The regulatory network for petal anthocyanin pigmentation is shaped by the MYB5a/NEGAN transcription factor in *Mimulus*

Short title: Network dynamics altered by an R2R3 MYB in anthocyanin-pigmented petals

Xingyu Zheng[1], Kuenzang Om[2¤a], Kimmy A. Stanton[2¤b], Daniel Thomas[2], Philip A. Cheng[2], Allison Eggert[2], Emily Simmons[1], Yao-Wu Yuan[3], Gregory D. Conradi Smith[1], Joshua R. Puzey[1*], Arielle M. Cooley[2*]

1 Depts. of Biology and Applied Science, William & Mary, Williamsburg, VA, U.S.A.
2 Dept. of Biology, Whitman College, Walla Walla, WA, U.S.A.
3 Dept. of Ecology & Evolutionary Biology, University of Connecticut, Storrs, CT, U.S.A.

¤a Current Address: School of Biological Sciences, Washington State University, Pullman, WA, U.S.A.
¤b Current Address: Biology Dept, Duke University, Durham, NC, U.S.A..
*Corresponding author for transcriptome analysis: puzeyjr@wm.edu
*Corresponding author for sequence analysis, RT-PCR, and transgenics: cooleya@whitman.edu
ORCiDs: Xingyu Zheng, 0000-0003-4704-1072; Kuenzang Om, 0000-0002-5082-2489; Yao-Wu Yuan, 0000-0003-1376-0028; Gregory D. Conradi Smith, 0000-0002-1054-6790; Emily Simmons, 0000-0002-5630-0112; Joshua Puzey, 0000-0001-8019-9993; Arielle M. Cooley, 0000-0003-3730-0288

Supporting information

S1 Fig. RT-PCR indicates a lack of *MYB4* expression in developing *M. l. variegatus* petal lobes. A. *MYB4* primers 16F-17R and 18F-17R amplify a *MYB4* fragment from *M. l. luteus* and *M. l. variegatus* genomic DNA. Expected lengths are 310 base pairs (bp) and 320 bp. (-), no-template control; l, *M. l. luteus*; v, *M. l. variegatus*. *CHS* is included as a positive control (expected length 650 bp). B. The *MYB4* 16F-17R primers amplify a *MYB4* fragment out of *M. l. variegatus* leaf cDNA, but not out of developing *M. l. variegatus* floral bud cDNA. Expected length from cDNA, for this intron-spanning primer pair, is 190 base pairs. (-), no-template control; pl, petal lobe; ng, nectar guide. C. *Actin* was used as a positive control to verify cDNA quality. Labels are the same as in panel B.

S2 Fig. Offspring of line Vrnai1 show a 3:1 ratio of RNAi phenotype : wild-type, consistent with a single transgene insertion in Vrnai1. Eight seedlings were grown to flowering. Two flowers per plant are shown. Plants 1.1, 1.3, and 1.5 were used for RNA extraction and transcriptome sequencing.

S3 Fig. Maximum likelihood tree of anthocyanin-related R2R3 MYBs. Tree is based on genes listed in Supplemental Table S4.

S4 Fig. Principle components analysis illustrating transcriptome differences between *MYB5a* RNAi and wild-type *M. l. variegatus*. Unsupervised clustering of the 7 transcriptome libraries was performed using principle component analysis based on the differences between samples in normalized read count of all genes. The x and y axes represent the 1st and 2nd principal components.

S5 Fig. Transcript level of Early Biosynthetic Genes (EBGs) of the anthocyanin pathway in *MYB5a* RNAi line Vrnai1 compared to wild-type *M. l. variegatus*. No consistent pattern in expression level is observed comparing the RNAi lines to the wild-type. The bars represent the average expression level, and upper and lower error bars represent maximum and minimum expression level among the samples (n=3). RPKM (per million mapped reads) is used as the normalized unit of transcript expression.

S6 Fig. Transcript level of Late Biosynthetic Genes (LBGs) of the anthocyanin pathway in *MYB5a* RNAi line Vrnai1 compared to wild-type *M. l. variegatus*. All genes identified as LBGs show a decreased expression in the RNAi line compared to the wild type. The bars represent the average expression level, and upper and lower error bars represent maximum and minimum expression level among the samples (n=3). RPKM (per million mapped reads) is used as the normalized unit of transcript expression.

S1 Table. Primers used for qualitative and quantitative RT-PCR and transgene construction. Primers are named by their target gene(s), an identification number, and "F" for forward primers or "R" for reverse primers.

S2 Table. Raw read count of all transcriptome libraries from HTSeq. S1 indicates the outgroup, *M. naiandinus*. S2-S4 are from the RNAi line Vrnai1, and S5-S7 are from wild-type *M. l. variegatus*.

S3 Table. Normalized transcript expression level of all libraries in RPKM. Mlv-wt, wild-type [untransformed] *M. l. variegatus* line RC6. Vrnai1, *MYB5a* RNAi line 1. Mna, *M. naiandinus* line 105.

S4 Table. Species and gene IDs for the sequences used for tree construction. See Figure 6 for the trimmed tree or Supplemental Figure S3 for the full tree.

S5 Table. Significantly differentially expressed genes from DESeq2 results. Result of all 632 differentially expressed genes from DESeq2 analysis, giving base means across samples, log2 fold changes, standard errors, test statistics, p-values and adjusted p-values.

S6 Table. Significance of GO terms in the differentially expressed genes. First we did a classical enrichment analysis by testing the over-representation of GO terms within the group of differentially expressed genes (Fisher's Test) and then performed Kolmogorov-Smirnov test using the both the "classic" and the "elim" method. S5 gives a data frame containing the top GO terms identified by the elim algorithm with p-value cut of at .01.

S7 Table. KEGG Pathway Enrichment Testing of the differentially expressed genes. We accessed KEGG pathway assignments for *Arabidopsis* through the KEGGREST Bioconductor package, and then applied Wilcoxon rank-sum test to each pathway for enrichment testing. Columns features pathway code, pathway name, the p value of being enriched and the number of annotated genes in the pathway.

S8 Table. Forward and reciprocal BLAST results for anthocyanin pathway genes. The column "code" contains the short-hand annotation for the structural genes of the anthocyanin pathway; the column "blastp-besthit" shows the best hit in Arabidopsis; the column "rcp-blastp-besthit" shows the reciprocal results for each best-hit Arabidopsis genes. Genes are highlighted if the reciprocal BLAST identifies the same *M. l. luteus* gene.

S9 Table. Forward and reciprocal BLAST results for regulatory genes. The column "code" contains the short-hand annotation for the potential regulatory genes of the anthocyanin pathway; the column "blastp-besthit" shows the best hit in Arabidopsis; the column "rcp-blastp-besthit" shows the reciprocal results for each best-hit Arabidopsis genes. Genes are highlighted if the reciprocal blast identifies the same *M. l. luteus* gene.

S10 Table. Orthologs across species based on reciprocal BLAST hits. Gene IDs taken from the *Mimulus luteus* transcriptome are designated as Mlu_xxxxx. Gene names are given in parentheses where available.

S1 File. Mathematical modeling of *RTo* gene expression changes in response to *MYB5a* gene expression changes.