* accessions and each of their progenitors to identify those shared by light-pink or dark-pink flowers that differ from progenitor expression levels. We also examined overlap in pairwise comparisons between *N. tabacum* accessions with light-pink and dark-pink flowers (095-55 versus Chulumani, 095-55 versus QM24, and Chulumani versus QM25).

11.2.9 Comparative Quantitative Expression of Flavonoid Biosynthetic Genes Across Development

To establish whether differences in developmental timing of *DFR* and *ANS* activation

underlie changes in floral color, we examined patterns of expression of flavonoid biosynthetic genes across development within each accession by generating strip charts overlaid on violin plots and heatmaps, as described in the Sect. 11.2.7.

11.3 Results

11.3.1 Patterns of Differential Gene Expression Among Accessions and Across Development

To elucidate the genetic basis of floral color variation among white, light-pink, and dark-pink *Nicotiana* flowers, we used transcriptome analyses and ddPCR to quantify the expression of flavonoid biosynthetic pathway genes from two natural and two synthetic accessions of *N. tabacum* and their diploid progenitors. Because of difficulty in mapping the transcriptomes to the publicly available *Nicotiana* genomes, we created a de novo assembly for sequence analysis. Sequencing coverage was low for one biological replicate of QM24 at 60% of anthesis length; therefore, we excluded this replicate from further analyses. The number of reads mapped from each replicate ranged from 3.66 to 9.42 million (Table 11.2).

In pairwise comparisons of accession transcriptomes at a given developmental time point, we found the highest number of DEGs between progenitor species and the smallest number between synthetic lines (Table 11.3). There was also a higher number of DEGs between *N. tabacum* accessions and their progenitors than between *N. tabacum* accessions (Table 11.3). In comparisons across developmental time points within each accession, the highest number of DEGs was found between 60 and 95% of anthesis length, followed by 60 and 85% and between 85 and 95% (Table 11.4). Not surprisingly, far more genes were differentially expressed between accessions than across development within an accession.

11.3.2 Higher Levels of Anthocyanin-Specific Biosynthetic Transcripts Are Correlated with Pink Versus White Phenotype Comparisons, but Variation Between Pink Phenotypes Is More Complex

To determine whether relatively higher expression levels of *DFR* and *ANS* in dark-pink flowers might underlie differences in floral color, we compared DEGs involved in the flavonoid biosynthetic pathway between pairs of accessions at each developmental time point (Fig. 11.2). In comparisons between white (*N. sylvestris*) and pink (all other accessions) flowers, most of the differentially expressed flavonoid biosynthetic transcripts were higher in pink flowers. These included genes that produce anthocyanins (*DFR*, *ANS*, *AN1*, and *AN2*) as we hypothesized, but also other genes in the pathway (*CHS*, *CHI*, *F3H*, *F3'H*, and *FLS*). In comparisons between light-pink and dark-pink flowers, some of the DEGs were up-regulated in lighter flowers, whereas others were up-regulated in darker flowers. This contrasted with the findings for comparisons involving *N. sylvestris* (Fig. 11.2), suggesting that the production of dark- versus light-pink flowers is not simply the result of higher expression of anthocyanin biosynthetic genes, but that more complex expression dynamics in the pathway are involved.

To determine whether relative expression levels across the pathway as a whole are consistent with a role in determining different color phenotypes, we generated heatmaps for the summed contig counts for each flavonoid biosynthetic gene across accessions. *CHS* expression levels were higher than expression levels in the other genes, so to facilitate interpretation of expression patterns among the genes present at lower levels, we also generated heatmaps without *CHS* (Fig. 11.3). Some patterns were consistent with white/light-pink (*N. sylvestris*, Chulumani, and QM24) versus dark-pink

Table 11.2 Number of reads mapped to the de novo reference assembly

Sample	Reads mapped
sylv a 60	5,674,537
sylv b 60	6,280,955
sylv c 60	5,309,214
sylv a 85	4,224,300
sylv b 85	8,116,442
sylv c 85	8,450,752
sylv a 95	5,207,567
sylv b 95	8,235,247
sylv c 95	4,501,618
tomf a 60	6,540,640
tomf b 60	4,660,980
tomf c 60	6,062,876
tomf a 85	6,782,341
tomf b 85	8,585,868
tomf a 95	7,746,705
tomf b 95	6,238,237
tomf c 95	6,602,617
095-55 a 60	6,439,524
095-55 b 60	4,198,419
095-55 c 60	8,176,370
095-55 a 85	5,331,654
095-55 b 85	6,718,258
095-55 c 85	7,064,007
095-55 a 95	6,138,186
095-55 b 95	7,087,995
095-55 c 95	7,136,716
Chulu a 60	4,296,617
Chulu b 60	5,880,425
Chulu c 60	5,626,046
Chulu a 85	6,004,335
Chulu b 85	6,861,648
Chulu c 85	9,198,169
Chulu a 95	5,297,129
Chulu b 95	8,899,176
Chulu c 95	5,928,235
QM24a 60	4,957,023
QM24c 60	9,365,190
QM24a 85	9,415,003
QM24b 85	8,618,508
QM24c 85	4,650,289
QM24a 95	6,162,663
QM24b 95	6,400,574

(continued)

Table 11.2 (continued)

Sample	Reads mapped
QM25a 60	5,161,262
QM25b 60	6,854,781
QM25c 60	4,610,085
QM25a 85	4,661,285
QM25b 85	3,854,284
QM25c 85	4,946,783
QM25a 95	5,768,444
QM25b 95	3,660,452
QM25c 95	5,885,153

095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

(*N. tomentosiformis*, 095-55, and QM25) floral color categories. Early in development (60% of anthesis length), *DFR* levels were lower in white/light-pink flowers. At 60 and 85% of anthesis length, *F3H* levels were higher in white/light-pink flowers. We also saw, however, patterns that did not correspond to floral color categories. The low *DFR* and *F3H* levels were not maintained in light-pink flowers at later developmental stages. In addition, at 60 and 85% of anthesis length, *FLS* expression was higher in *N. tomentosiformis*, Chulumani, and both synthetic lines than in *N. sylvestris* and 095-55, corresponding to lower concentration of flavonols in the white *N. sylvestris* and dark-pink 095-55 flowers. At 95% of anthesis length, *FLS* expression was similar across accessions. Expression of the regulatory genes *AN1* (at all

Table 11.3 Differentially expressed contigs across the transcriptomes of *Nicotiana* accessions

Comparison	60% of anthesis length	85% of anthesis length	95% of anthesis length
sylv versus tomf	35,517	32,920	35,638
095-55 versus sylv	16,037	15,930	18,661
Chulu versus sylv	16,352	17,260	17,171
QM24 versus sylv	10,771	11,323	11,895
QM25 versus sylv	12,080	14,726	14,316
095-55 versus tomf	18,643	17,869	20,353
Chulu versus tomf	17,976	18,896	18,918
QM24 versus tomf	13,615	14,420	11,268
QM25 versus tomf	13,430	14,778	14,562
QM24 versus 095-55	7,478	7,308	5,487
QM24 versus Chulu	6,212	7,573	4,012
QM25 versus 095-55	6,207	7,592	7,257
QM25 versus Chulu	3,688	7,024	5,238
095-55 versus Chulu	510	297	35
QM24 versus QM25	0	0	0

095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

Table 11.4 Differentially expressed contigs across developmental time points for each *Nicotiana* accession

Comparison	*N. sylvestris*	*N. tomentosiformis*	*N. tabacum* 095-55	*N. tabacum* "Chulumani"	QM24	QM25
60% versus 85% of anthesis length	52	234	851	2,092	64	291
60% versus 95% of anthesis length	538	1,900	4,198	4,858	3,511	1,736
85% versus 95% of anthesis length	5	0	545	0	86	180

QM24 and QM25 are first-generation synthetic lines

developmental time points) and *AN2* (at 60 and 85% of anthesis length) was higher in *N. tomentosiformis*, 095-55, and Chulumani than in the other accessions (Fig. 11.3).

Because *FLS* and *DFR* encode proteins that function in different branches of the pathway and because their expression levels vary across accessions at 60% of anthesis length, we determined whether differences in the expression of these genes correlated with differences in floral color. Consistent with floral color differences, *DFR* levels at 60% of anthesis length were significantly higher in dark-pink flowers than in light-pink or white flowers (with the exception of 095-55 versus Chulumani). However, *DFR* levels were similar in white versus light-pink, light-pink versus light-pink, and dark-pink versus dark-pink comparisons (Fig. 11.2a). In contrast, *FLS* expression differences at 60% of anthesis length were not consistent with differences between flower colors; for example, in light-pink versus dark-pink comparisons, *FLS* expression was significantly higher in QM24 versus 095-55 but similar in Chulumani versus QM25 (Fig. 11.2a). Because FLS and DFR act on the same substrates and may compete for them, we calculated *FLS*:*DFR* expression ratios to establish whether this ratio was consistent with a role in determining floral color. In both the transcriptome and ddPCR data, the *FLS*:*DFR* ratio was high (>35 fold change) at 60% of anthesis length in light-pink flowers (Chulumani and QM24), whereas other accessions (with the exception of *N. sylvestris* in the ddPCR data) at 60% and all accessions at 85 and 95% of anthesis length had much lower *FLS*:*DFR* ratios (Fig. 11.4a, b). To determine whether the high ratios in light-pink (Chulumani and QM24) and white (*N. sylvestris*) flowers were consistent with lower levels of anthocyanidins, suggesting diversion of the pathway to flavonols by FLS, we plotted total anthocyanidin concentration against *FLS*:*DFR* ratio at 60% of anthesis length (Fig. 11.4c). Although the relationship was not linear, the accessions with high *FLS*:*DFR* ratios tended to have lower anthocyanidin concentration, suggesting that a high *FLS*:*DFR* ratio early in floral development may play a role in preventing the accumulation of anthocyanidins.

11.3.3 Different Pathway Modifications Underlie Shifts Between Light-Pink and Dark-Pink Flowers

To determine whether expression differences of specific genes in the flavonoid biosynthetic pathway at specific developmental time points underlie the differences between light-pink and dark-pink flowers, we used pairwise DEG comparisons between accessions. We generated Venn diagrams of all four *N. tabacum* accessions versus each of their diploid progenitors within developmental time points (Fig. 11.5a). We then extracted the identity of the contigs from the sectors of the Venn diagram that were shared between light-pink and between dark-pink accessions. Dark-pink accessions (095-55 and

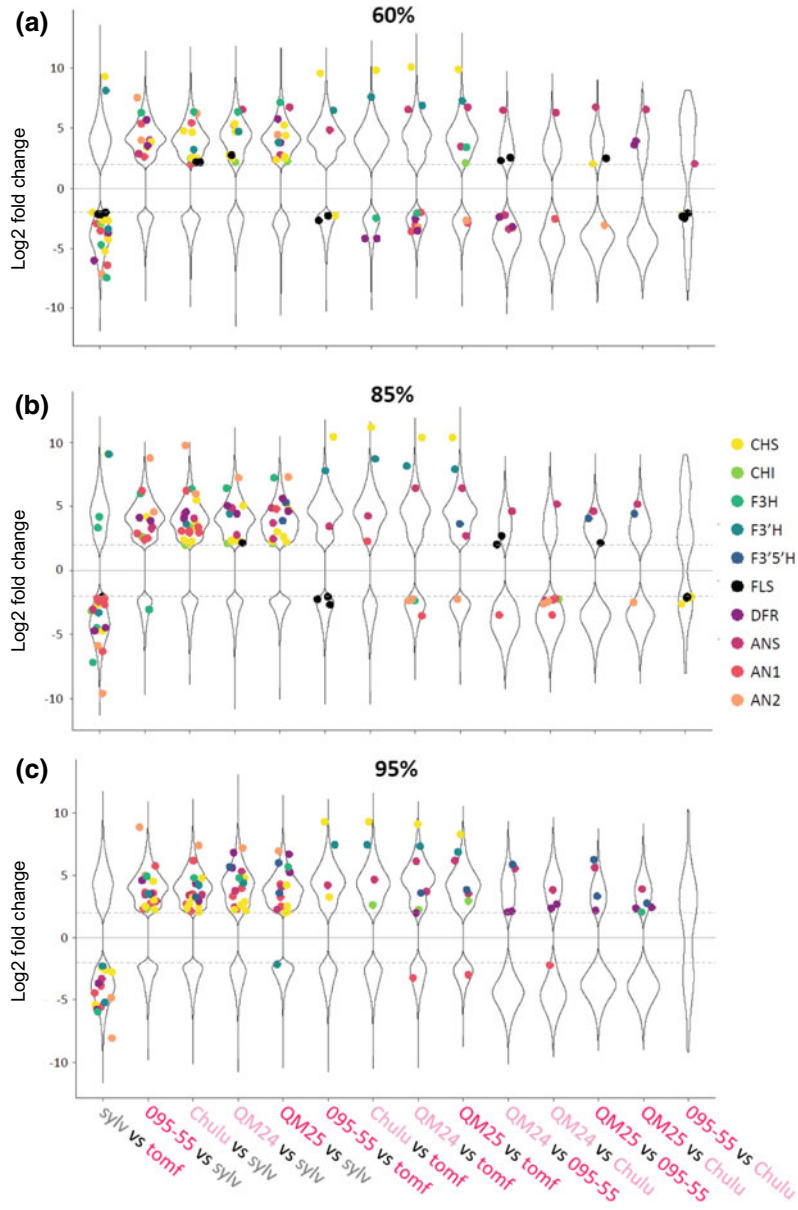

Fig. 11.2 Differential gene expression across *Nicotiana* accessions at **a** 60%, **b** 85%, and **c** 95% of anthesis length. The violin plots represent transcriptome-wide differentially expressed contigs (log2 fold change > 2 (absolute value)). The points represent differentially expressed contigs from the flavonoid biosynthetic pathway, colored for each gene. Dashed lines at 2 and −2 represent the cutoff for differentially expressed contigs. Pairwise comparisons are shown along the x-axis, and the accessions are color-coded according to flower color: gray = white; light pink = light pink; magenta = darkpink. 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines. Differentially expressed genes that are up-regulated in the first accession of the pair are in the positive (upper) section of each graph and those that are up-regulated in the second accession of the pair are in the negative (lower) section of each graph. 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; ANS, anthocyanidin synthase; AN1, anthocyanin 1; AN2, anthocyanin 2

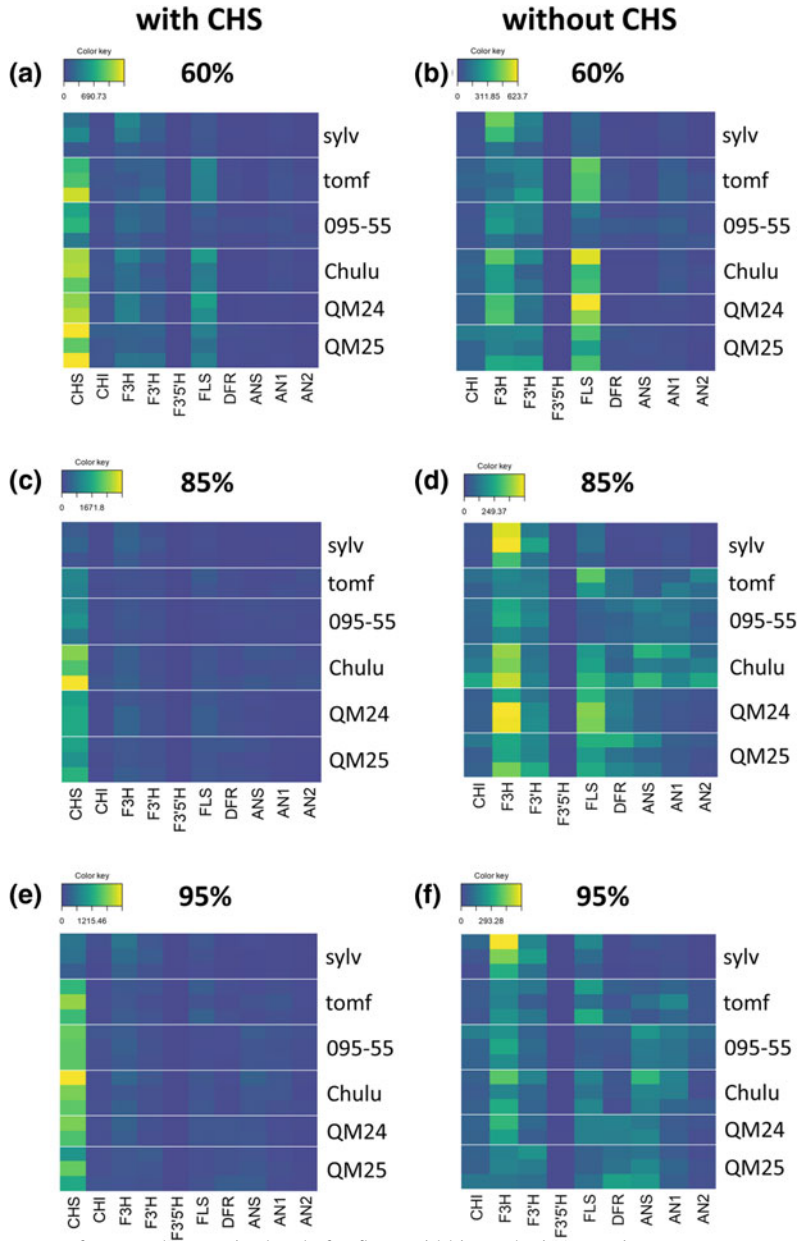

Fig. 11.3 Heatmaps of summed transcript levels for flavonoid biosynthetic genes in *Nicotiana* accessions. Transcript levels were compared at 60% (**a**, **b**), 85% (**c**, **d**), and 95% (**e**, **f**) of anthesis length, with (**a**, **c**, **e**) and without (**b**, **d**, **f**) *CHS*, which encodes 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; ANS, anthocyanidin synthase; AN1, anthocyanin 1; AN2, anthocyanin 2. Accessions: 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

QM25) had higher *DFR* and *ANS* expression levels than *N. sylvestris* (Fig. 11.5b), which was expected because of the presence of cyanidin in the dark-pink flowers (Fig. 11.1a). Similar differences were expected for light-pink versus white comparisons, perhaps to a lesser

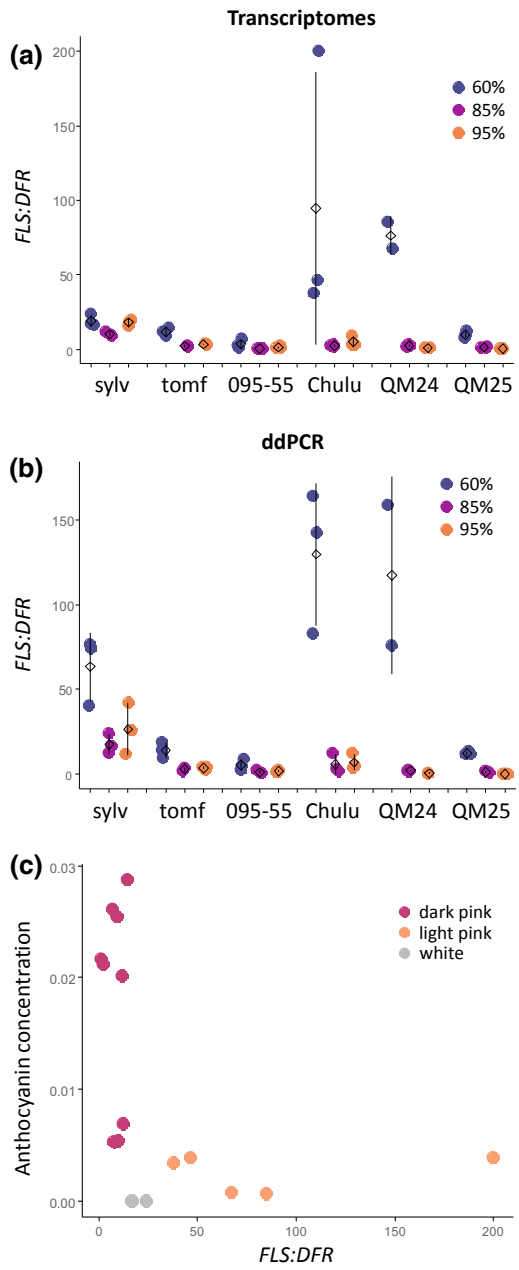

Fig. 11.4 *FLS:DFR* expression ratio from transcriptome (**a**) and digital droplet PCR (ddPCR) (**b**) data at 60%, 85%, and 95% of anthesis length in *Nicotiana* accessions. *FLS* encodes flavonol synthase and *DFR* encodes dihydroflavonol 4-reductase. Black diamonds and lines represent the mean and standard deviation, respectively. **c** Total anthocyanidin concentration versus transcriptome *FLS:DFR* ratio with points representing flower color. 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

magnitude, but this was not observed. Light-pink accessions (Chulumani and QM24) had higher *FLS* expression levels than *N. sylvestris*, but dark-pink accessions (095-55 and QM25) did not (Fig. 11.5b). The higher *FLS* expression levels exclusively in light-pink flowers were not expected. This suggests that in the development of light-pink flowers, the flux of the flavonoid biosynthetic pathway may be shunted toward the production of flavonols, whereas in dark-pink flowers it may be directed toward anthocyanins. In the comparisons of the *N. tabacum* accessions to *N. tomentosiformis*, light-pink (but not dark-pink) accessions had lower *F3'H* and *DFR* expression levels, whereas dark-pink (but not light-pink) accessions had higher *ANS* expression levels (Fig. 11.5b). These findings were consistent with lower cyanidin production in light-pink flowers but the higher *ANS* expression in dark-pink *N. tabacum* accessions does not seem to result in higher cyanidin production compared to *N. tomentosiformis*, perhaps because there is no accompanying increase in *DFR* expression levels.

To ascertain whether the same genes were differentially expressed in all comparisons of light-pink and dark-pink flowers, we generated Venn diagrams to compare differential expression in pairwise comparisons between light-pink and dark-pink *N. tabacum* accessions (095-55 versus Chulumani, 095-55 versus QM24, and Chulumani versus QM25). No DEGs were shared among all three comparison sets (Fig. 11.5c, d), suggesting that similar colors were achieved in different ways. In comparisons between 095-55 and both light-pink accessions at 60 and 85% of anthesis length, however, *FLS* expression levels were higher in the light-pink flowers. In contrast, light-pink flowers had lower *DFR* levels in 095-55 versus QM24 and QM25 versus Chulumani comparisons at 60% of anthesis length. Either high *FLS* or low *DFR* expression levels can result in a high *FLS:DFR* ratio, suggesting that the factor driving differences between light-pink and dark-pink flowers may be the ratio of *FLS* to *DFR* expression. No other consistent differences were detected between light-pink and dark-pink flowers, but

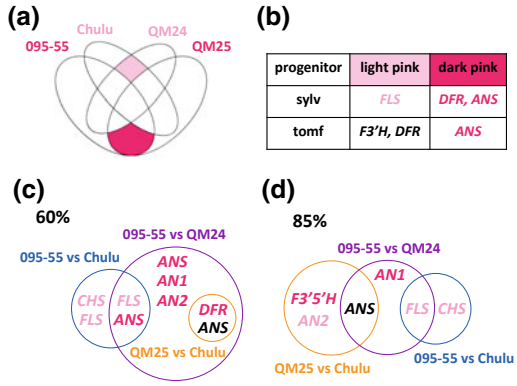

Fig. 11.5 Shared differentially expressed genes from the flavonoid biosynthetic pathway in comparisons between *Nicotiana* accessions. **a** Example Venn diagram of differentially expressed genes from pairwise comparisons of each *N. tabacum* accession with one diploid progenitor; analyses were performed for comparisons with both *N. sylvestris* and *N. tomentosiformis*, but only an example diagram is shown here. Light-pink and dark-pink sections highlight genes that are differentially expressed exclusively in light-pink accessions versus progenitor or dark-pink accessions versus progenitor, respectively. **b** Summary of the Venn diagram data for *N. tabacum* accessions versus progenitor at 60% of anthesis length. The genes in the table are those found in the light-pink and dark-pink sectors of (**a**) for comparisons between the *N. tabacum* accessions and *N. sylvestris* or *N. tomentosiformis*. These represent differentially expressed genes exclusive to light-pink or dark-pink accessions versus progenitor comparisons; pink text indicates higher expression in the *N. tabacum* accession, and black text indicates higher expression in the progenitor. (**c**, **d**) Venn diagrams of comparisons between light-pink and dark-pink *N. tabacum* accessions at 60% (**c**) and 85% (**d**) of anthesis length. Light-pink text represents genes that are up-regulated in light flowers; dark-pink text represents genes that are up-regulated in dark flowers; black text represents genes for which the expression levels do not correlate with flower color. Blue, purple, and orange labels and circles represent separate light-pink versus dark-pink comparisons. CHS, chalcone synthase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; AN1, anthocyanin 1; AN2, anthocyanin 2. Accessions: 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

lower levels of the regulatory genes *AN1* (at 60 and 85% of anthesis length) and *AN2* (at 60% of anthesis length) were observed in QM24 than in dark-pink 095-55, although not in other light-pink versus dark-pink comparisons. In addition, *CHS* was found at higher levels in the light-pink flowers of Chulumani versus 095-55 at 60 and 85% of anthesis length, but not in any other comparisons. This variation in results suggests that the mechanism behind the differences between light-pink and dark-pink floral color may be a combination of several factors.

11.3.4 DFR Up-Regulation Occurs Later in Development in Light-Pink Flowers

Gene expression comparisons between light-pink and dark-pink flowers found lower *DFR* expression in light-pink flowers at 60% of anthesis length; we therefore assessed whether light-pink floral color may be the result of activation of *DFR* later in development. *DFR* expression was up-regulated at 85% of anthesis length compared to 60% of anthesis length and remained at high levels at 95% in *N. tomentosiformis*, Chulumani, and both synthetic lines (Fig. 11.6a–f). In heatmaps showing relative transcript levels across development within each accession, very low levels of *DFR* expression were observed in *N. sylvestris*, Chulumani, and QM24 at 60% of anthesis length (Fig. 11.6g–l). To compare expression across development in all accessions, we generated strip plots of *DFR* expression levels from the six transcriptomes (Fig. 11.7). *N. sylvestris* displayed low levels of *DFR* throughout development, which is consistent with its lack of anthocyanidin pigmentation. *DFR* levels in *N. tomentosiformis*, 095-55, and QM25 were significantly higher than those of *N. sylvestris* at 60% of anthesis length (Fig. 11.2a), indicating that *DFR* had already been activated in these accessions at this time point. In contrast, light-pink-flowered accessions (Chulumani and QM24) at 60% of anthesis length had *DFR* expression levels similar to those of *N. sylvestris*, suggesting that activation was delayed in light-pink flowers (after 60% but before 85% of anthesis length). This delay in *DFR* activation

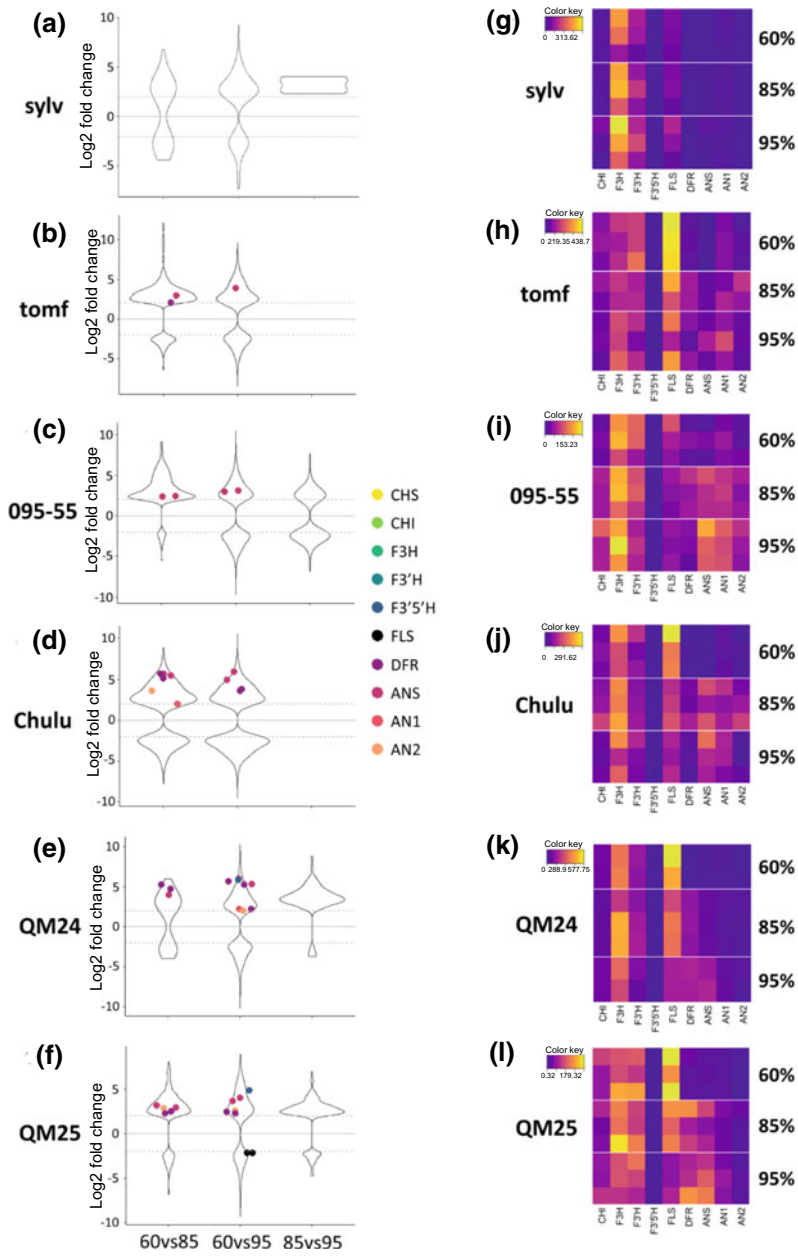

Fig. 11.6 Gene expression dynamics across developmental time points in *Nicotiana* accessions. Differentially expressed genes between different time points in **a** *N. sylvestris* (sylv), **b** *N. tomentosiformis* (tomf), **c** *N. tabacum* 095-55 (095-55), **d** Chulumani (Chulu), **e** the synthetic line QM24, and **f** the synthetic line QM25. The violin plots represent transcriptome-wide differentially expressed genes (log2 fold change > 2 (absolute value)). The points represent differentially expressed flavonoid biosynthesis genes, colored for each gene. Dashed lines at 2 and −2 represent the cutoff for differentially expressed genes. The positive (upper) portion of the graph represents genes that are up-regulated at the earlier developmental time point;the negative (lower) portion represents genes up-regulated at the later developmental time point. Heatmaps of summed transcript levels for flavonoid biosynthetic genes across development for **g** *N. sylvestris*, **h** *N. tomentosiformis*, **i** 095-55, **j** Chulumani, **k** QM24, and **l** QM25 without *CHS*, which encodes 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; ANS, anthocyanidin synthase; AN1, anthocyanin 1; AN2, anthocyanin 2

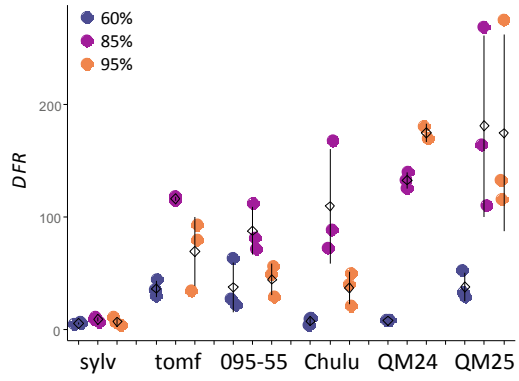

Fig. 11.7 *DFR* expression levels for each *Nicotiana* accession at 60, 85, and 95% of anthesis length. Black diamonds and lines represent the mean and standard deviation, respectively. 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

may underlie the increased *FLS*:*DFR* ratio observed in these accessions, which may play a role in the decrease of anthocyanidin accumulation.

11.4 Discussion

11.4.1 Higher Expression Levels of Anthocyanin-Specific Genes Differentiate White from Pink Flowers

In comparisons between white-flowered *N. sylvestris* and each pink accession, expression of some or all anthocyanin-specific genes (*DFR*, *ANS*, *AN1*, and *AN2*) and of other genes in the pathway was higher in pink flowers. These results suggest that the difference between white and pink flowers arises from changes in pathway regulation and/or from expression of the genes that produce anthocyanins, consistent with reports that overexpression of either *AN1* or *AN2* increases anthocyanin concentration (Pattanaik et al. 2010; Bai et al. 2011). However, no *AN2* ortholog was detected in the *N. sylvestris* genome. Thus, it is not surprising that *AN2* expression was higher in all pink-flowered accessions than in *N. sylvestris* at all developmental time points examined. *AN2* activates *DFR* and *ANS* expression in *Nicotiana* (Pattanaik et al. 2010); therefore, the absence of the *AN2* gene from *N. sylvestris* is consistent with the lower levels of *DFR* and *ANS* observed in this species. Silencing or inactivation of *AN2* orthologs has been shown to yield a white-flowered phenotype in *N. tabacum* (Pattanaik et al. 2010), *Petunia* (Quattrocchio et al. 1999; Hoballah et al. 2007), and *Antirrhinum* (Schwinn et al. 2006). Taken together, these results suggest that lack of *AN2* expression due to loss of the gene from the genome may be responsible for the white floral phenotype in *N. sylvestris*.

11.4.2 Competition Between FLS and DFR May Underlie the Light-Pink Floral Phenotype

Variation in pink floral phenotypes appears to be driven by complex dynamics across the flavonoid biosynthetic pathway, consistent with differences in the degree of pigmentation rather than presence or absence. In comparisons between light-pink and dark-pink flowers, the latter exhibit increased expression of *DFR*, *ANS*, *AN1*, and/or *AN2*, whereas light-pink flowers have higher levels of *CHS* and/or *FLS*. These contrasting patterns suggest that the flux of the pathway may be directed toward producing anthocyanins in dark-pink flowers, but shunted toward flavonols in light-pink flowers. Light-pink accessions displayed high *FLS*:*DFR* ratios at 60% of anthesis length, whereas dark-pink accessions did not. In pairwise comparisons between light-pink and dark-pink accessions, this high *FLS*:*DFR* ratio in light-pink flowers was attributable to lower expression levels of *DFR*, higher expression levels of *FLS*, or both. Across development, *FLS* expression levels were high in *N. tomentosiformis*, Chulumani, and both synthetic lines at 60% of anthesis, but these accessions display both light- and dark-pink phenotypes, suggesting that high *FLS* expression alone is not sufficient to produce a light-pink floral phenotype. *DFR*

expression levels at 60% of anthesis length in the light-pink accessions Chulumani and QM24 were similar to those observed for *N. sylvestris* throughout development, suggesting that *DFR* activation is developmentally delayed until 85% of anthesis length in these accessions. In contrast, *N. tomentosiformis*, 095-55, and QM25 had higher levels of *DFR* expression than *N. sylvestris* at 60% of anthesis length, indicating that *DFR* activation had already occurred by 60% of anthesis length in these accessions. These results suggest that both high expression levels of *FLS* and low expression levels of *DFR* early in development, most likely due to delayed activation of *DFR*, may be required to generate the light-pink floral phenotype; future studies should test this hypothesis.

FLS and DFR are in competition for the same substrates, the dihydroflavonols dihydrokaempferol, dihydroquercetin, and dihydromyricetin, to produce flavonols and anthocyanins, respectively. This competition, however, involves more than only enzyme affinity for substrates. Both FLS and DFR have been shown to interact with CHS in vivo, and FLS may interfere with the DFR and CHS interaction (Crosby et al. 2011). In addition, a study in grape (*Vitis vinifera*) reported that flavonols can bind to DFR and inhibit its activity (Trabelsi et al. 2008). Several studies have shown that up-regulation or silencing of *FLS* can alter the visible floral color phenotype, despite its regulation of the production of flavonols, which are not visible to the human eye. Overexpression of *FLS* yields white or light-pink floral phenotypes in *Mimulus* (Yuan et al. 2016), *Petunia* (Sheehan et al. 2016), and *N. tabacum* (Luo et al. 2015). Similarly, silencing *FLS* increases anthocyanin content in *Arabidopsis* (Stracke et al. 2009), pink-flowered *Petunia* and *N. tabacum* (Holton et al. 1993), and white-flowered *Petunia*, although, in this study, anthocyanin pigmentation was only observed in floral buds and was not maintained at anthesis (Davies et al. 2003). This evidence supports our conclusion that differences between light- and dark-pink flower phenotypes may be attributable to competition between FLS and DFR, although several questions remain. Does FLS have a higher affinity for the shared substrates than DFR? If so, why does FLS not always outcompete DFR? How high does the *FLS*:*DFR* ratio need to be to tip the balance toward producing flavonols at the expense of anthocyanins? Further studies are required to address these questions.

Acknowledgements We thank James Giovannoni and Yimin Xu for transcriptome sequencing, Zhangjun Fei for help with sequence analysis, Loyal Goff for assistance with cummerbund, and Christopher Fiscus, Dinusha Maheepala, Glen Morrison, and Alex Rajewski for help with R programming.

Data Availability

Raw sequencing reads used in this study have been submitted to the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) with the following accession numbers: *Nicotiana sylvestris* (SRR6434939, SRR6434941–SRR6434948), *N. tabacum* 095-55 (SRR6434933–SRR6434937, SRR6434953, SRR6434966, SRR6434967, SRR6434696, SRR6434698, SRR6434699), *N. tabacum* 51789 (SRR6434694, SRR6434695, SRR6434697), *N. tabacum* "Chulumani" (SRR6434929–SRR6434932, SRR6434960–SRR6434964), synthetic *N. tabacum* QM24 (SRR6434921, SRR6434922, SRR6434949, SRR6434950, SRR6434965, SRR6434968, SRR6434969), synthetic *N. tabacum* QM25 (SRR653419, SRR6434920, SRR6434923–SRR6434928, SRR6434951), and *N. tomentosiformis* (SRR6434940, SRR6434952, SRR6434954–SRR6434959). Analysis code is available on GitHub (https://github.com/jblandis/Nicotiana_transcriptomics).

References

Bai Y, Pattanaik S, Patra B et al (2011) Flavonoid-related basic helix-loop-helix regulators, *NtAn1a* and *NtAn1b*,

of tobacco have originated from two ancestors and are functionally active. Planta 234:363–375
Cao L, Rohart F, Gonzalez I et al (2016) mixOmics: Omics data integration project. R package version 6 (1):1
Chase MW, Knapp S, Cox AV et al (2003) Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). Ann Bot 92:107–127
Chen H, Boutros PC (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinform 12:35
Clarkson JJ, Dodsworth S, Chase MW (2017) Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in *Nicotiana* (Solanaceae). Plant Syst Evol 303:1001–1012
Clarkson JJ, Kelly LJ, Leitch AR et al (2010) Nuclear glutamine synthetase evolution in *Nicotiana*: phylogenetics and the origins of allotetraploid and homoploid (diploid) hybrids. Mol Phylogenet Evol 55:99–112
Clarkson JJ, Knapp S, Garcia VF et al (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. Mol Phylogenet Evol 33:75–90
Crosby KC, Pietraszewska-Bogiel A, Gadella TWJ Jr, Winkel BSJ (2011) Förster resonance energy transfer demonstrates a flavonoid metabolon in living plant cells that displays competitive interactions between enzymes. FEBS Lett 585:2193–2198
Davies KM, Schwinn KE, Deroles SC et al (2003) Enhancing anthocyanin production by altering competition for substrate between flavonol synthase and dihydroflavonol 4-reductase. Euphytica 131:259–268
Des Marais DL, Rausher MD (2010) Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. Evolution 64:2044–2054
Edwards KD, Fernandez-Pozo N, Drake-Stowe K et al (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genom 18:448
Finn RD, Coggill P, Eberhardt RY et al (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–85
Gates DJ, Olson BJSC, Clemente TE, Smith SD (2017) A novel R3 MYB transcriptional repressor associated with the loss of floral pigmentation in *Iochroma*. New Phytol. https://doi.org/10.1111/nph.14830
Goff L, Trapnell C, Kelley D (2013) cummeRbund: analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version
Goodspeed TH (1954) The genus *Nicotiana*. Chronica Botanica, Waltham, Massachusetts, USA
Grabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652
Grotewold E (2006) The genetics and biochemistry of floral pigments. Annu Rev Plant Biol 57:761–780
Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512
Hoballah ME, Gübitz T, Stuurman J et al (2007) Single gene-mediated shift in pollinator attraction in *Petunia*. Plant Cell 19:779–790
Holton T, Brugliera F, Tanaka Y (1993) Cloning and expression of flavonol synthase in *Petunia hybrida*. Plant J 4:1003–1010
Joshi NA, Fass JN (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]
Knapp S, Chase MW, Clarkson JJ (2004) Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae). Taxon 53:73–82
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359
Luo P, Ning G, Wang Z et al (2015) Disequilibrium of flavonol synthase and dihydroflavonol-4-reductase expression associated tightly to white vs. red color flower formation in plants. Front Plant Sci 6:1257
Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17:10–12
McCarthy EW, Arnold SEJ, Chittka L et al (2015) The effect of polyploidy and hybridization on the evolution of floral colour in *Nicotiana* (Solanaceae). Ann Bot 115:1117–1131
McCarthy EW, Berardi AE, Smith SD, Litt A (2017) Related allopolyploids display distinct floral pigment profiles and transgressive pigments. Am J Bot 104:92–101
Pattanaik S, Kong Q, Zaitlin D et al (2010) Isolation and functional characterization of a floral tissue-specific R2R3 MYB regulator from tobacco. Planta 231:1061–1076
Quattrocchio F, Wing J, van der Woude K et al (1999) Molecular analysis of the *anthocyanin2* gene of Petunia and its role in the evolution of flower color. Plant Cell 11:1433–1444
Rasband WS (1997) ImageJ. US National Institutes of Health, Bethesda, MD
Ritchie ME, Phipson B, Wu D et al (2015) *limma* powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47
Schwinn K, Venail J, Shang Y et al (2006) A small family of *MYB*-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. Plant Cell 18:831–851
Sheehan H, Moser M, Klahre U et al (2016) *MYB-FL* controls gain and loss of floral UV absorbance, a key trait affecting pollinator preference and reproductive isolation. Nat Genet 48:159–166
Sierro N, Battey JND, Ouadi S et al (2013) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol 14:R60
Smith SD, Rausher MD (2011) Gene loss and parallel evolution contribute to species difference in flower color. Mol Biol Evol 28:2799–2810

Stracke R, De Vos RCH, Bartelniewoehner L et al (2009) Metabolomic and genetic analyses of flavonol synthesis in *Arabidopsis thaliana* support the in vivo involvement of leucoanthocyanidin dioxygenase. Planta 229:427–445

Streisfeld MA, Rausher MD (2009) Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. New Phytol 183:751–763

The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169

Trabelsi N, Petit P, Manigand C et al (2008) Structural evidence for the inhibition of grape dihydroflavonol 4-reductase by flavonols. Acta Crystallogr D Biol Crystallogr D64:883–891

Trapnell C, Roberts A, Goff L et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578

Warnes GR, Bolker B, Bonebakker L, Gentleman R (2009) gplots: Various R programming tools for plotting data. R package version

Wessinger CA, Rausher MD (2015) Ecological transition predictably associated with gene degeneration. Mol Biol Evol 32:347–354

Wickham H (2009) ggplot2: Elegant Graphics for Data Analysis Springer-Verlag. New York

Yuan Y-W, Rebocho AB, Sagawa JM et al (2016) Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between *Mimulus* species. Proc Natl Acad Sci U S A 113:2448–2453

Yuan Y-W, Sagawa JM, Young RC et al (2013) Genetic dissection of a major anthocyanin QTL contributing to pollinator-mediated reproductive isolation between sister species of *Mimulus*. Genetics 194:255–263

Zhong S, Joung J-G, Zheng Y et al (2011) High-throughput illumina strand-specific RNA sequencing library preparation. Cold Spring Harb Protoc 2011:940–949

Zufall RA, Rausher MD (2004) Genetic changes associated with floral adaptation restrict future evolutionary potential. Nature 428:847–850