New Phytologist **Supporting Information**

Article title: **Single-cell RNA-sequencing of *Nicotiana attenuata* corolla cells reveals the biosynthetic pathway of a floral scent**

Authors: Moonyoung Kang[1], Yuri Choi[1], Hyeonjin Kim[1], and Sang-Gyu Kim[1*]

[1] Department of Biological Sciences, Korea Advanced Institute for Science and Technology, Daejeon 34141, Republic of Korea

[*]Corresponding author: Sang-Gyu Kim (sgkim1@kaist.ac.kr, +82-42-350-2645)

Article acceptance date: 05 January 2022

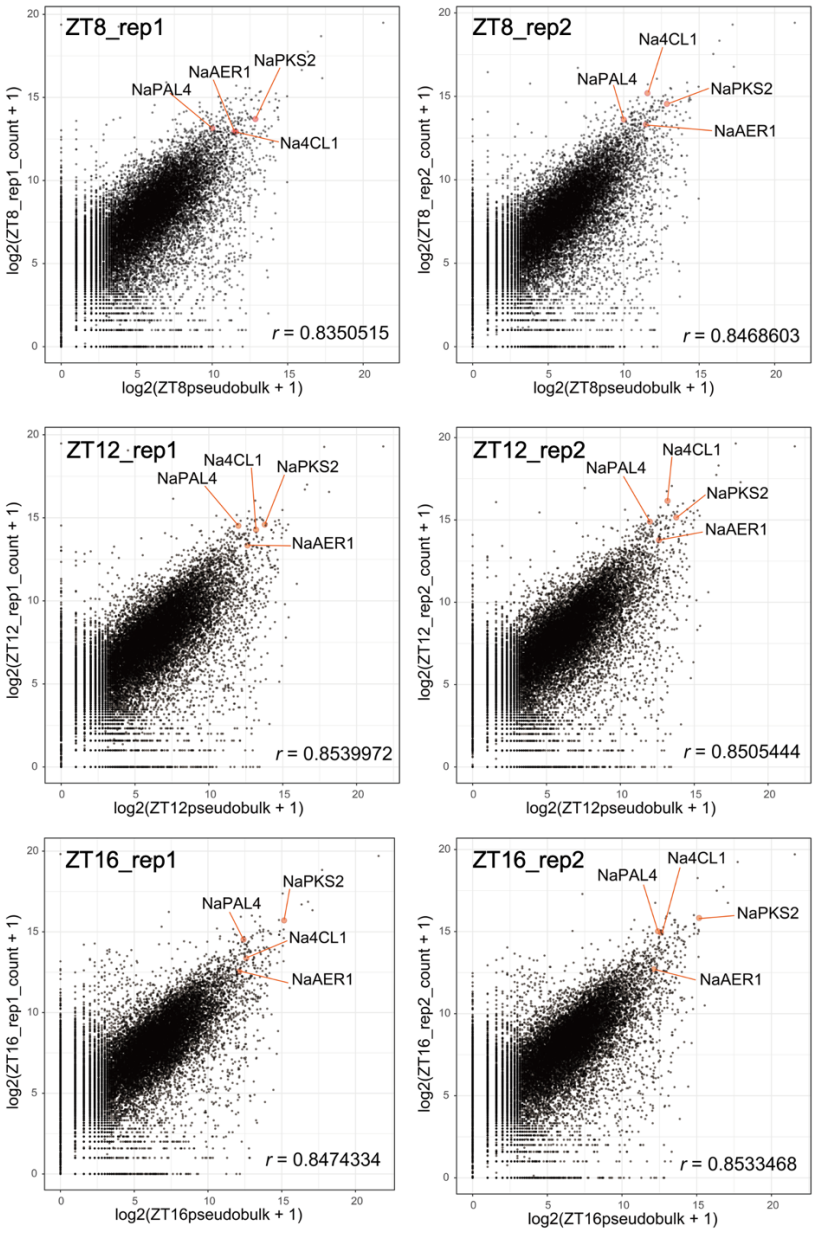


Figure S1. Correlation analysis between bulk-RNA sequencing and scRNA-seq datasets

Scatter plots representing the correlation between bulk RNA sequencing and scRNA-seq datasets in each time point. Pseudo-bulk datasets of scRNA-seq were generated by pooling unique molecular identifier (UMI) across cells, and bulk RNA-seq count numbers were log2-transformed. Pearson's correlation coefficients between two sequencing techniques were calculated and annotated inside figures. The global transcriptome profiles from the two datasets were highly correlated with coefficients over 0.8.

Timepoints	Estimated Number of Cells	Mean Reads per Cell	Median Genes per Cell	Number of Reads	Valid Barcodes	Sequencing Saturation	Q30 Bases in Barcode	Q30 Bases in RNA Read	Q30 Bases in Sample Index	Q30 Bases in UMI	Reads Mapped to Genome	Reads Mapped Confidently to Genome	Reads Mapped Confidently to Intergenic Regions	Reads Mapped Confidently to Intronic Regions	Reads Mapped Confidently to Exonic Regions	Reads Mapped Confidently to Transcriptome	Reads Mapped Antisense to Gene	Fraction Reads in Cells	Total Genes Detected in scRNA Sequencing	Median UMI Counts per Cell	Total Genes Detected in Bulk RNA Sequencing	Gene Discovery % Compared to Bulk RNA Sequencing
ZT 8	1,024	405,136	1,233	414,859,988	95.50%	83.60%	97.20%	93.40%	95.40%	97.40%	25.70%	22.40%	6.60%	0.30%	15.50%	15.00%	0.10%	80.40%	19,553	4,263	22,846	85.59%
ZT 12	1,038	415,875	1,307	431,678,469	95.40%	84.50%	97.50%	93.90%	96.40%	97.60%	33.00%	29.20%	6.10%	0.60%	22.50%	21.90%	0.20%	84.00%	19,842	5,428	22,626	87.70%
ZT 16	1,718	247,401	1,073	425,035,816	96.20%	90.00%	97.70%	94.00%	96.90%	97.70%	45.50%	40.10%	7.10%	0.70%	32.30%	31.40%	0.20%	79.80%	18,917	3,246	22,803	82.96%

Figure S2. QC metrics of scRNA-seq datasets

Mean reads per cell are over 200,000 for each dataset, and median genes per cell exceed one thousand. In the single cell population, we detected 82 ~ 87 % of genes detected in bulk RNA sequencing. Mapping ratio is relatively low due to complexity and incomplete gene annotation of *N. attenuata* draft genome (assembly NIATTr2).

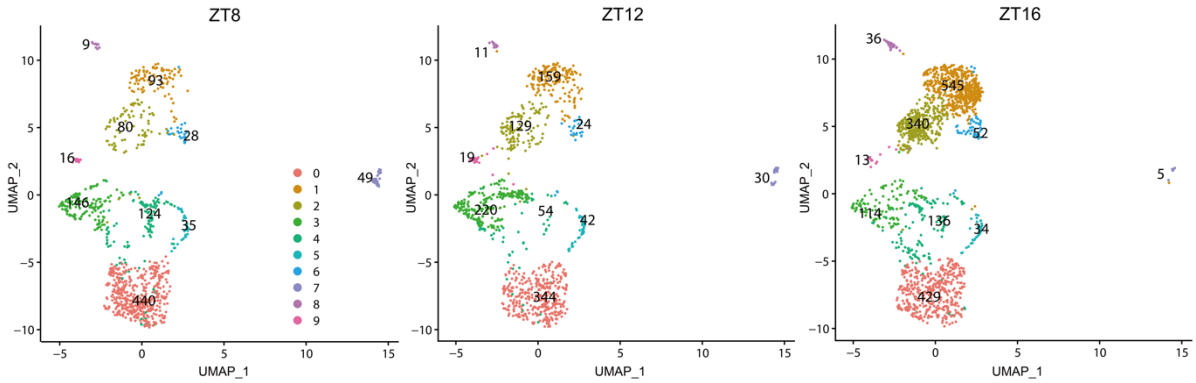


Figure S3. Cells are displayed by an integrated uniform manifold approximation and projection (UMAP) plot in two dimensions

UMAP plots showing dimensional reduction of the distribution of 1020, 1032, and 1704 cells collected at ZT 8, ZT 12, and ZT 16, respectively. UMAP projection yielded overlapping distributions of cells from each scRNA-seq time point. The exact numbers of cells are represented inside each cluster.

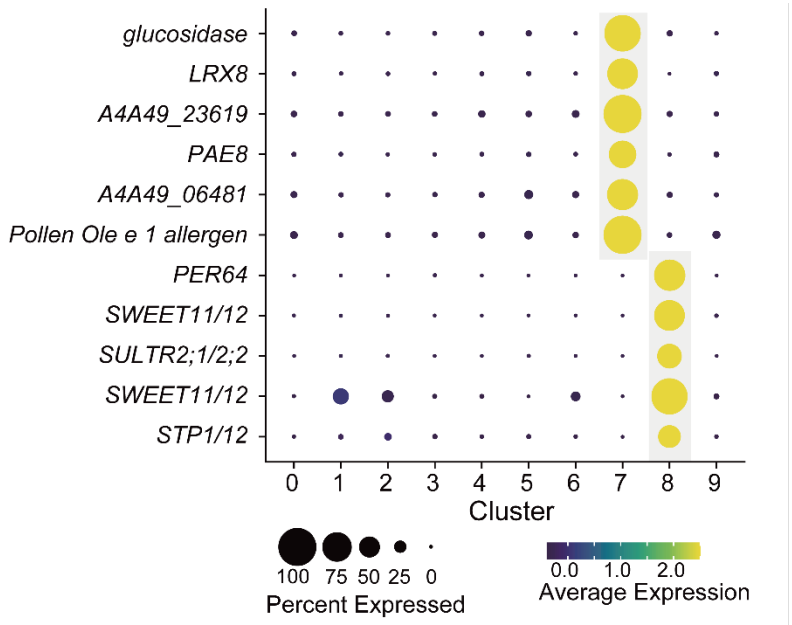


Figure S4. Pollen and vasculature-specific marker genes are exclusively expressed in clusters 7 and 8, respectively

Pollen marker genes and phloem/xylem marker genes are exclusively detected in clusters 7 and 8, respectively. *SULTR2;1/2;2*, a sulfate transporter, and *SWEET11/12* and a *sugar transporter1/12* (*STP1/12*) are known to be exclusively expressed in *A. thaliana* vasculatures (Chen *et al.*, 2012; Maruyama-Nakashita *et al.*, 2015; Kim *et al.*, 2021). All marker genes in cluster 7 are specifically expressed in pollen or pollen tubes as shown in figure S5 (Brockmöller *et al.*, 2017). Dot size and color indicate the percentage of cells expressing a marker gene and mean expression value across cells in each cluster, respectively.

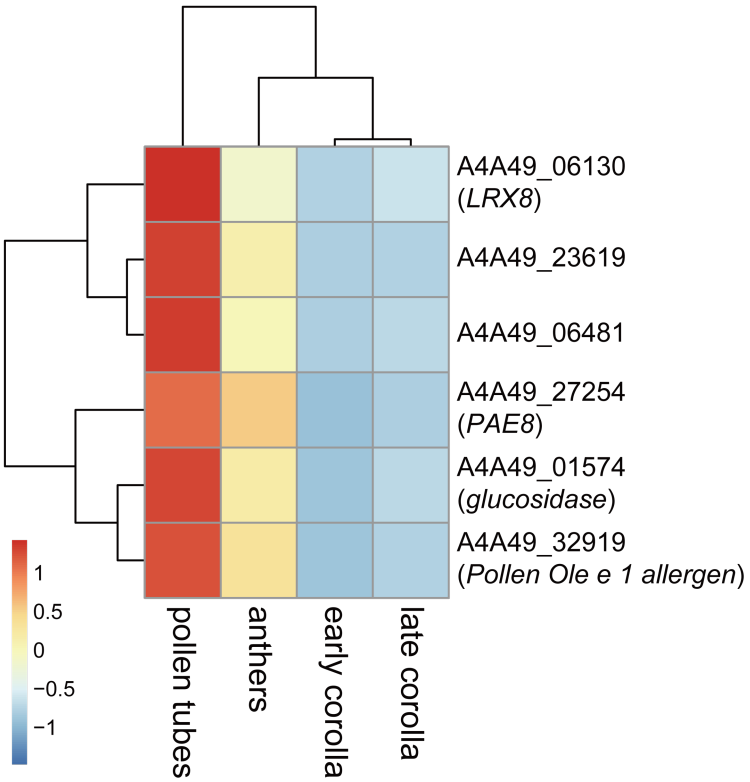


Figure S5. Validation of pollen-specific marker genes

Pollen specific marker genes are exclusively expressed in anthers and pollens in public bulk RNA-seq datasets (Li *et al*., 2016). TPM values were log2-transformed and scaled for heatmap.

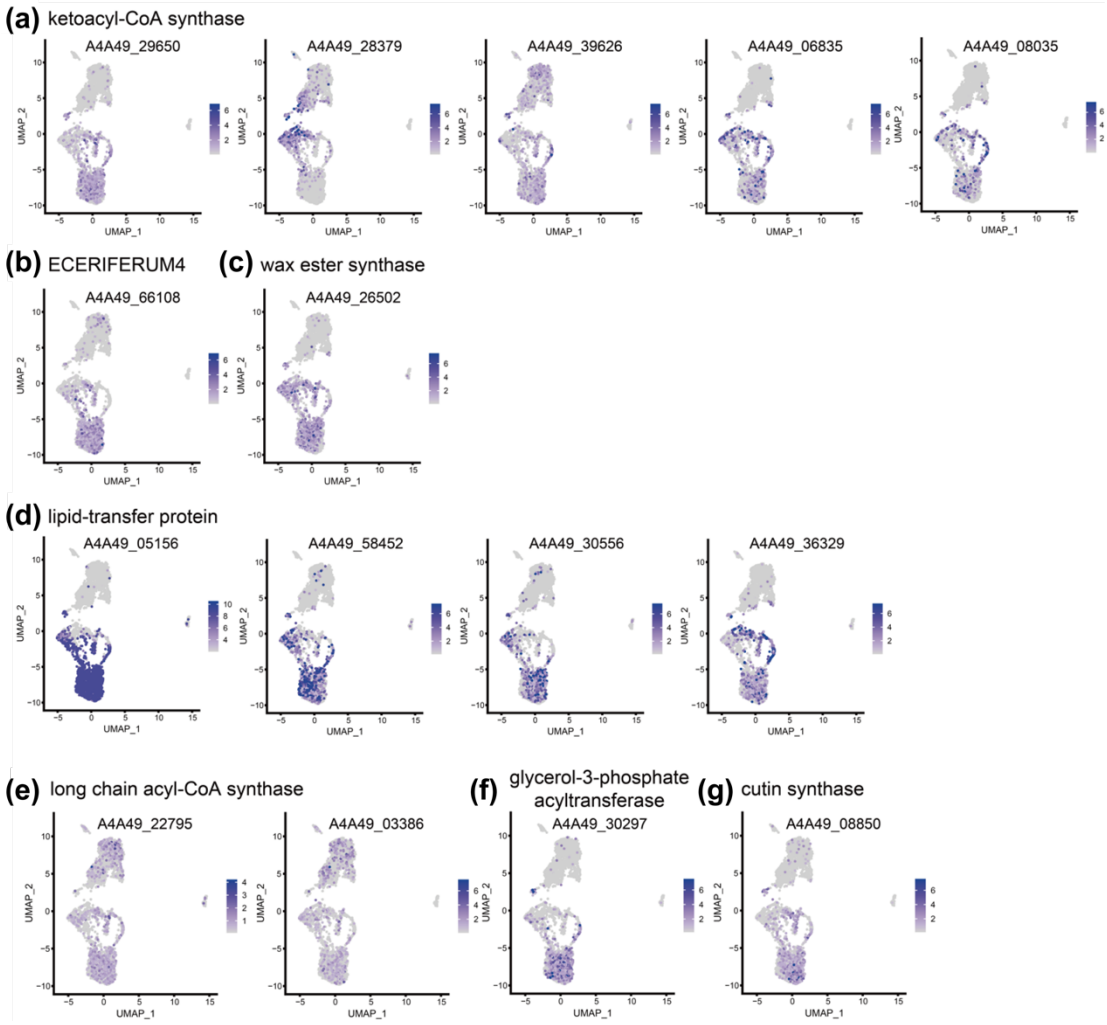


Figure S6. Putative wax and cutin biosynthesis genes are more expressed in clusters 0, 3, 4, 5, and 9

(a) Ketoacyl-CoA synthase (A4A49_29650, 28379, 39626, 06835, 08035), **(b)** ECERIFERUM4 (A4A49_66108), **(c)** wax ester synthase/diacylglycerol acyltransferase 1 (A4A49_26502), and **(d)** lipid-transfer proteins (A4A49_05156, 58452, 30556, 36329), which are likely to participate in wax transport. **(e)** Long chain acyl-CoA synthetase 4 (A4A49_22795, 03386), **(f)** Glycerol-3-phosphate acyltransferase 6 (A4A49_30297), **(g)** Cutin synthase-like (A4A49_08850), which are likely to participate in cutin biosynthesis. These results suggest that the clusters 0, 3, 4, 5, and 9 are epidermal cells. Each dot indicates single cells, and transcript levels of genes are displayed in color scale.

identity	short name	locus tag	*A. thaliana*	% Similarity (Blosum62)	E-values of blastp
vasculature	*PER64*	A4A49_03359	AT4G33420 (Per47), AT5G42180 (Per64)	84.29%, 74.32%	6e-163, 1e-92
	SWEET11/12	A4A49_06893	AT3G48740 (SWEET11), AT5G23660 (SWEET12)	75.768%, 74.49%	4e-92, 3e-91
	SULTR2;1/2;2	A4A49_32258	AT5G10180 (SULTR2;1), AT1G77990 (SULTR2;2)	89.788%, 88.123%	0.0, 0.0
	SWEET11/12	A4A49_14806	AT3G48740 (SWEET11), AT5G23660 (SWEET12)	73.597%, 74.503%	2e-98, 3e-97
	STP1/12	A4A49_31174	AT1G11260 (STP1), AT4G21480 (STP12)	90.244%, 91.865%	0.0, 0.0
Epidermis	*NaKCS10*	A4A49_29650	AT2G26250 (KCS10)	89.331%	0.0
	NaKCS11-like	A4A49_28379	AT1G04220 (DAISY/KCS2)	90.038%	0.0
	NaKCS19-like	A4A49_39626	AT5G43760 (KCS20)	81.992%	0.0
	NaKCS6	A4A49_06835	AT1G68530 (KCS6)	96.378%	0.0
	NaKCS11-like	A4A49_08035	AT5G43760 (KCS20)	90.737%	0.0
	NaFAR3-like	A4A49_66108	AT4G33790 (CER4/FAR3)	78.442%	5e-125
	NaWSD1-like	A4A49_26502	AT5G53390 (WSD11)	75.143%	2e-160
	NaNLTP2_2	A4A49_05156	AT2G38530 (LTP2)	72.034%	1e-23
	NaNLTP2_4	A4A49_58452	AT2G38530 (LTP2)	74.576%	2e-20
	NaNLTP2_5	A4A49_30556	AT2G38530 (LTP2)	73.729%	5e-20
	NaLTP1-like	A4A49_36329	AT5G59310 (LTP4)	82.051%	2e-33
	NaLACS4-like	A4A49_22795	AT1G64400 (LACS3)	90.881%	0.0
	NaGPAT6	A4A49_30297	AT2G38110 (GPAT6)	92.843%	0.0
	AT2G04570-like	A4A49_08850	AT2G04570 (OSP1)	83.427%	3e-158
parenchyma	*photosystem subunit H-1*	A4A49_26930	AT3G16140 (photosystem I subunit H-1)	93.103%	4e-55
	photosystem I LHC gene 3	A4A49_09113	AT1G61520 (photosystem I LHC gene 3)	97.81%	4e-150
	photosystem II subunit S	A4A49_26955	AT1G44575 (photosystem II subunit S)	86.861%	5e-82
	photosystem I subunit K	A4A49_02826	AT1G30380 (photosystem I subunit K)	96.923%	6e-66
	photosystem II reaction center W	A4A49_05181	AT2G30570 (photosystem II reaction center W)	85.185%	9e-41
	pigment defective 334-ATPase	A4A49_01739	AT4G32260 (pigment defective 334)	85.845%	7e-66

Figure S7. Protein sequences of marker genes are similar to their *A. thaliana* orthologs

Orthologs in *Arabidopsis thaliana* were searched by BLASTp at The Arabidopsis Information Resource site (https://www.arabidopsis.org/Blast/), based on Araport11 protein sequences datasets. Marker genes are similar to their orthologs over 72 % similarity. Expect values (E-values) from BLASTP results were near 0.0, suggesting conserved functions. For example, *AtGPAT6*, a gene involved in nanoridge formation on the surface of flowers, is homologous to *NaGPAT6*. (Li-Beisson *et al.*, 2009). A4A49_08850 is homologous to OSP1, a GDSL lipase playing roles in wax biosynthesis (Tang *et al.*, 2020). PER, peroxidase; SWEET, sugars will eventually be exported transporters; KCS, ketoacyl-CoA synthase; FAR3, fatty acid reductase 3 (same as ECERIFERUM 4); NLTP, nonspecific lipid transfer protein; LTP, lipid transfer protein; LACS, long chain acyl-CoA synthetase 3; WSD1, wax ester synthase/diacylglycerol acyltransferase 1; GPAT6, glycerol-3-phosphate sn-2-acyltransferase 6; OSP1, occluded stomatal pores 1.

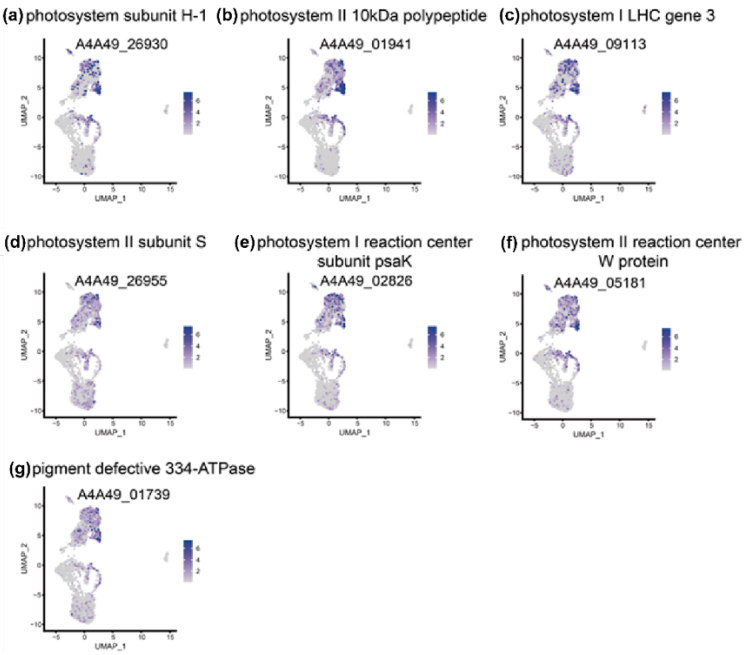


Figure S8. Subunits of photosystem I and II complex are enriched in clusters 1, 2, and 6

(a) Photosystem subunit H-1 (A4A49_26930), **(b)** Photosystem II 10kDa polypeptide (A4A49_01941), **(c)** Photosystem I LHC gene 3 (A4A49_09113) **(d)** Photosystem II subunit S (A4A49_26955), **(e)** Photosystem I reaction center subunit psaK (A4A49_02826), **(f)** Photosystem II reaction center W protein (A4A49_05181), **(g)** Pigment defective 334-ATPase (A4A49_01739): all are likely to be subunits of photosystem complex. Each dot indicates single cells, and the expression levels of genes are displayed on color scale. Cluster 6 is defined as a chlorenchyma cell in the throat of corolla, which retained chloroplasts, and clusters 1 and 2 are defined as parenchymal cells.

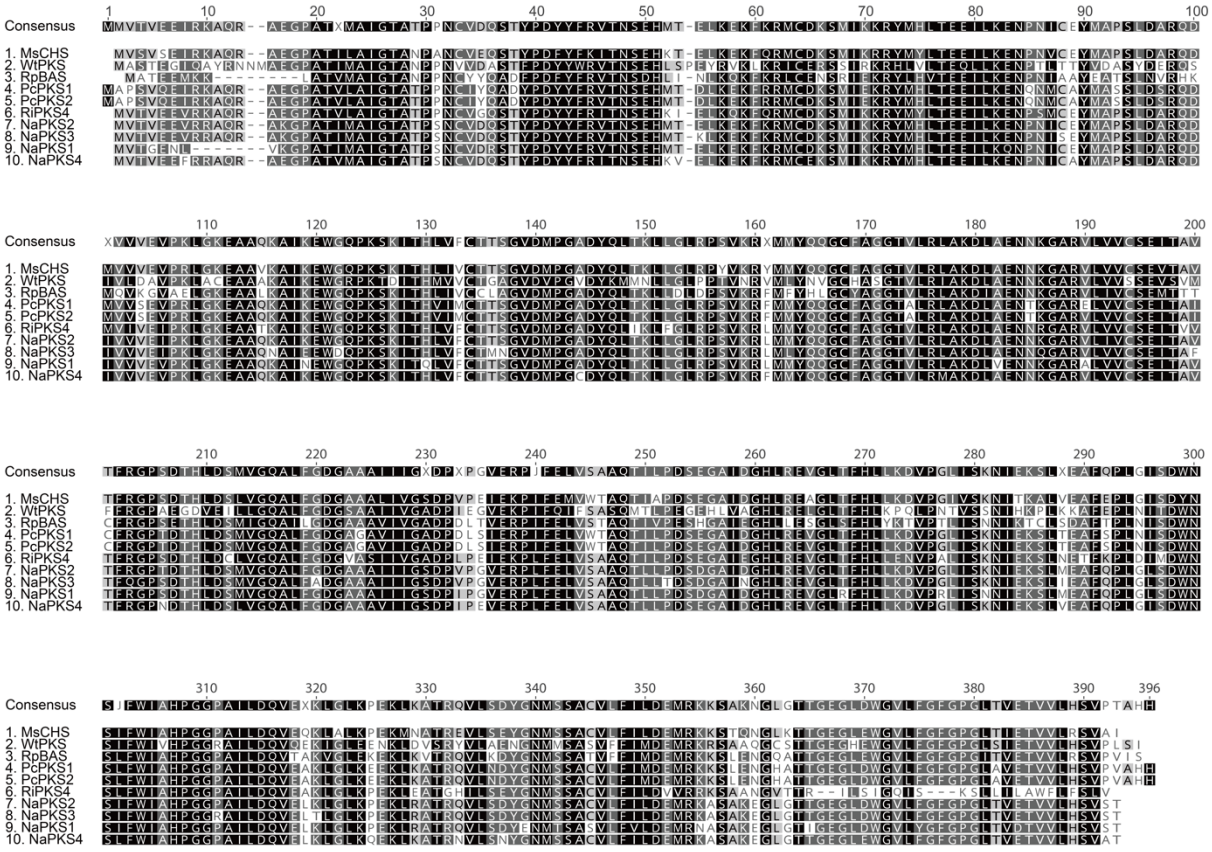


Figure S9. NaPKS1/2/3/4 share high sequence similarity with polyketide synthase proteins of other species

Protein sequence alignments of polyketide synthase (PKS)-related of proteins in several plant species. Full-length amino acid sequences were aligned using Clustal W implemented in the Geneious software. MsCHS, *Medicago sativa* CHS, L02902; WtPKS, *Wachendorfia thyrsiflora* polyketide synthase, AY727928; RpBAS, *Rheum palmatum* benzalacetone synthase, AAK82824; PcPKS1, *Polygonum cupidatum* PKS1, ABK92281; PcPKS2, *Polygonum cupidatum* PKS12; NaPKS1, *Nicotiana attenuata* PKS1 (renamed from chalcone synthase 1), A4A49_08280; NaPKS2, *N. attenuata* PKS2 (renamed from chalcone synthase 2), A4A49_34074; NaPKS3, *N. attenuata* PKS3 (renamed from chalcones synthase 3), A4A49_39367; NaPKS4, *N. attenuata* PKS4 (renamed from chalcone synthase 4), A4A49_65921.

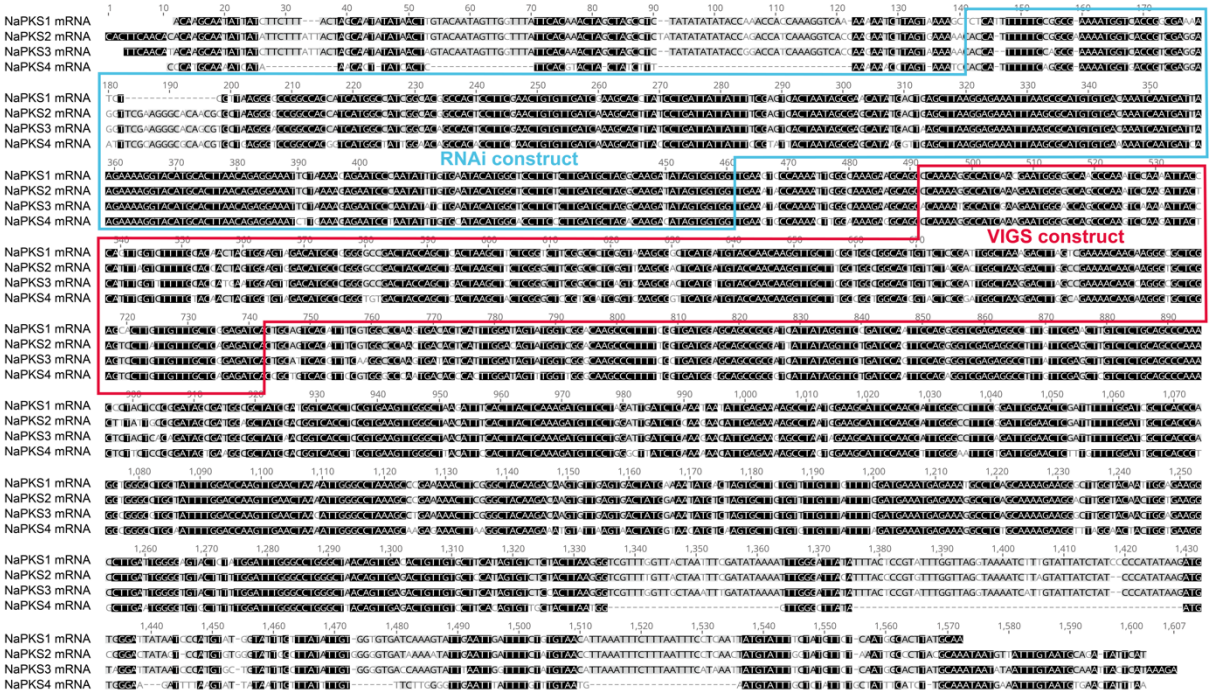


Figure S10. Target sequences for NaPKS RNAi experiments (used in previous studies, NaCHAL) and VIGS experiments (used in this study)

Because *NaPKS1/2/3/4* are highly similar to each other, it is hard to design the specific construct of RNA interference (RNAi) and VIGS for one gene. Only 10 nucleotides are different between *NaPKS3* (A4A49_39367) and *NaPKS2* (A4A49_34074) in the RNAi construct target sequence (total 317 base pairs) used previously in Kessler *et al*. (2008). The difference between *NaPKS3* and *NaPKS1* (A4A49_08280) is 34 nucleotides, and the difference between *NaPKS3* and *NaPKS4* (A4A49_65921) is 39 nucleotides in the RNAi construct. Our VIGS sequence (250 bp) also contained small differences between those four genes; *NaPKS3* sequences in the target region differ from *NaPKS1*, *NaPKS2*, and *NaPKS4* by 29, 26, and 39 nucleotides, respectively.

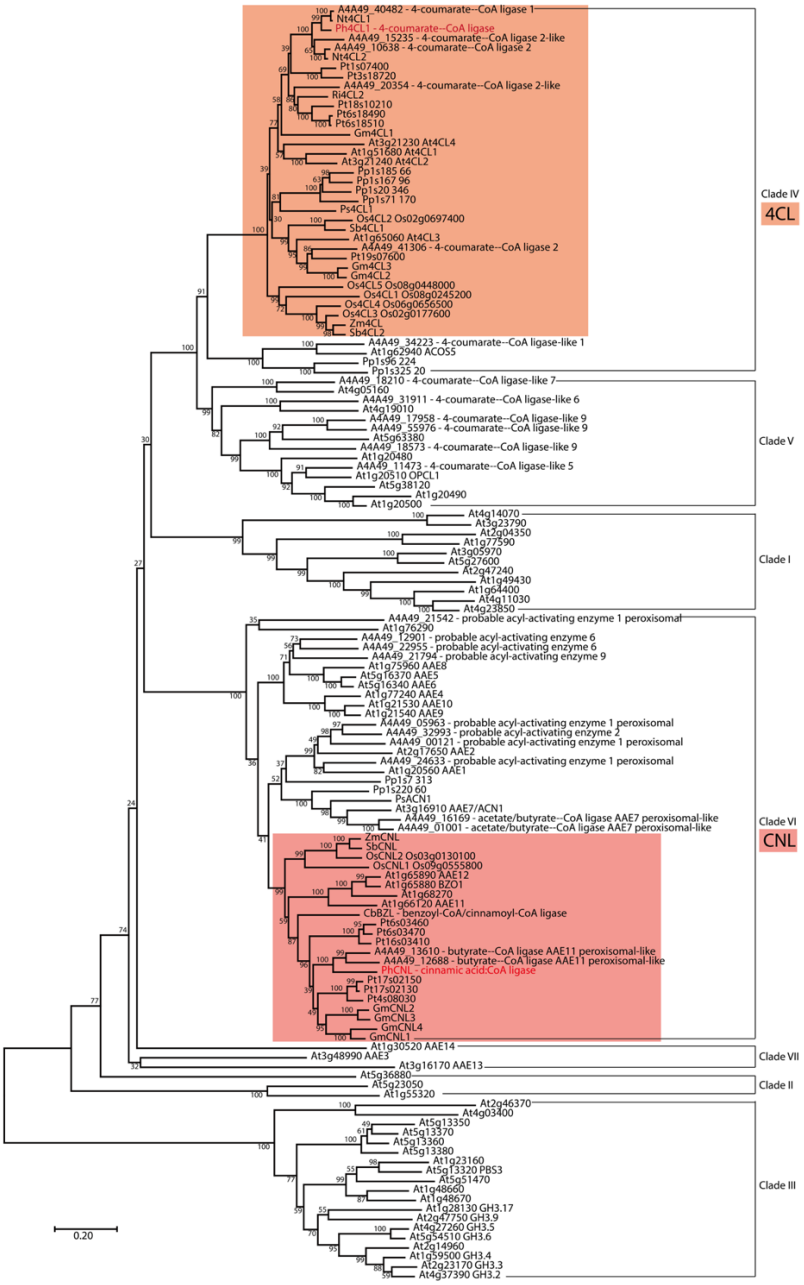


Figure S11. Cladogram of *N. attenuata* 4CL, CNL homologs within acyl-activating enzyme (AAE) superfamily

Protein sequences of Na4CL and NaCNL homologs (found by BLASTp) are aligned with AAE superfamily and *Petunia hybrida* homologs, by neighbor-joining method (Saitou & Nei, 1987). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site (ZUCKERKANDL & PAULING, 1965). Evolutionary analyses were conducted in MEGA X (Kumar *et al.*, 2018).

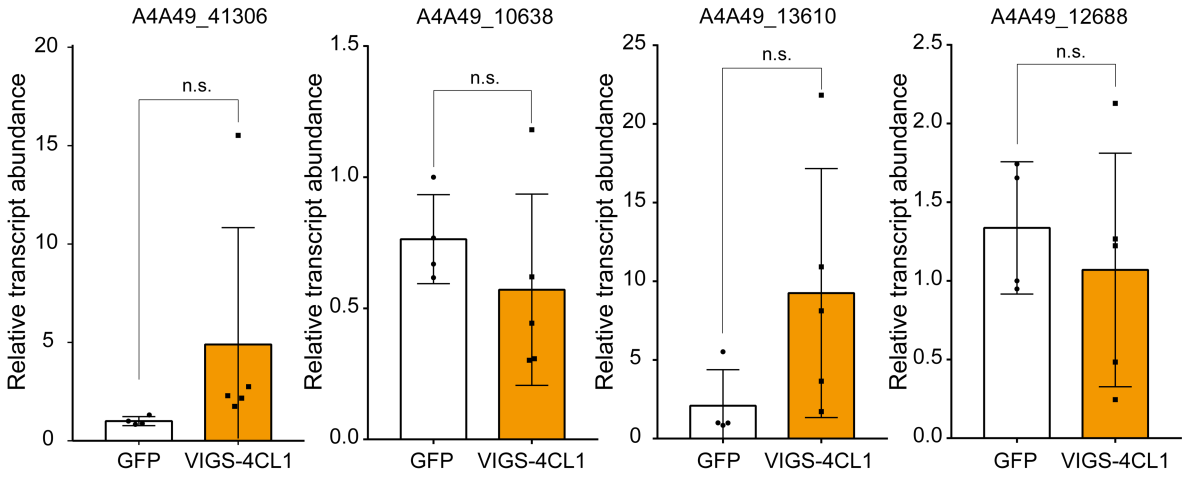


Figure S12. 4CL and CNL homologs are not co-silenced in *Na4CL1*-silenced flowers

Transcript accumulation of two *Na4CL1* homologues (A4A49_41306 and A4A49_10638) and two *NaCNL* homologues (A4A49_13610 and A4A49_12688) in *Na4CL1*-silenced flowers. *Na4CL1* was silenced without co-silencing of other homologs in corolla limbs. Mean (± SE) levels of transcript accumulation (two-tailed Students' t-test; n.s., not significant, $p \geq 0.05$).

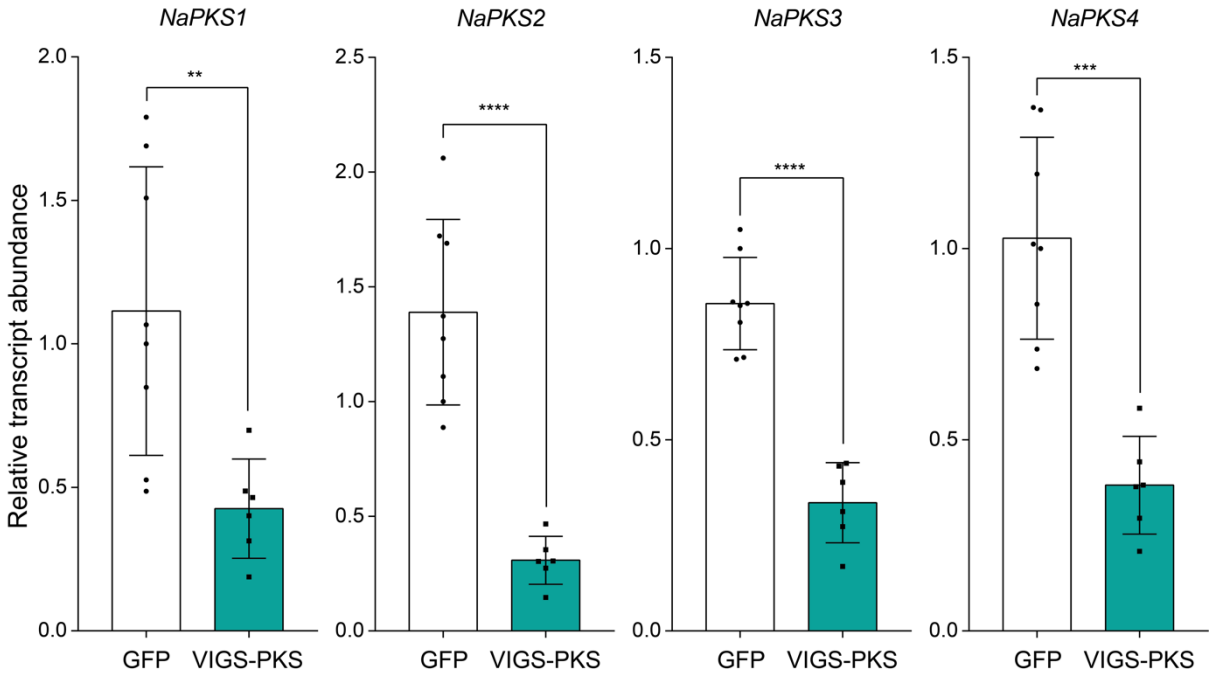


Figure S13. Co-silencing of *NaPKS1/2/3/4* in *NaPKS3*-VIGS flowers

After *NaPKS3* was silenced by VIGS, other *PKSs* were also significantly co-silenced in corolla limbs. Values represent individual flowers (biological replicates). (mean ± SE , two-tailed Students' t-test; **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$) NaPKS1, A4A49_08280; NaPKS2, A4A49_34074; NaPKS3, A4A49_39367; NaPKS4, A4A49_65921.

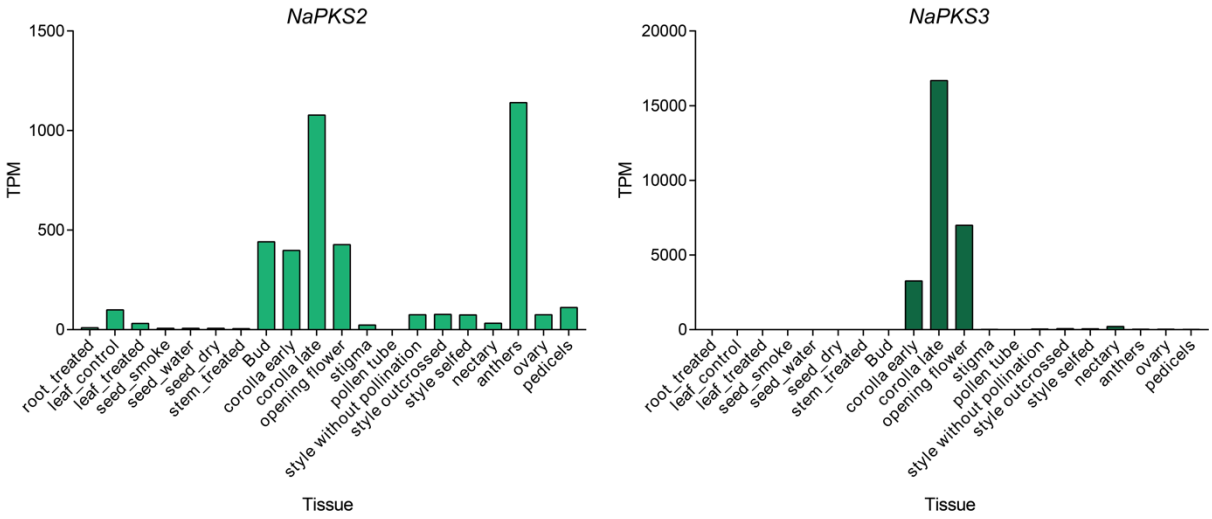


Figure S14. Tissue-specific expression of *NaPKS2* and *NaPKS3* in *N. attenuata*

TPM values from tissue-specific bulk RNA-seq datasets show that both *NaPKS2* and *NaPKS3* are highly expressed in flower tissues, especially in the corolla. *NaPKS2* is also highly accumulated in anthers. (http://nadh.ice.mpg.de/NaDH/chart/expression_genes)

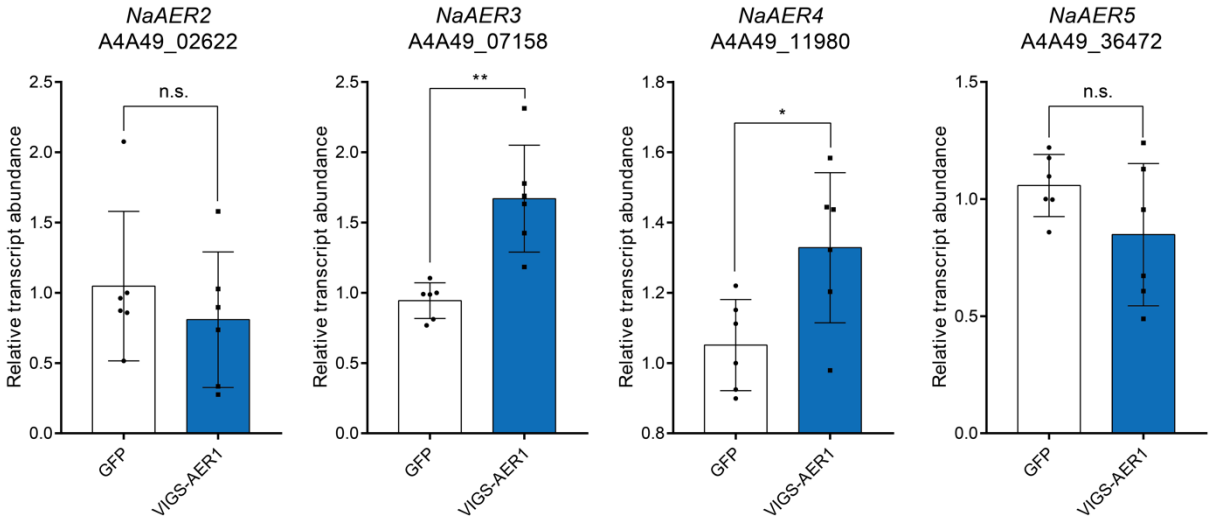


Figure S15. AER homologs are not co-silenced in *NaAER1*-silenced flowers.

NaAER1 was silenced without co-silencing of other 2-alkenal reductases in corolla limbs. Mean (± SE) levels of transcript accumulation (two-tailed Students' t-test; n.s., not significant, *, $p<0.05$; **, $p<0.01$).

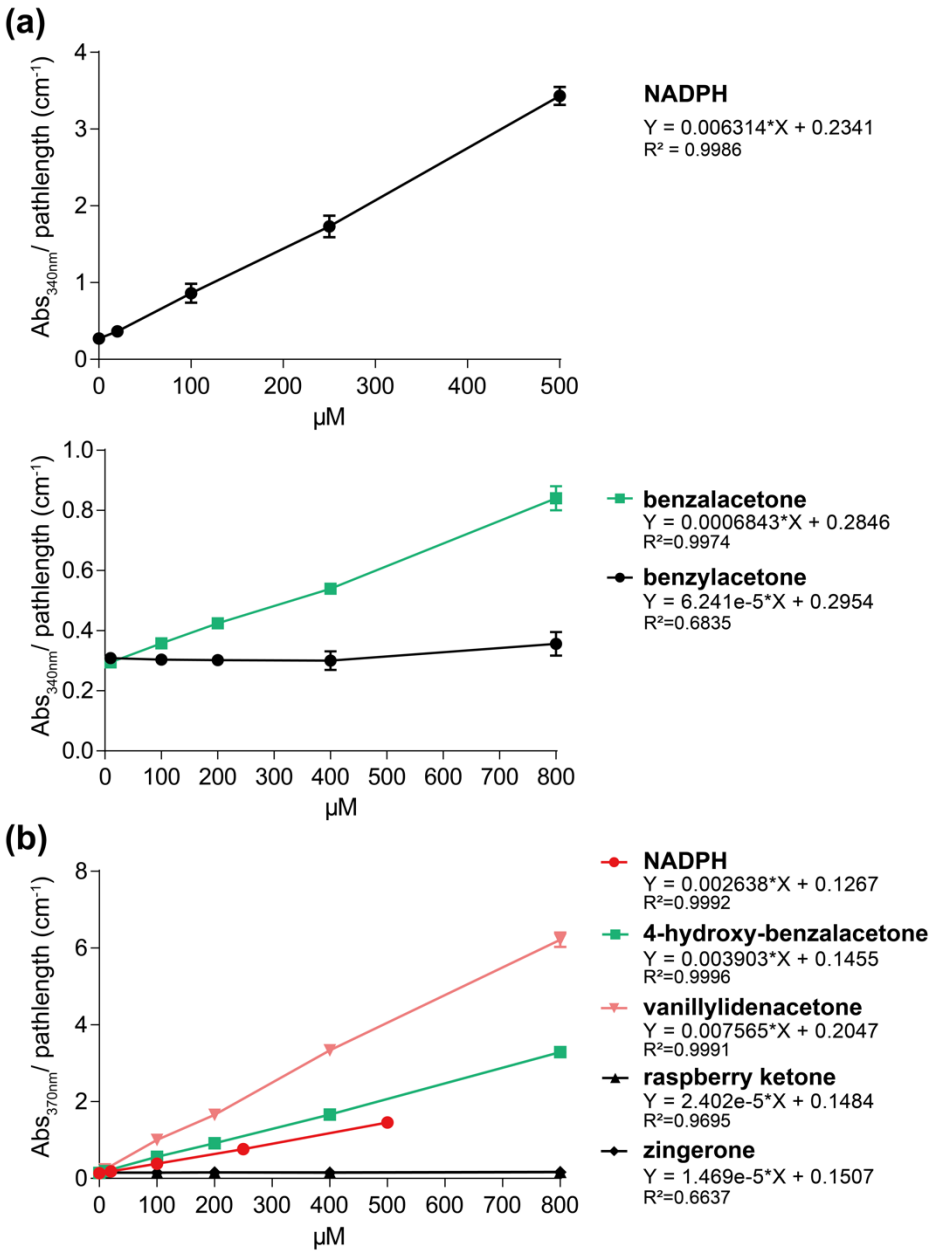


Figure S16. Extinction coefficients used for NaAER1 kinetics

Extinction coefficients of putative substrates and products of NaAER1 enzyme reaction were determined experimentally (mean ± SD). **(a)** Extinction coefficient of benzalacetone at 340 nm was about 0.1-fold the coefficient of NADPH, which has an absorption maximum at 340 nm. Benzylacetone, which is a putative reduced product of benzalacetone, shows low absorption at 340 nm. **(b)** Coefficients of benzalacetone derivatives, 4-hydroxy-benzalacetone and vanillylidenacetone, were determined at 370 nm. NADPH has a lower extinction coefficient, at 370 nm than at 340 nm. Raspberry ketone and zingerone, reduced forms of 4-hydroxy-benzalacetone and vanillylidenacetone respectively, absorb less light at 370 nm.

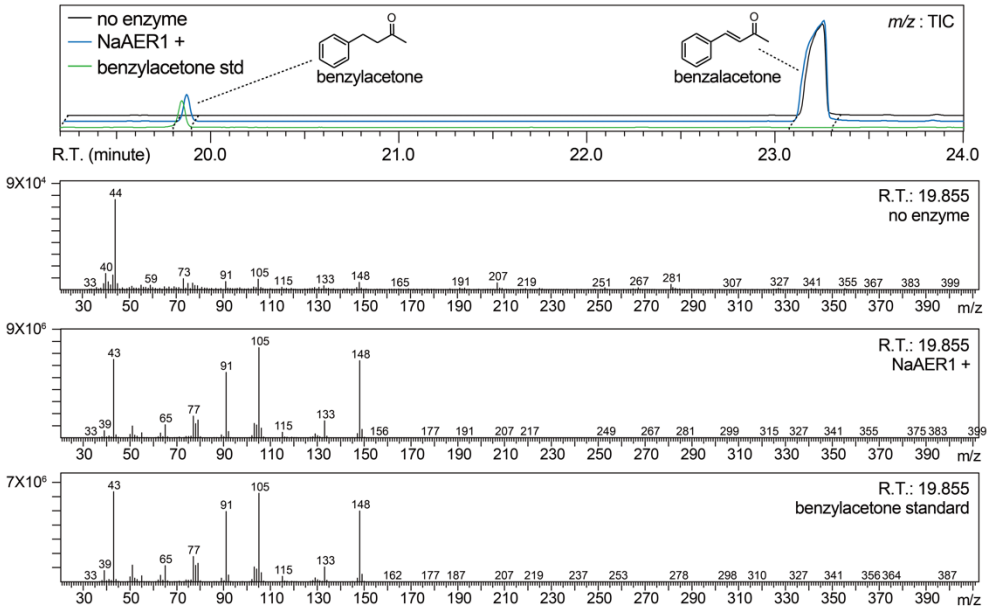


Figure S17. NaAER1 can reduce benzalacetone to benzylacetone

Benzalacetone can be reduced to benzylacetone, when incubated with NADPH and NaAER1 protein. After NaAER1-benzalacetone incubation, a new peak was eluted at retention time (RT) 19.855 min, which corresponds to RT of benzylacetone standard. The mass spectrometry fragmentation pattern of RT 19.855 min is also the same as that for standard.

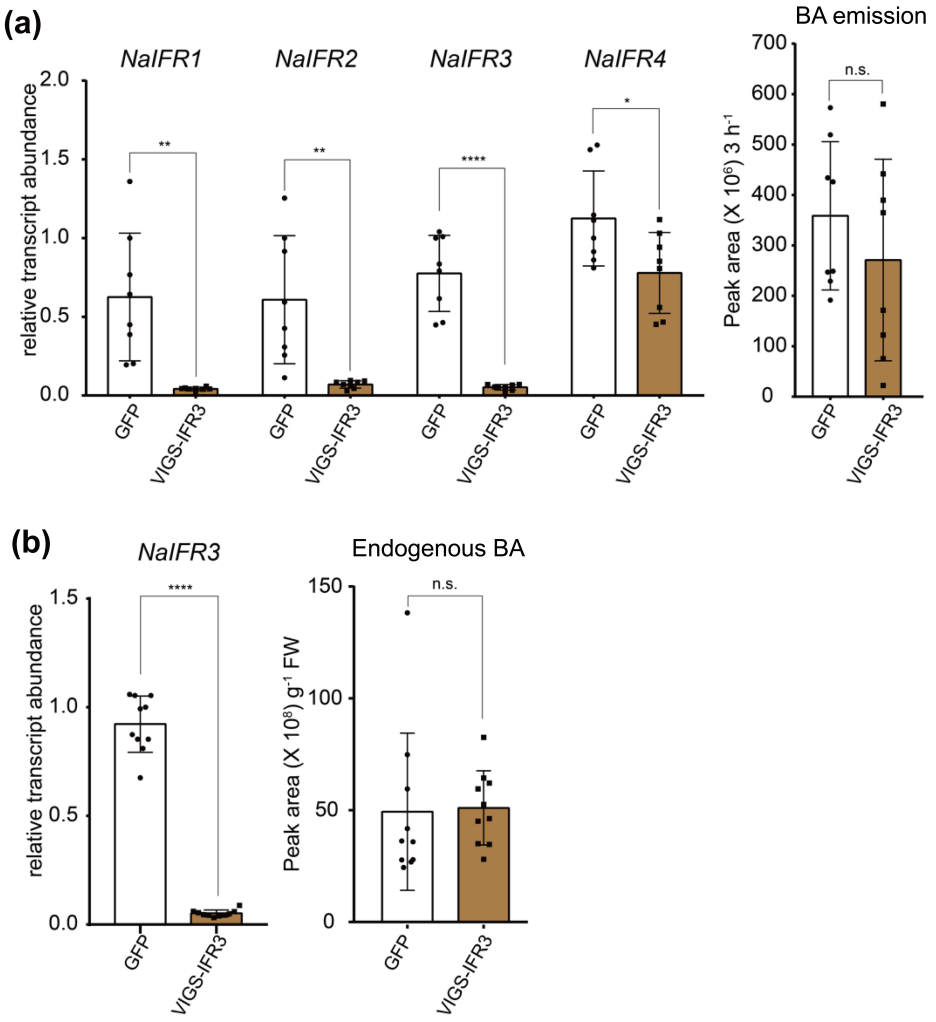


Figure S18. Silencing *NaIFR* transcription does not alter BA biosynthesis in corolla

Isoflavone reductase 3 (NaIFR3, A4A49_30373) was previously identified as a reductase involved in BA biosynthesis. We silenced the expression of *NaIFR3* and its homologs to verify whether they redundantly function for BA biosynthesis. **(a)** Virus-induced gene silencing was used to reduce the level of *NaIFR1*/2/3/4 in the corolla. Silencing efficiency of individual replicates was confirmed by qRT-PCR, with a 15.1-fold decrease in *NaIFR1,* an 8.7-fold decrease in *NaIFR2*, a 14.8-fold decrease in *NaIFR3*, and a 1.44-fold decrease in *NaIFR4* (mean ± SE, two-tailed Students' t-test; *, p<0.05, **, p<0.005, ****, p<0.0001). BA emission was slightly reduced in *NaIFRs*-silenced flowers (mean ± SE, two-tailed Students' t-test, p=0.3348). **(b)** To further confirm the function of NaIFRs for BA biosynthesis, we measured endogenous BA levels in *NaIFRs*-silenced flowers (mean ± SE, two-tailed Students' t-test; n.s, not significant, ****, p<0.0001). *NaIFRs*-silenced flowers retained the same levels of BA as control flowers (mean ± SE, two-tailed Students' t-test, p=0.8912). NaIFR1, A4A49_26421; NaIFR2, A4A49_30375; NaIFR3, A4A49_30373; NaIFR4, A4A49_18736.

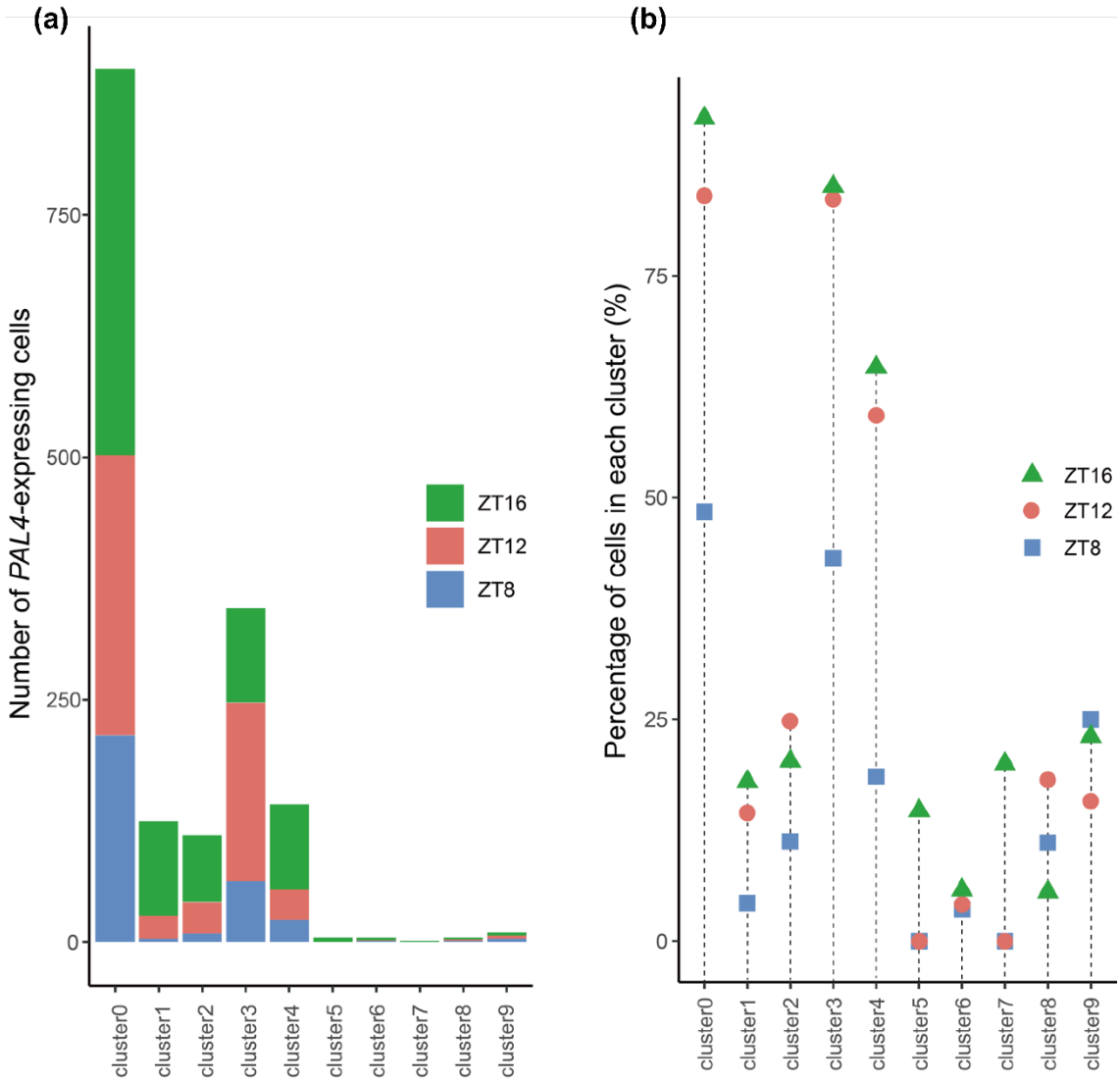
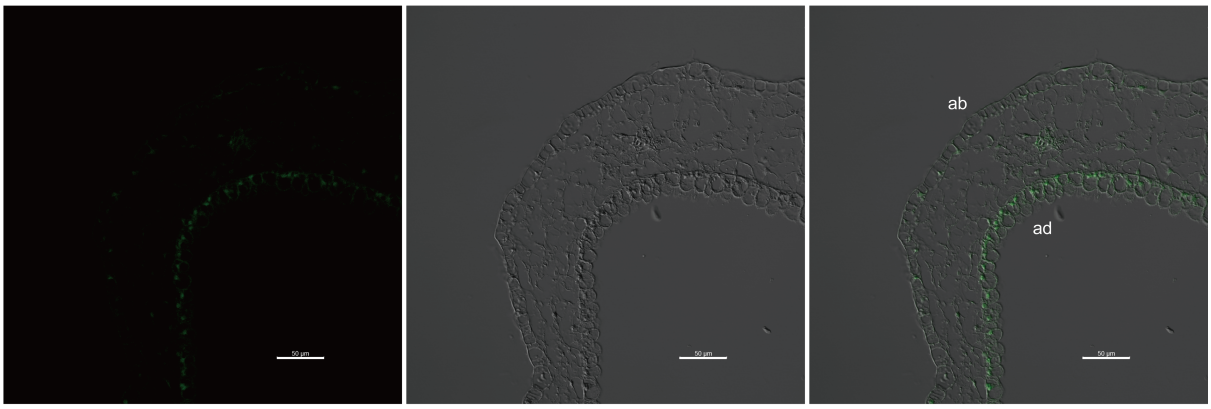


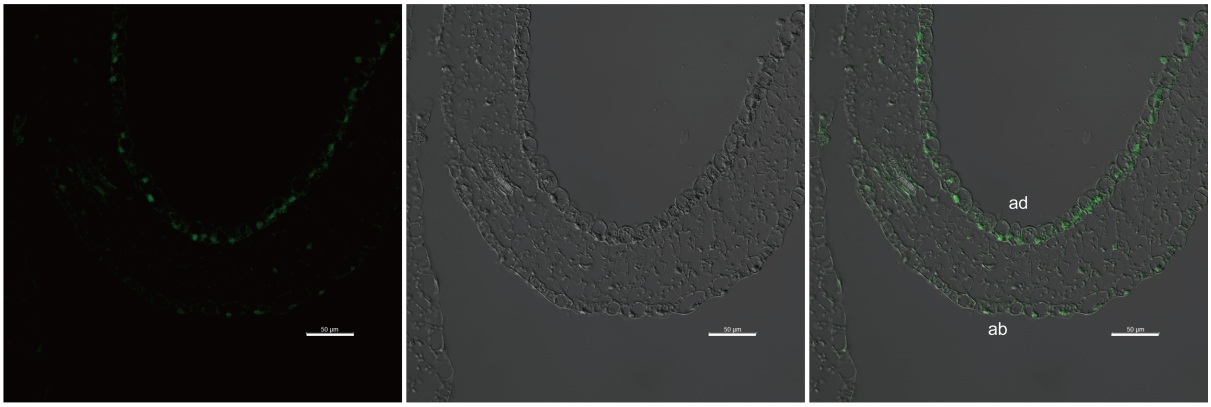
Figure S19. *NaPAL4* transcripts are highly expressed in clusters 0, 3, and 4

(a) Numbers of cells expressing *NaPAL4* transcripts across all clusters at three time points. **(b)** Percentage of cells expressing *NaPAL4* across all clusters at three time points.

(a) *pNaPAL4::eGFP*



(b) *pNaPKS2::eGFP*



(c) WT

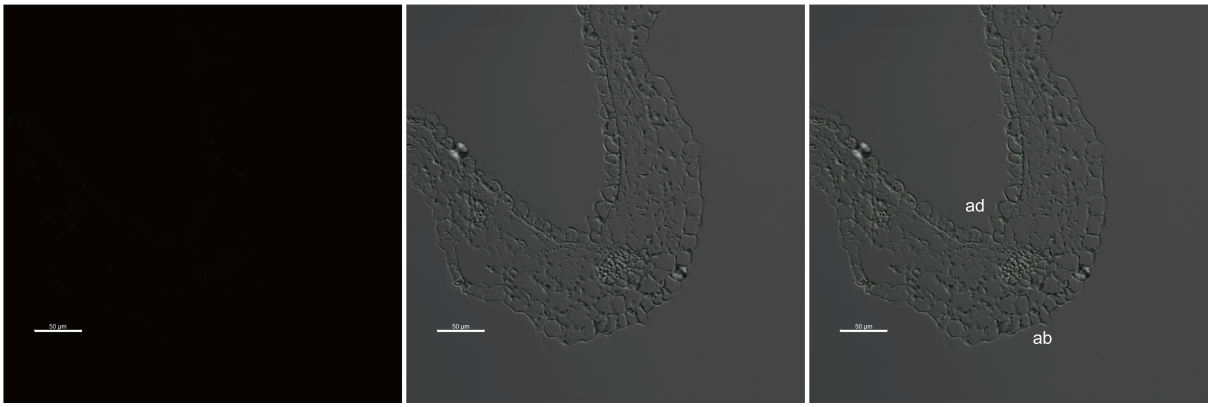


Figure S20. Expression of *NaPAL4* and *NaPKS2* in *N. attenuata* flowers

Confocal laser-scanning microscopy images (scale bar = 50 μm) of *pNaPAL4::eGFP* and *pNaPKS2::eGFP* in *N. attenuata* flowers collected at ZT 8. **(a, b)** Fluorescence signals driven by two promoters are localized in the epidermal cells, and are stronger in the adaxial epidermis than in the abaxial epidermis. ad, adaxial; ab, abaxial.

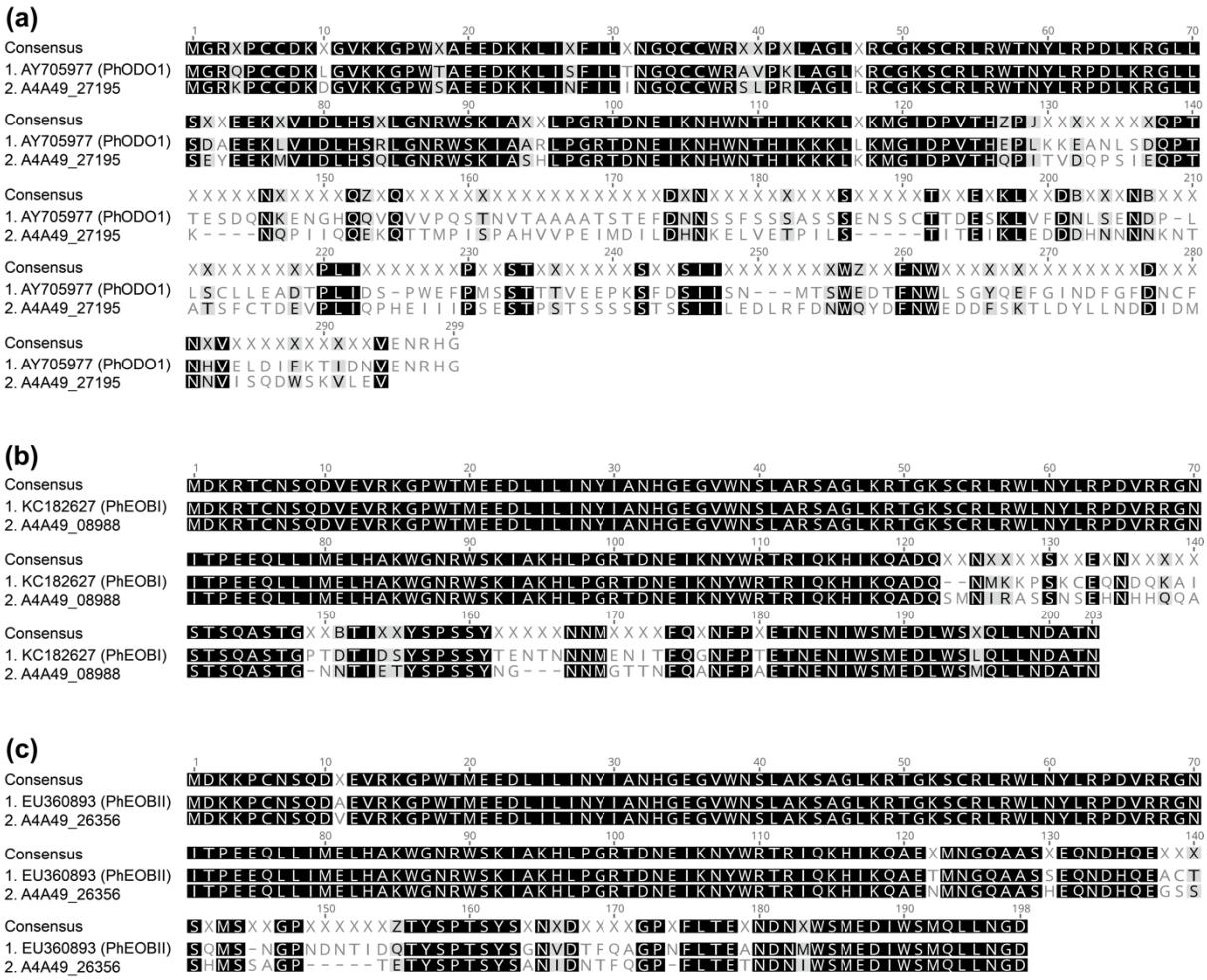


Figure S21. *N. attenuata* homologs of *Petunia×hybrida* MYB transcription factors

Protein sequence alignments of *P. hybrida* ODORANT1 (PhODO1) and *P. hybrida* EMISSION OF BENZENOID I/II (PhEOBI/II) with their *N. attenuata* homologs. Full-length amino acid sequences were aligned using Clustal W implemented in the Geneious software. **(a)** Among the PhODO1 homologs searched by BLASTP on *N. attenuata* protein sequences, A4A49_27195 (NaODO1) is a gene that are meaningfully expressed in *N. attenuata* flowers. **(b, c)** A4A49_08988 (NaEOBI) and A4A49_26356 (NaEOBII) are the MYB transcription factors homologous to PhEOBI and PhEOBII.

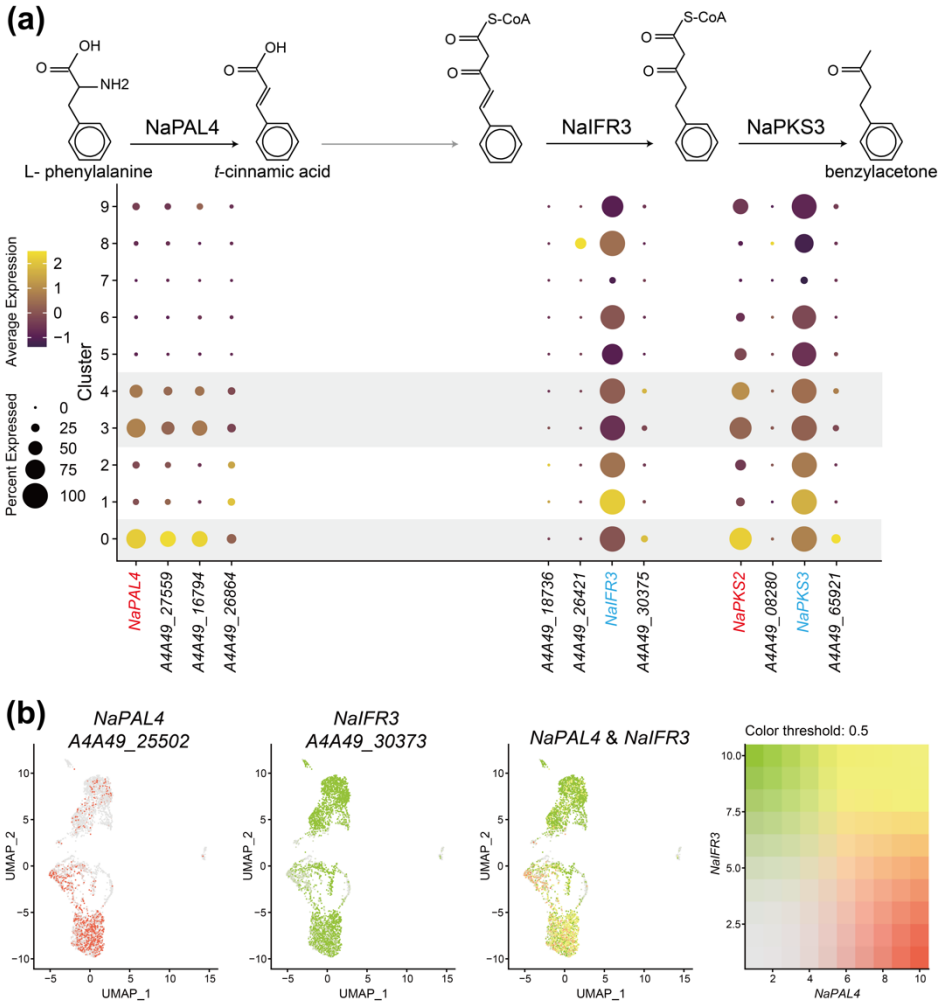


Figure S22. *NaIFR3* expression patterns in scRNA-seq dataset

(a) Dotplot showing previous BA biosynthesis model, including IFR3 (Guo *et al.*, 2020). Genes marked in red (*NaPKS2*) are components of new BA biosynthetic model suggested in this study, and blue-colored genes (*NaIFR3* and *NaPKS3*) are components of previous model. Homologous genes of each gene are denoted next to them in black. *NaIFR3* as well as *NaPKS3* are expressed across almost all clusters. **(b)** *NaPAL4* and *NaIFR3* are merged on featureplots. *NaIFR3* is expressed also in upper clusters, where *NaPAL4* is less expressed.

Reference

Brockmöller T, Ling Z, Li D, Gaquerel E, Baldwin IT, Xu S. **2017**. *Nicotiana attenuata* Data Hub (*Na*DH): An integrative platform for exploring genomic, transcriptomic and metabolomic data in wild tobacco. *BMC Genomics* **18**: 1–11.

Chen LQ, Qu XQ, Hou BH, Sosso D, Osorio S, Fernie AR, Frommer WB. **2012**. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* **335**: 207–211.

Felsenstein J. **1985**. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783.

Guo H, Lackus ND, Köllner TG, Li R, Bing J, Wang Y, Baldwin IT, Xu S. **2020**. Evolution of a novel and adaptive floral scent in wild tobacco. *Molecular Biology and Evolution* **37**: 1090–1099.

Kessler D, Gase K, Baldwin IT. 2008. Field experiments with transformed plants reveal the sense of floral scents. *Science* **321**:1200–1202. doi:10.1126/science.1160072

Kim J-Y, Symeonidi E, Pang TY, Denyer T, Weidauer D, Bezrutczyk M, Miras M, Zöllner N, Hartwig T, Wudick MM, *et al.* 2021. Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *The Plant Cell* **33**:511–530. doi: 10.1093/plcell/koaa060

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. **2018**. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**: 1547–1549.

Li-Beisson Y, Pollard M, Sauveplane V, Pinot F, Ohlrogge J, Beisson F. **2009**. Nanoridges that characterize the surface morphology of flowers require the synthesis of cutin polyester. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 22008–22013.

Li D, Heiling S, Baldwin IT, Gaquerel E. 2016. Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory. *Proceedings of the National Academy of Sciences of the United States of America* **113**:E7610–E7618. doi:10.1073/pnas.1610218113

Maruyama-Nakashita A, Watanabe-Takahashi A, Inoue E, Yamaya T, Saito K, Takahashi H. **2015**. Sulfur-responsive elements in the 3′-nontranscribed intergenic region are essential for the induction of *SULFATE TRANSPORTER 2;1* gene expression in *Arabidopsis* roots under sulfur deficiency. *Plant Cell* **27**: 1279–1296.

Saitou N, Nei M. **1987**. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**: 406–425.

Suh MC, Samuels AL, Jetter R, Kunst L, Pollard M, Ohlrogge J, Beisson F. **2005**. Cuticular lipid composition, surface structure, and gene expression in *Arabidopsis* stem epidermis. *Plant Physiology* **139**: 1649–1665.

Tang J, Yang X, Xiao C, Li J, Chen Y, Li R, Li S, Lü S, Hu H. **2020**. GDSL lipase occluded stomatal pore 1 is required for wax biosynthesis and stomatal cuticular ledge formation. *New Phytologist* **228**: 1880–1896.

ZUCKERKANDL E, PAULING L. **1965**. Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJBT-EG and P, eds. Academic Press, 97–166.

Table S1. 121 genes excluded for analysis

Locus tag ID	NCBI_gene description
A4A49_04901	exocyst complex component exo70a1
A4A49_08097	uncharacterized WD repeat-containing protein alr3466-like
A4A49_08976	NAC domain-containing protein 2-like
A4A49_16875	E3 ubiquitin-protein ligase ATL31-like
A4A49_35076	putative calcium-binding protein CML19
A4A49_01686	hypothetical protein
A4A49_05562	u-box domain-containing protein 21
A4A49_42257	uncharacterized LOC109211123
A4A49_13806	NDR1/HIN1-like protein 3-like
A4A49_08885	nodulin-related protein 1-like
A4A49_01593	ethylene-responsive transcription factor erf109
A4A49_41221	exocyst complex component EXO70A1-like
A4A49_30015	ethylene-responsive transcription factor erf019
A4A49_38493	uncharacterized LOC109223861
A4A49_37763	uncharacterized LOC109223885
A4A49_37762	uncharacterized LOC109223886
A4A49_08704	hypothetical protein A4A49_08704
A4A49_22454	uncharacterized LOC109225120
A4A49_01661	zinc finger protein ZAT10-like
A4A49_09742	hypothetical protein A4A49_09742
A4A49_03446	uncharacterized LOC109227747
A4A49_07063	hypothetical protein A4A49_07063
A4A49_25717	hypothetical protein A4A49_25717
A4A49_33780	e3 ubiquitin-protein ligase pub23
A4A49_59916	RING-H2 finger protein ATL3-like
A4A49_00112	protein UPSTREAM OF FLC-like
A4A49_01093	cycloartenol Synthase-like
A4A49_30862	dehydration-responsive element-binding protein 1f
A4A49_42272	ethylene-responsive transcription factor 4-like
A4A49_28282	E3 ubiquitin-protein ligase Hakai-like
A4A49_33511	uncharacterized LOC109231204
A4A49_04398	uncharacterized LOC109232053
A4A49_30397	putative nuclease HARBI1
A4A49_27533	probable carboxylesterase 16
A4A49_51721	hypothetical protein A4A49_51721
A4A49_12278	hypothetical protein A4A49_12278
A4A49_18295	glutaredoxin-c9
A4A49_58651	uncharacterized LOC109235426
A4A49_10323	EG45-like domain containing protein
A4A49_30267	protein PLANT CADMIUM RESISTANCE 2-like
A4A49_39253	probable protein phosphatase 2C 25
A4A49_08252	transcription factor bHLH35

A4A49_00712	hypothetical protein A4A49_00712
A4A49_28978	dehydration-responsive element-binding protein 1f
A4A49_51246	hypothetical protein A4A49_51246
A4A49_65778	17.6 kDa class I heat shock protein-like
A4A49_04460	17.6 kda class i heat shock protein
A4A49_01061	hypothetical protein A4A49_01061
A4A49_25741	classical arabinogalactan protein 5-like
A4A49_43070	hypothetical protein A4A49_43070
A4A49_52264	hypothetical protein A4A49_52264
A4A49_35025	hypothetical protein A4A49_35025
A4A49_27124	uncharacterized LOC109241637
A4A49_19893	uncharacterized LOC109241892
A4A49_27795	putative wrky transcription factor 53
A4A49_64938	protein yls9
A4A49_16953	ndr1hin1-like protein 3
A4A49_64937	ndr1hin1-like protein 3
A4A49_16956	protein YLS9-like
A4A49_27136	cytochrome P450 94A1-like
A4A49_24482	hypothetical protein A4A49_24482
A4A49_24481	uncharacterized LOC109243614
A4A49_28238	exocyst complex component exo70a1
A4A49_02919	E3 ubiquitin-protein ligase RMA1H1-like
A4A49_05755	uncharacterized LOC109243962
A4A49_25653	ethylene-responsive transcription factor erf017
A4A49_09712	iaa-amino acid hydrolase ilr1-like 1
A4A49_09948	putative calcium-binding protein cml31
A4A49_09949	calcium-binding protein cml37
A4A49_09950	putative calcium-binding protein CML19
A4A49_10806	zinc finger protein zat11
A4A49_52374	zinc finger protein ZAT12-like
A4A49_10804	zinc finger protein zat11
A4A49_22093	uncharacterized LOC109206584
A4A49_07999	uncharacterized LOC109206786
A4A49_08106	ethylene-responsive transcription factor erf109
A4A49_23019	ethylene-responsive transcription factor erf109
A4A49_23016	ethylene-responsive transcription factor erf109
A4A49_08257	oleosin 21.2 kda
A4A49_26226	hypothetical protein A4A49_26226
A4A49_34506	hypothetical protein A4A49_34506
A4A49_33923	cucmber peeling cupredoxin
A4A49_10619	zinc finger AN1 domain-containing stress-associated protein 12
A4A49_33179	uncharacterized LOC109208622
A4A49_11952	uncharacterized LOC109209072

A4A49_11575	uncharacterized LOC109209257
A4A49_12259	dehydration-responsive element-binding protein 1f
A4A49_14244	uncharacterized LOC109209866
A4A49_12965	calcium-binding protein cml38
A4A49_25324	lysine histidine transporter-like 7
A4A49_13871	hypothetical protein A4A49_13871
A4A49_21385	protein hyper-sensitivity-related 4
A4A49_16537	hypothetical protein A4A49_16537
A4A49_35987	putative glycine-rich cell wall structural protein 1
A4A49_25048	hypothetical protein A4A49_25048
A4A49_41474	ethylene-responsive transcription factor erf109
A4A49_31384	probable WRKY transcription factor 53
A4A49_24534	chaperone protein dnaj 11, chloroplastic
A4A49_25089	dehydration-responsive element-binding protein 1A-like
A4A49_28100	ethylene-responsive transcription factor erf017
A4A49_37911	hypothetical protein A4A49_37911
A4A49_62051	zinc finger protein zat11
A4A49_29168	peroxidase n1
A4A49_37140	uncharacterized LOC109216478
A4A49_36228	ethylene-responsive transcription factor ERF026-like
A4A49_39300	putative glutathione s-transferase
A4A49_38427	17.3 kDa class II heat shock protein-like
A4A49_64179	auxin-responsive protein saur32
A4A49_32032	hypothetical protein A4A49_32032
A4A49_32862	u-box domain-containing protein 21
A4A49_36195	glutathione s-transferase u10
A4A49_33478	uncharacterized LOC109218635
A4A49_37711	heavy metal-associated isoprenylated plant protein 3-like
A4A49_38235	17.3 kda class ii heat shock protein
A4A49_38953	hypothetical protein A4A49_38953
A4A49_38797	calmodulin-binding protein 25-like
A4A49_41099	exocyst complex component EXO70A1-like
A4A49_64622	uncharacterized LOC109221284
A4A49_41019	cytochrome p450 94b3
A4A49_41206	uncharacterized LOC109221416
A4A49_42489	u-box domain-containing protein 21

Table S2. Primers used in this study

Primer sequences and their purposes are listed in table. Red-colored sequences are overhang sequences added for gibson cloning, which overlap with the ends of linearized vectors digested by restriction enzymes. Black sequences anneal to the PCR template.

Name	Sequence (Red: overhang sequence for cloning)	Purpose
EF1a_qPCR_F	TGCTGTGAGGGACATGCGT	qPCR control
EF1a_qPCR_R	TCTTCTGGGCAGCCTTGGT	qPCR control
IFR1_qPCR_F	TCTCCAATGCCATTGCCTGTC	A4A49_26421_qPCR
IFR1_qPCR_R	AGGCCTCAACCCCAAATGATG	A4A49_26421_qPCR
IFR2_qPCR_F	TCAATGTCGCGTTGGCACTTA	A4A49_30375_qPCR
IFR2_qPCR_R	GGTACTCCTCCGCAGTAGTGT	A4A49_30375_qPCR
IFR3_qPCR_F	GGAACATGATGCTTGCACCTACA	A4A49_30373_qPCR
IFR3_qPCR_R	TCTCCCACATAGCCACTAACTCA	A4A49_30373_qPCR
IFR4_qPCR_F	GGGAGTTGATAGGCAAGTGGG	A4A49_18736_qPCR
IFR4_qPCR_R	GACTCCTACCTGCCCAGCATA	A4A49_18736_qPCR
PKS1_qPCR_F	TAGTGGAGTAGACATGCCGGG	A4A49_08280_qPCR
PKS1_qPCR_R	ACATCATGAGCCGCTTTACCG	A4A49_08280_qPCR
PKS2_qPCR_F	GATGGAGCAGCCGCGATTATT	A4A49_34074_qPCR
PKS2_qPCR_R	CGGCAATAAAGTTTGGGCTGC	A4A49_34074_qPCR
PKS3_qPCR_F	GGACTTAGCCGAAAACAACCAGG	A4A49_39367_qPCR
PKS3_qPCR_R	CACTTGGGCCTTGAAACGTGAAT	A4A49_39367_qPCR
PKS4_qPCR_F	TGTTTGCTCAGAGATCACCGC	A4A49_65921_qPCR
PKS4_qPCR_R	AACAAAGGCCTCTCGACCTCT	A4A49_65921_qPCR
4CL1_qPCR_F	GTTTACGCCAGCAGAGGTTGT	A4A49_40482_qPCR
4CL1_qPCR_R	AGCGTAATCCTTCACTTTGCCC	A4A49_40482_qPCR
4CL2_qPCR_F	CCCTTGGTTCTTGCGTTGGC	A4A49_41306_qPCR
4CL2_qPCR_R	CGCTTCCTCCAACTCTTTCCC	A4A49_41306_qPCR
AER1_qPCR_F	TCCAGACTTCAAGCCAGGTGA	A4A49_31470_qPCR
AER1_qPCR_R	GCATGCCAAGAAGACCGACAT	A4A49_31470_qPCR
A4A49_07158_qPCR_F	TGTGGGAGGGAAGATGCTTGA	AER homologs qPCR
A4A49_07158_qPCR_R	GCGAGATCATCCCACACACAG	AER homologs qPCR
A4A49_11980_qPCR_F	TGGAGGAAAGATGCTTGACGC	AER homologs qPCR
A4A49_11980_qPCR_R	ACTCCTTCAGACTGCTCGAGG	AER homologs qPCR
A4A49_02622_qPCR_F	CAATTGGGATTGCGTTGCCC	AER homologs qPCR
A4A49_02622_qPCR_R	AATTCCAACTCCTCCCGCTG	AER homologs qPCR
A4A49_36472_qPCR_F	CGATTTTGCTGGAGCCTTGAAA	AER homologs qPCR
A4A49_36472_qPCR_R	CAAGAGCACCTCATCCAGCATA	AER homologs qPCR
A4A49_27195_qPCR_F	AACAATGCCAATTTCCCCAGC	ODO1-like qPCR
A4A49_27195_qPCR_R	TTGTTGTGGTCGTCGTCTTCT	ODO1-like qPCR
A4A49_08988_qPCR_F	ACAAGCCAAGCTTCTACTGGAA	EOBI homolog qPCR
A4A49_08988_qPCR_R	TTTCAGCTGGAAAATTGGCCTG	EOBI homolog qPCR
A4A49_26356_qPCR_F	GGAGGACAAGGATTCAGAAGCA	EOBII homolog qPCR
A4A49_26356_qPCR_R	CAGCAGACGACATATGGCTACT	EOBII homolog qPCR
A4A49_25502_Gibson_F	ATTCGAGCTCTCCCATATGGAGCTTAAAAGAAGGCATT	promoter (2kb) Gibson pcr_pPAL4::eGFP
A4A49_25502_Gibson_R	CTAGTGCGGCCGCCTGCAGGTGCTAAAGTACTAAATATGTAT	promoter (2kb) Gibson pcr_pPAL4::eGFP
A4A49_34074_Gibson_F	ATTCGAGCTCTCCCATATGGGACTATTTCTATACGTGTCGG	promoter (2kb) Gibson pcr_pPKS2::eGFP
A4A49_34074_Gibson_R	CTAGTGCGGCCGCCTGCAGGTTTTCGCCGGAAAAAATGG	promoter (2kb) Gibson pcr_pPKS2::eGFP
pTRV2:eGFP_F	AGGTTACCGAATTCTCTGTTCACCGGGGTG	pTRV2:eGFP gibson assembly (XbaI/XhoI) - for negative control of VIGS
pTRV2:eGFP_R	GGACATGCCCGGGCCCTTGTCGGCCATGAT	pTRV2:eGFP gibson assembly (XbaI/XhoI) - for negative control of VIGS
pTRV2:A4A49_40482-F	AGGTTACCGAATTCTACATCCCTAAGCATT	pTRV2:A4A49_40482 gibson assembly (XbaI/XhoI) - VIGS - 4-Coumarate CoA ligase 1
pTRV2:A4A49_40482-R	GGACATGCCCGGGCCATGGCTCCTAAATAT	pTRV2:A4A49_40482 gibson assembly (XbaI/XhoI) - VIGS - 4-Coumarate CoA ligase 1
pTRV2:A4A49_30373_F	AGGTTACCGAATTCTGAAGGAATTCCTTATACT	pTRV2:A4A49_30373 gibson assembly (XbaI/XhoI) - VIGS - Isoflavone reductase 3
pTRV2:A4A49_30373_R	GGACATGCCCGGGCCACTCATTGAACGATAAT	pTRV2:A4A49_30373 gibson assembly (XbaI/XhoI) - VIGS - Isoflavone reductase 3
pTRV2:A4A49_39367_F	AGGTTACCGAATTCTACAAAATGCCATCGA	pTRV2:A4A49_39367 gibson assembly (XbaI/XhoI) - VIGS - Polyketide synthase 3
pTRV2:A4A49_39367_R	GGACATGCCCGGGCCGATCTCCGAGCAAAC	pTRV2:A4A49_39367 gibson assembly (XbaI/XhoI) - VIGS - Polyketide synthase 3
pTRV2:A4A49_31470_F	AGGTTACCGAATTCTGTTGAGCTACTAGGA	pTRV2:A4A49_31470 gibson assembly (XbaI/XhoI) - VIGS - 2-alkenal reductase 1
pTRV2:A4A49_31470_R	GGACATGCCCGGGCCCTTCAATTACCTCTCCA	pTRV2:A4A49_31470 gibson assembly (XbaI/XhoI) - VIGS - 2-alkenal reductase 1
pET50-4CL1_gibson_F	TTTAAGAAGGAGATATACAATGCCTATGGAGAGTACA	codon optimized sequence cloned into NdeI and PstI cutted pET50 (gibson)
pET50-4CL1_gibson_R	CGGAGCTCGCCTGCACATTTGGAACGCCCG	codon optimized sequence cloned into NdeI and PstI cutted pET50 (gibson)
pET50-PKS3_gibson_F	TTTAAGAAGGAGATATACAATGGTCACTGTAGAGGA	codon optimized sequence cloned into NdeI and PstI cutted pET50 (gibson)
pET50-PKS3_gibson_R	CGGAGCTCGCCTGCACAGTTGAGACACTATGTAAT	codon optimized sequence cloned into NdeI and PstI cutted pET50 (gibson)
pET50-PKS2_cloning_F	TTTAAGAAGGAGATATACAATGGTCACCGTCGAGG	PKS2 sequence (without codon optimization) cloned into NdeI and PstI cutted pET50 (gibson)
pET50-PKS2_cloning_R	CGGAGCTCGCCTGCAAAGTAGAGACACTATGGAGCA	PKS2 sequence (without codon optimization) cloned into NdeI and PstI cutted pET50 (gibson)
pET50-AER1_cloning_F	TTTAAGAAGGAGATATACAATGGAGAAAAAAGCAGGAAAAG	AER1 sequence (without codon optimization) cloned into NdeI and PstI cutted pET50 (gibson)
pET50-AER1_cloning_R	CGGAGCTCGCCTGCAAATCGTGCGCCACGC	AER1 sequence (without codon optimization) cloned into NdeI and PstI cutted pET50 (gibson)