Article

Current Biology

Pseudogenization and Resurrection of a Speciation Gene

Highlights

- The R2R3-MYB transcription factor AN2 is a hotspot for evolutionary change
- Resurrection of AN2 by a 2-bp deletion during reversal to bee pollination
- Complex trait modified by two regulatory mutations of large effect
- Trait reversal is possible in recent radiations

Authors

Korinna Esfeld, Andrea E. Berardi, Michel Moser, Eligio Bossolini, Loreta Freitas, Cris Kuhlemeier

Correspondence

cris.kuhlemeier@ips.unibe.ch

In Brief

Floral color is a crucial signal to attract pollinators. The R2R3-MYB transcription factor AN2 is a hotspot for color evolution in *Petunia*. Esfeld, Berardi, Moser, et al. document the resurrection of *AN2* from a pseudogene background that led to the re-gain of purple pigmentation in *Petunia secreta*.

Esfeld et al., 2018, Current Biology *28*, 1–11
December 3, 2018
https://doi.org/10.1016/j.cub.2018.10.019

CellPress

Please cite this article in press as: Esfeld et al., Pseudogenization and Resurrection of a Speciation Gene, Current Biology (2018), https://doi.org/10.1016/j.cub.2018.10.019

Current Biology
Article

Pseudogenization and Resurrection of a Speciation Gene

Korinna Esfeld,[1,4] Andrea E. Berardi,[1,4] Michel Moser,[1,4] Eligio Bossolini,[1,3] Loreta Freitas,[2] and Cris Kuhlemeier[1,5,*]
[1]Institute of Plant Sciences, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland
[2]Department of Genetics, University Fed Rio Grande do Sul, POB 15053, Porto Alegre, 91501970 Rio Grande do Sul, Brazil
[3]Present address: BASF Agricultural Solutions Belgium NV, Technologiepark-Zwijnaarde, 389052 Gent, Belgium
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: cris.kuhlemeier@ips.unibe.ch
https://doi.org/10.1016/j.cub.2018.10.019

SUMMARY

A persistent question in evolutionary biology is how complex phenotypes evolve and whether phenotypic transitions are reversible. Multiple losses of floral pigmentation have been documented in the angiosperms, but color re-gain has not yet been described, supporting that re-gain is unlikely. Pollinator-mediated selection in *Petunia* has resulted in several color shifts comprised of both losses and gains of color. The R2R3-MYB transcription factor *AN2* has been identified as a major locus responsible for shifts in pollinator preference. Whereas the loss of visible color has previously been attributed to repeated pseudogenization of *AN2*, here, we describe the mechanism of an independent re-gain of floral color via *AN2* evolution. In *P. secreta*, purple color is restored through the improbable resurrection of *AN2* gene function from a non-functional AN2-ancestor by a single reading-frame-restoring mutation. Thus, floral color evolution in *Petunia* is mechanistically dependent on AN2 functionality, highlighting its role as a hotspot in color transitions and a speciation gene for the genus.

INTRODUCTION

There is accumulating evidence for mutations of large phenotypic and fitness effect as the major driver of adaptation and speciation, in contrast to historically favored theory, where infinite numbers of small effect mutations drive evolution and speciation [1, 2] (but see [3]). Although mutations of small phenotypic effect are typically difficult to identify, with present technology, those of large phenotypic effect can be identified more easily. This provides an opportunity to study the molecular mechanisms underlying evolutionary changes in complex traits. Trait evolution is often directional [4], and in particular, trait reversal is thought to be rare. This is the essence of Dollo's second law, which states that the loss of a complex trait will rapidly become irreversible [5, 6].

Pollinator-mediated speciation provides an evolutionary scenario in which combinations of polygenic traits, known as pollination syndromes, are under strong selection, providing a natural system in which to look for mutations and loci of evolution. Evolutionary shifts in pollination systems have been observed frequently and are marked by changes in pigmentation, scent, nectar, and floral morphology. Such adaptations to pollinators are thought to have driven the remarkable floral diversity of the angiosperms and are accompanied by lineage splitting [7–11]. However, changes in a single floral trait can be enough to strongly alter pollinator preference, with color being no exception [12–16]. The color of a flower advertises the presence of a reward to a prospective pollinator and is therefore crucial in plant-pollinator interactions. Bees, for example, are attracted by pink, yellow, or purple flowers and avoid red ones, whereas red flowers are associated with bird pollination [17–19]. Patterns of UV color can give pollinator guidance on where to find a reward on the flower [20]. Additionally, whole flowers being either completely UV absorbent or UV reflective can influence nighttime pollinator preference between "human white" flowers [16].

The majority of angiosperms use flavonoids for floral pigmentation, with anthocyanins contributing to visible color and flavonols being responsible for UV color. Flavonoids are synthesized as part of a complex network of phenylpropanoids that includes a large variety of primary and secondary compounds, such as lignins, volatile signals, developmental regulators, and defense compounds [21]. Flavonoid biosynthesis has been studied in great detail, with core enzymes and transcriptional regulators characterized in many taxa [21, 22]. In the flavonoid pathway, the last common substrates of anthocyanins and flavonols are the dihydroflavonols (Figure 1). Flavonol synthase (FLS) action yields the flavonols, whereas anthocyanins are produced by the sequential action of dihydroflavonol-4-reductase (DFR) and ANS and are further modified by six enzymes, which produce the orange-red, magenta, or blue end products (for details, see Figure 1) [21]. In eudicots, the pathway is regulated by a complex of basic-helix-loop-helix (bHLH), WD40, and MYB transcription factors that targets the pathway genes in early and late sets, with MYB transcription factors yielding spatial and temporal specificity [23–25].

Evolutionary changes in genes of the flavonoid pathway have been implicated in adaptation to pollinators in many plant taxa, and the in-depth knowledge of the flavonoid pathway makes informed prediction of the evolutionary mechanisms of color change possible. For instance, mutations in regulatory genes predominantly lead to loss of visible flower color [13, 26–30]. In

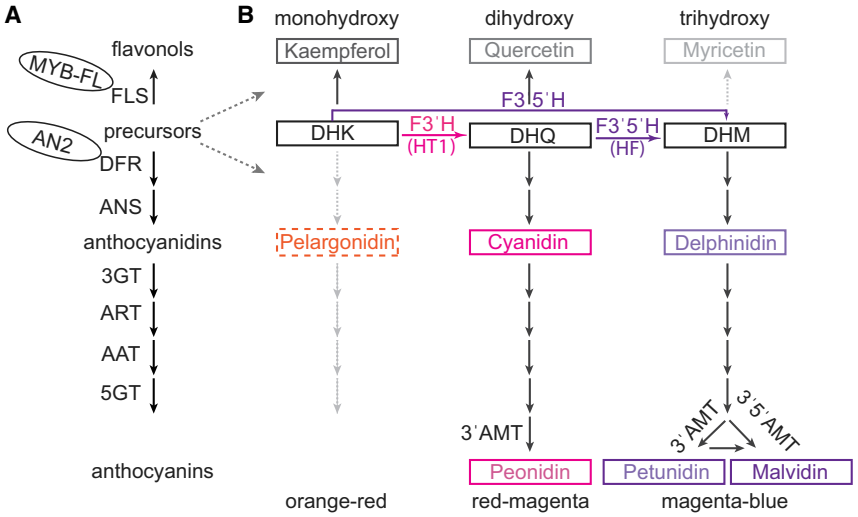

Figure 1. Flavonoid Pathway in *Petunia*

(A) Schematic representation of the flavonoid biosynthetic pathway genes with transcription factors MYB-FL and AN2 (circled). Transcription factor MYB-FL induces FLS, whereas AN2 activates DFR and downstream anthocyanin biosynthetic genes.

(B) Depiction of the *Petunia* flavonoid biosynthetic pathway where flavonol (UV color) as well as anthocyanin (visible color) pigments are produced (division indicated by the dashed arrows). DHK, DHQ, and DHM are the last common dihydroflavonol precursors to both flavonols and anthocyanins. Dihydroflavonols are modified by the pathway branching enzymes F3′H and F3′5′H, which add hydroxyl groups to the flavonoid B ring. At each branching node of this series of hydroxylating steps, FLS and DFR can compete for dihydroflavonol substrate. FLS action yields the flavonols kaempferol, quercetin, and myricetin, in increasing B-ring hydroxylation order. The anthocyanins are produced by the sequential action of DFR and ANS. Further modifications by a suite of six enzymes yield, in increasing B-ring hydroxylation order: the brick-red to orange pelargonidin; the red to magenta cyanidin and peonidin; as well as the blue to purple delphinidin, petunidin, and malvidin. In *Petunia*, the enzyme DFR does not accept DHK as a substrate, causing the absence of the orange-red pelargonidin (indicated in light gray arrows). The enzyme FLS has low activity on DHM, causing the severely reduced presence of myricetin. Increased hydroxylation of anthocyanidins through the action of F3′H/HT1 and F3′5′H/HF shift color from red toward purple-blue. Glycosylation, acylation, and methylation of the anthocyanidin backbone by a set of six enzymes have a similar effect.

Abbreviations: 3′AMT, 3′-anthocyanin methyltransferase; 3′5′AMT, 3′5′-anthocyanin methyltransferase; 3GT, anthocyanin-3-glucosyltransferase; 5GT, anthocyan 5 glucosyltransferase; AAT, anthocyanin acyltransferase; ART, anthocyanin rhamnosyltransferase; DFR, dihydroflavonol-4-reductase; DHK, dihydrokaempferol; DHM, dihydromyricetin; DHQ, dihydroquercetin; F3′H/HT1, flavonoid-3′-hydroxylase, encoded by *HT1*; F3′5′H/HF, flavonoid-3′5′-hydroxylase, encoded by *HF1* and *HF2*; FLS, flavonol synthase.

contrast, mutations in the biosynthetic genes, whether by inactivation, deletion, or downregulation, predominate in color transitions [9, 14, 31–36]. Thus, these findings help to pinpoint mutations involved in changing floral color between species, whether it is a loss, a gain, or a shift.

To identify mutations that drive floral color evolution, we chose the recently and adaptively radiated South American genus *Petunia* (Solanaceae). Twenty species of *Petunia* are recognized and are separated into two distinct clades based on major phenotypic differences; they are thus named the short-tube clade and long-tube clade (also known as clade I and II, respectively), which split ~2.85–1.3 million years ago (MYA) [37, 38]. Like in many flowering genera, flower color in *Petunia* is associated with different pollination syndromes. The fourteen species of the short-tube clade are bee pollinated and have the ancestral purple flowers, whereas the species of the long-tube clade are diverse in flower color and pollinators [39]. Specifically, *P. axillaris* flowers are white, UV absorbent, and hawkmoth pollinated; *P. exserta* is bright red, UV reflective, and hummingbird pollinated; and the bee-pollinated *P. secreta* is purple and UV reflective (Figures 2A and 2B). Hybridization between *Petunia* species is possible but, in nature, only observed between species of the long-tube clade [37, 40, 41]. Thus, the main isolating mechanisms in *Petunia* are thought to be geography and pollinator preference, with no evidence for postzygotic barriers [42].

Petunia is a model system for the study of floral pigmentation and genetic architecture of evolutionary shifts in pigmentation [43]. For example, the loss of color (purple to white) during the transition from bee to hawkmoth pollination in *P. axillaris* is simple and involves inactivating mutations in the gene encoding the R2R3-MYB transcription factor *AN2* [13, 26]. Likewise, loss of UV pigmentation in the transition from hawkmoth to hummingbird pollination was accompanied by a frameshift mutation in the R2R3-MYB transcription factor *MYB-FL*, leading from UV-absorbent flowers in *P. axillaris* to UV-reflective ones in *P. exserta* [16]. Thus, shifts in visible and UV color are attributed to mutations in MYB transcription factors with large phenotypic effect, allowing for subsequent adaptation to different pollinators. However, our understanding of the evolution of flower color in *Petunia* remains incomplete. The observed color variation among species of the long-tube clade could be the result of either an independent loss of color in the white *P. axillaris* or a reacquisition of color in the purple *P. secreta* and red *P. exserta*. Re-gain of anthocyanin pigmentation is a less likely scenario: first, because reacquiring a gene function is more demanding than losing it and, second, because once a pathway becomes obsolete, release from selective pressure will leave its genes free to accumulate deleterious mutations [44, 45], making reversal less likely with the passing of time. Indeed, observed losses of floral pigmentation appear to be largely unidirectional, reviewed in [4].

Here, we used phylogenetic reconstructions, sequence comparisons, as well as functional analyses to demonstrate that *P. secreta* is colored due to the improbable and independent reacquisition of AN2 function. This rare example of gene resurrection and trait reversal conforms to the essence of Dollo's second law [46–49] and contributes to long-standing questions of how evolution proceeds.

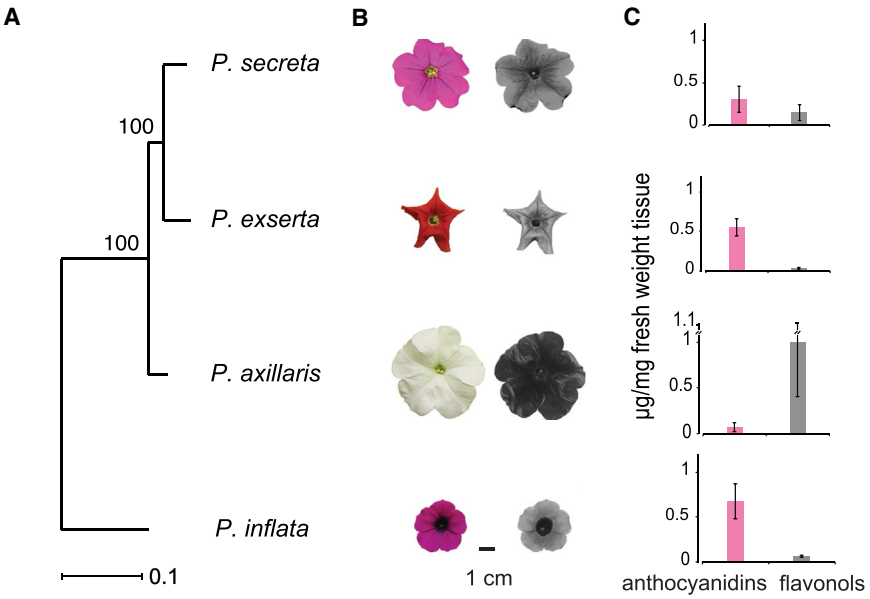

Figure 2. Phylogenetic Relationship and Color Phenotypes of the Species of Petunia

(A) Species tree adapted from a maximum likelihood analysis based on 399,673 nucleotide sites revealed from RNA sequencing (RNA-seq) data of the four species. The tree was rooted with *P. inflata*.

(B) The four species are shown photographed under visible (left) and UV light (right).

(C) Concentrations of anthocyanidins (pink) and flavonols (gray) present in the four species as determined by liquid chromatography-mass spectrometry (LC-MS) (mean ± SD; n = 3 per species).

RESULTS

Colored *Petunia* Species in the Long-Tube Clade Are Sister to the White-Flowered Species

Phylogenic trees of recent radiations often lack resolution due to ongoing gene flow, incomplete lineage sorting, and standing genetic variation [50, 51]. A robust phylogenetic reconstruction is, however, a prerequisite for determining the direction of evolutionary changes. Previously, the separation of the short-tube and the long-tube clades has been well supported, but relationships within the clades are uncertain and highly influenced by the loci used [38, 39]. We constructed a species tree of our four studied species with the goal of resolving long-tube species relationships based on two different datasets. First, we used a concatenated maximum likelihood approach with approximately 400,000 SNPs (40,758 informative sites) positioned in genic regions spread over the genome. Second, amino acid sequences of 984 shared single-copy orthologs were used to estimate a genomic-scale coalescent-based species tree. Both approaches yielded the same topology. The species tree is strongly supported at its nodes and shows, in contrast to the previously published species tree [39], that, in the long-tube clade, the pigmented *P. secreta* and *P. exserta* form a species pair that is sister to the white-flowered *P. axillaris* (SNP tree shown in Figure 2A; see STAR Methods).

The R2R3-MYB Transcription Factor AN2 Is a Hotspot for Color Evolution

The transcription factor AN2 has played a major role in *Petunia* flower color trait evolution [13, 26], but its sequence and functionality have not yet been established for most of the genus. We therefore reconstructed the evolutionary history of this gene across *Petunia* with the multispecies coalescent method using both coding and genomic AN2 sequences of all species. Similar to the previously published *Petunia* species tree [39], our phylogenetic analysis of the *AN2* gene showed a robust separation in the short-tube and long-tube clades (rather than separating according to flower color), but relationships within these clades remain unresolved (Figures 3A and S1).

The AN2 sequences could be further grouped into eleven haplotypes (H1–H11) by considering exclusively those SNPs that affect AN2 protein function ignoring other sequence variation (Figures 3A, 3B, and S1; Table S1). H1 is shared by all short-tube clade species and encodes an intact AN2 protein, as expected for these purple-flowered species [13]. In 123 *P. axillaris* accessions, seven distinct haplotypes are present (Tables S1 and S2; see also [13]). H2–H4 contain premature stop codons; H5–H7 and H10 have frame shifts (Figure 3B). Therefore, these AN2 haplotypes encode truncated, non-functional AN2 proteins in line with the loss of anthocyanins in *P. axillaris* [13, 26]. Among 42 *P. exserta* accessions, three haplotypes were found, all containing nonsense or frameshift mutations (Figures 3A and 3B; Tables S1 and S2). H8 and H9 are unique to *P. exserta*, whereas H10 is shared with *P. axillaris* (Figures 3A, 3B, and S1; Tables S1 and S2). Interestingly, even though *P. exserta* flowers are intensely pigmented with anthocyanins (Figures 2B and 2C), they contain inactive *AN2* genes and must synthesize anthocyanins by an AN2-independent mechanism. In contrast, all 25 *P. secreta* accessions investigated have a single *AN2* haplotype: H11, which contains deletions of nucleotide (nt) positions 380 and 383–384 relative to the coding region of the active AN2 protein of the short-tube clade. This results in the replacement of amino acids ASN-LYS by THR in *P. secreta* (Figure 3C). *P. secreta AN2* is highly expressed in immature petals (Figure 4; Table S3) and stable expression of *P. secreta AN2* in a *P. axillaris* background restored the purple flower color, whereas downregulation of *AN2* by virus-induced gene silencing (VIGS) resulted in white sectors in the purple flowers (Figure 3D). Therefore, H11 encodes a functional AN2 protein that is required for the purple flower color of *P. secreta.*

The Functional H11 Arose Independently within the Long-Tube Clade

Although *P. secreta* is the only species in the long-tube clade with an intact AN2 protein, H11 did not derive its functionality from the short-tube clade H1. In *AN2* gene trees with and without introns, *P. secreta AN2* is placed well within the long-tube clade with high posterior support (Figures 3A and S1). When comparing all short-tube AN2 sequences with all long-tube

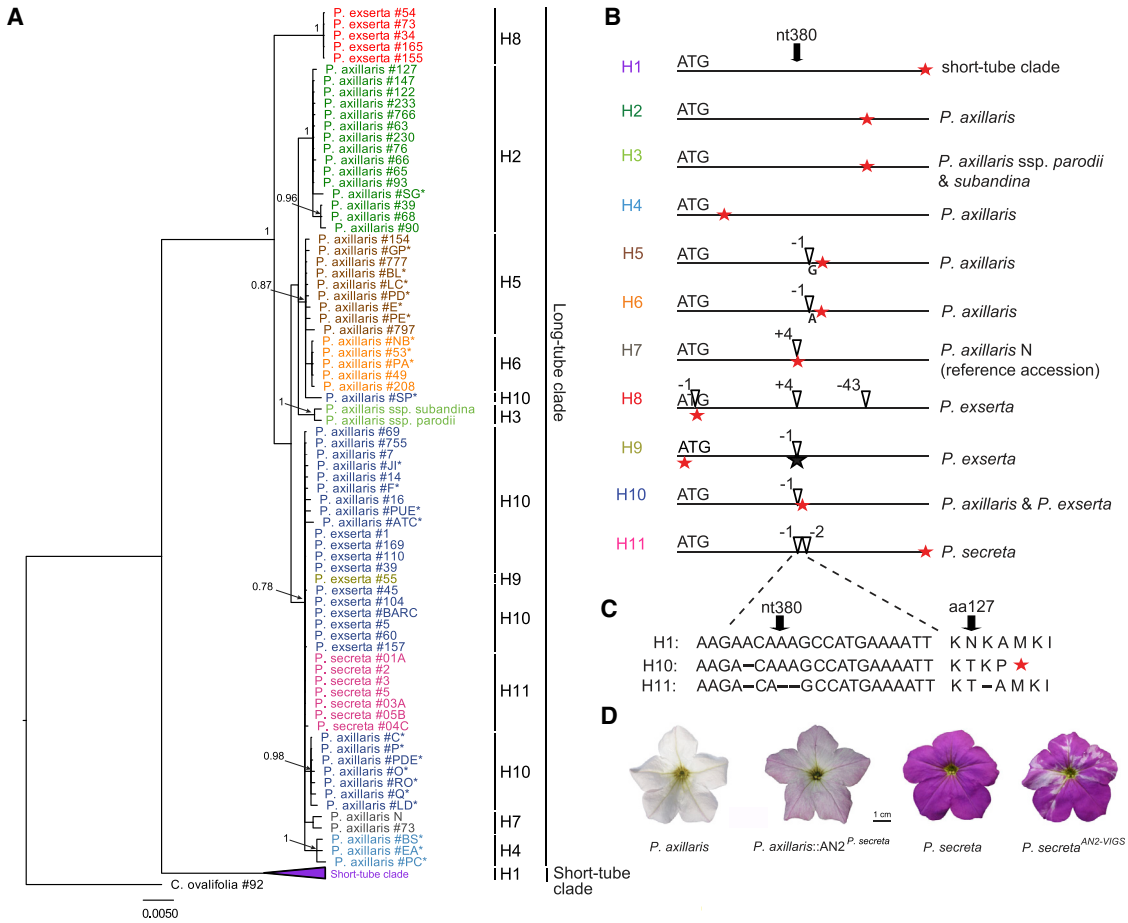

Figure 3. The R2R3-MYB Transcription Factor AN2 Is a Hotspot for Color Evolution

(A) Bayesian inference of coding sequences of the *Petunia AN2* gene. H1 and H11 encode functional proteins, whereas H2–H10 are pseudogenes.

(B) Graphical display of the coding region of *AN2* ordered into the eleven haplotypes (H) classified according to the nature and position of the inactivating mutations. ATG, start codon: nt380: nucleotide 380; inverted triangle, indel, with number above indicating number of bases deleted or inserted; red star, stop codon.

(C) Partial nucleotide and translated protein sequences of the short-tube species *P. inflata* (H1), *P. axillaris*/*P. exserta* (H10), and *P. secreta* (H11). The 1-bp deletion in *P. axillaris*/*P. exserta* translates into a truncated protein. A compensatory 2-bp deletion leads to the functional AN2 of *P. secreta*.

(D) Introduction and silencing of the *P. secreta AN2* gene modify floral color (from left to right): *P. axillaris;* transgenic *P. axillaris*::*AN2*$^{P.\ secreta}$; *P. secreta;* *P. secreta*$^{AN2\text{-VIGS}}$; note the white spots on the purple flower. Starred accessions are based on the coding sequence, see Table S1.

See also Figures S1 and S2.

AN2 sequences, 50 fixed SNPs and six fixed indels were found across the gene. From these 50 SNPs, 21 are found in the exons and 12 lead to amino acid replacements, of which seven are non-conservative missense mutations (Figure S2A). Thus, *P. secreta* contains long-tube-specific differences, distinguishing the *AN2* of *P. secreta* from the *AN2* of the short-tube species and placing it unequivocally in the long-tube clade (Figures 3A, S1, and S2). Taken together, it is highly improbable that H11 directly descended from the short-tube H1 based on sequence diversity and phylogenetic distance.

What, then, could be the origin of H11? *P. axillaris* and *P. exserta* share the non-functional H10, and H11 of *P. secreta* is most closely related to this haplotype (Figures 3A, 5, S1, and S3). H10 is characterized by its 1-bp deletion at nucleotide position 380, which puts a stop codon in-frame truncating the protein. H11 shares this 1-bp deletion, but the additional 2-bp deletion at position 383-384 restores the reading frame. This 2-bp deletion is the only fixed difference between H11 (*P. secreta*) and H10 (*P. axillaris* and *P. exserta*; Figure 3). There are three possible scenarios for the origin of H11 (Figure S4): (1) both H10 and H11 were present as standing genetic variation in the ancestral population leading to the long-tube clade and were differentially fixed in the observed extant species (which excludes neither incomplete lineage sorting nor subsequent gene flow), (2) H11 is an ancestral haplotype that gave rise to H10, or (3) the contrary where H11 is a novel allele that arose from H10 or an H10-like ancestor. The ancestral state reconstruction (Figure 5A) rules out the first hypothesis of standing genetic variation because there is no posterior support for H11 being present at the ancestral node of the long-tube clade (Table S4). For (2) to be true, the independent insertion of the same two base pairs (nt 383 and 384) must have occurred at exactly

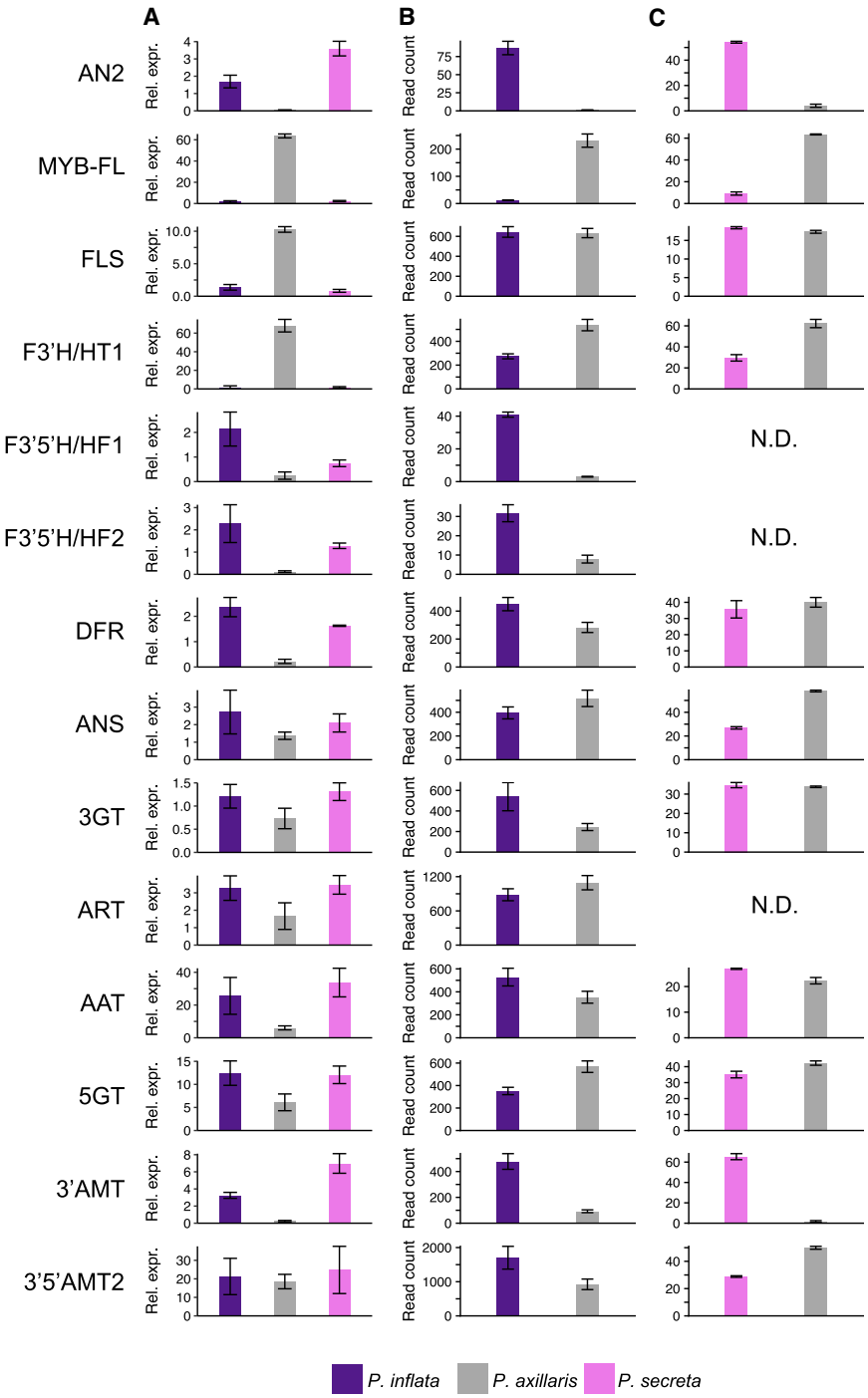

Figure 4. AN2 and MYB-FL Redesign Flavonoid Pathway Expression in *P. secreta*

(A) Relative expression as obtained by qRT-PCR of the two transcription factor genes *AN2* and *MYB-FL* as well as of biosynthetic genes of the flavonoid pathway conducted on stage 4 limb tissue. Expression is relative to expression of the reference gene *SAND*. Bars show means of relative expression from three biological replicates; error bars ± SE.

(B and C) Allele-specific expression (ASE) conducted on F_1 hybrids between *P. axillaris* and *P. inflata* (B) and ASE conducted on F_1 hybrids between *P. axillaris* and *P. secreta* (C). N.D., not determined due to lack of SNPs. For ASE, mapped reads were averaged over the SNPs between the parental alleles (see STAR Methods). Bars show means of mapped reads for the three biological replicates. Error bars represent ±SE.

See also Figure S5 and Tables S3 and S5.

the same position in two separate species whose divergence must have predated these events (Figures 2A and 3C) or H10 was subsequently shared between the two species by gene flow. In scenario (3) an ancestral H10 could have led to H11 by a simple 2-bp deletion only in the *P. secreta* lineage. Thus, from a parsimonious point of view, it is most likely that the 2-bp deletion in H11 occurred in a subpopulation containing H10 (or an H10-like ancestor). It became fixed in what we now know as *P. secreta*, and H10 was maintained in *P. axillaris* and *P. exserta* populations.

This hypothesis is further supported by our ancestral state reconstruction, which positions the fixation of functional H11 in a background of the non-functional H10-like ancestors (posterior probability of 1; Figure 5A; Table S4). Likewise, the haplotype network analysis places H11 within a larger group of H10 individuals (Figure 5B). When indels, such as these base pair deletions, are considered, H11 becomes completely distinct from H10 by this 2-bp deletion but still connected to the largest H10 node (Figure S3). Additionally, we observe that *AN2* is identical across all *P. secreta* populations, whereas other *Petunia* species and populations have more than one haplotype (Table S2). This raises questions regarding whether the observed frequency of H11 in *P. secreta* is due to allele age; patterns of selection (i.e., selective sweep); or the result of neutral processes, demographic events, and/or genetic hitchhiking, but the lack of polymorphism prevents further insight at this time. Based on the molecular data, the ancestral state reconstruction, and the haplotype network, we propose that H11 descended from H10. This implies that the purple color of *P. secreta* was re-gained through the mechanism of pseudogene resurrection.

AN2 and MYB-FL Redesign Flavonoid Pathway Expression in *P. secreta*

Introduction of *P. secreta AN2* into *P. axillaris* induces anthocyanin biosynthesis, and virus-induced *AN2* silencing in *P. secreta* inhibits it (Figure 3D). Therefore, *P. secreta* AN2 is expected to activate anthocyanin biosynthesis, as it does in

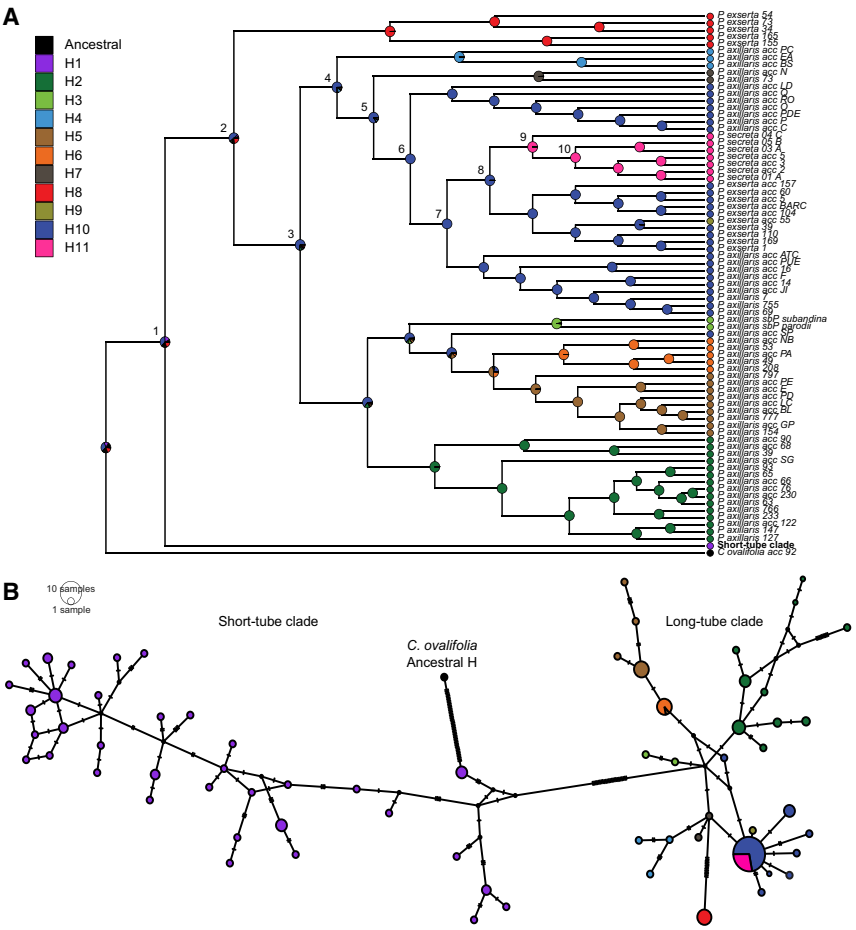

Figure 5. The Functional H11 Arose Independently within the Long-Tube Clade: Ancestral State Reconstruction of *AN2* Haplotypes Shows that the Novel *P. secreta* Functional Haplotype (H11) Arose from the Non-functional H10 of *P. axillaris* and *P. exserta*

(A) Stochastic mapping of haplotype evolution is shown on the *AN2* gene tree. Pies at nodes represent frequencies of node states across 1,000 simulations of haplotype character evolution. Circles at the tips represent the observed H1–H11. Posterior probabilities for the numbered nodes are given in Table S4. Haplotype 1 (H1) is found in all species of the short-tube clade and is represented as a single, collapsed tip. This analysis shows that the functional and derived *P. secreta* H11 arose from a background of non-functional H10 haplotypes, which is shared by *P. axillaris* and *P. exserta*.

(B) Median joining haplotype network of AN2 haplotypes with hatches representing mutations and node diameter proportional to the number of individuals, again showing that H11 is closely related to H10 based on sequence identity of unambiguous sites (see Figure S3 for haplotype network considering indels).

See also Figures S3 and S4 and Table S4.

P. inflata. Indeed, anthocyanin biosynthesis-specific loci, such as *DFR*, *ANS*, *3GT*, *ART*, and *AAT*, are strongly expressed at comparable levels in both purple-flowered species but are low in the white *P. axillaris* (Figure 4A; Table S3). In comparison, the genes encoding MYB-FL and FLS show the opposite expression profile, low in the UV-reflective species and high in the UV-absorbent *P. axillaris*, as expected for genes involved in flavonol biosynthesis. In order to determine whether the low expression of *MYB-FL* in *P. secreta* is caused by a mutation in the gene itself or by a *trans*-acting factor, we performed an allele-specific expression (ASE) analysis. In the *P. axillaris* × *P. secreta* F_1 hybrid, 88% of *MYB-FL* transcripts derived from the *P. axillaris* allele, indicating a *cis*-regulatory mutation (Figure 4C; Table S5). Indeed, the *P. secreta MYB-FL* gene contains a *Dad-one*-like transposon [52], replacing the last five amino acids and introducing a stop codon, one codon before the canonical stop codon in *P. axillaris* (Figure S5; GenBank: MH732766). This could cause functional impairment of the MYB-FL protein and reduced transcript stability due to nonsense-mediated decay (Figures 4A and S5; see also [53]).

Combinations of AN2 and MYB-FL regulate the relative amounts of anthocyanins and flavonols (quantity) but also their composition (spectral quality). Composition of flavonoids is a result of competition between FLS and DFR, as well as the pathway-branching enzymes F3′H and F3′5′H, which catalyze the synthesis of dihydroxylated and trihydroxylated flavonoids, respectively (Figure 1). *P. inflata* and *P. secreta* both synthesize the trihydroxylated malvidin and petunidin (Figures 6A and 6C). *P. axillaris* is unable to synthesize high amounts of trihydroxylated flavonols due to the low affinity of *Petunia* FLS for the precursor dihydromyricetin (Figures 1 and 6B) [54]. Instead, *P. axillaris* synthesizes large quantities of the dihydroxylated quercetin (Figures 1 and 6B). This implies that the reversal to bee pollination involved redirection of the biosynthetic flux back to the trihydroxylated branch of the pathway. In *Petunia*, F3′H is encoded by *HT1* (further referred to as *F3′H/HT1*) and *F3′5′H* is encoded by *HF1* and *HF2* (further referred to as *F3′5′H/HF1* and *F3′5′H/HF2*) [55]. In the purple species, *F3′H/HT1* is lowly expressed, whereas *F3′5′H/HF1/HF2* are highly expressed (Figure 4A; Table S3), which is consistent with the synthesis of trihydroxylated malvidin and petunidin in these species (Figures 6A and 6C). In *P. axillaris*, this pattern is reversed: expression of *F3′H/HT1* is very high and expression of the *F3′5′H/HF* genes is low, consistent with the presence of the dihydroxylated quercetin. In the *P. axillaris* × *P. inflata* F_1, as well as the *P. axillaris* × *P. secreta* F_1, the *F3′H/HT1* gene is expressed from both parental alleles, indicating *trans* activation, presumably by the *P. axillaris* copy of *MYB-FL*. In the *P. axillaris* × *P. inflata* F_1, 93% of the *F3′5′H/HF1* transcripts and 80% of the *F3′5′H/HF2* transcripts originate from the *P. inflata* allele, indicating that, in addition to the lack of *trans* activation by AN2, *cis*-regulatory mutations must be present in the *P. axillaris* copies of the genes (Figure 4B). Note that the sequence identity between the *F3′5′H/HF* loci of *P. axillaris* and *P. secreta* prevented the estimation of allele-specific expression in their F_1 hybrid, so we are unable to determine whether this difference is due to an effect in *cis* or *trans* (Figure 4C). Therefore, the change in

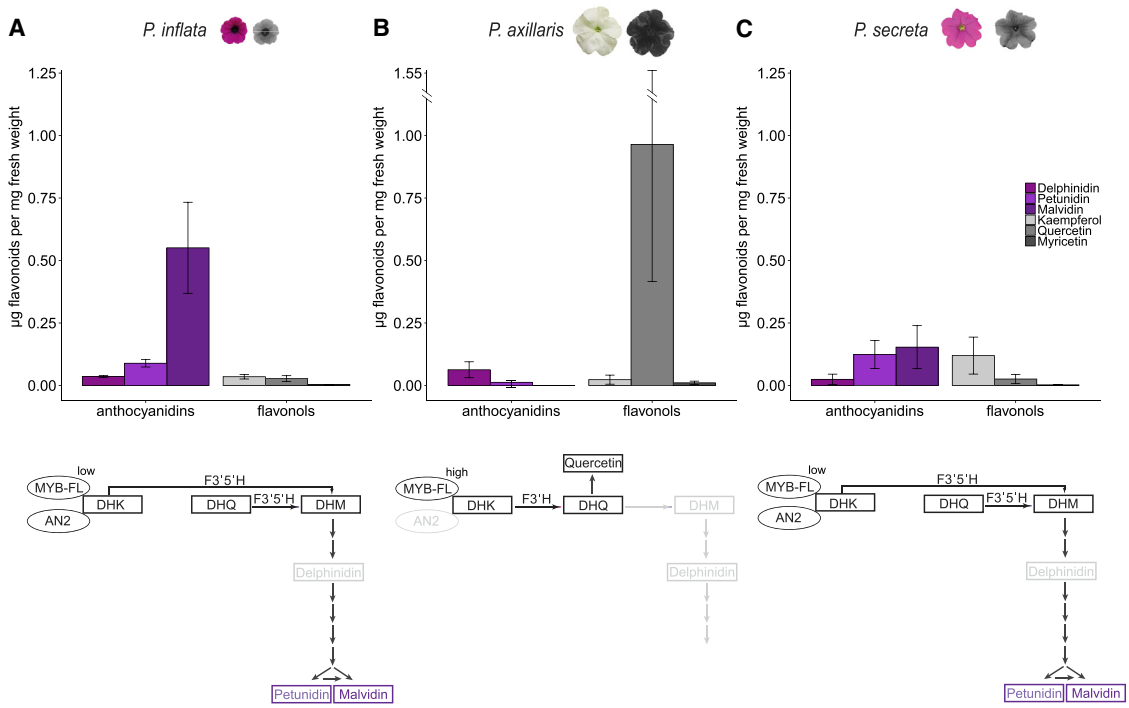

Figure 6. Summary Flavonoid Pathway Fluxes after AN2 Inactivation and Resurrection in *Petunia*

Simplified flavonoid pathway (see also Figure 1) for each species explaining the different flower colors found for the species of the genus *Petunia*. Anthocyanidins (visible color pigments) are as follows: delphinidin; petunidin; and malvidin. Flavonols (UV color pigments) are as follows: kaempferol; quercetin; and myricetin. Flavonoid extracts were separated by LC-MS (see STAR Methods). Error bars represent ±SD.

(A) In *P. inflata*, the combination of active AN2, low expression of *MYB-FL* and *F3′H/HT1*, and high expression of the *F3′5′H/HF* genes direct flavonoid biosynthesis toward the purple petunidin and malvidin.

(B) In *P. axillaris*, the combination of inactive AN2, highly expressed *MYB-FL* and *F3′H/HT1*, as well as low expression of *F3′5′H/HF* directs flavonoid synthesis toward the flavonol quercetin.

(C) The resurrection of *AN2* in *P. secreta* restores the pathway as it operates in *P. inflata*.

See also Table S7.

hydroxylation status was brought about by *MYB-FL*-dependent activation of *F3′H/HT1* and strong reduction of the expression of the two *F3′5′H/HF* genes due to lack of *AN2* activity as well as potential *cis*-regulatory changes in the genes themselves. The reversal to bee pollination in *P. secreta* therefore required the inactivation of *MYB-FL* together with the resurrection of *AN2*, causing loss of flavonol synthesis and the reactivation of the synthesis of trihydroxylated anthocyanins.

DISCUSSION

The in-depth knowledge of the biochemistry and genetics of the flavonoid pathway offers unique opportunities to study the evolution of flower color and, in a broader context, the evolution of pollination syndromes. Although all species of the *Petunia* short-tube clade have the ancestral purple flower color and are pollinated by bees, three different pollination syndromes are present in the closely related species of the long-tube clade. Our phylogeny of the genus *Petunia* clarifies relationships within the long-tube clade, with the white *P. axillaris* being sister to the colored *P. secreta* and *P. exserta* (Figures 2A and 2B).

Knowing the evolutionary importance of AN2 in the loss of pigmentation in *P. axillaris* [13, 26], we used a much larger dataset to confirm that all *P. axillaris* accessions contain *AN2* sequences with nonsense or frameshift mutations leading to truncated proteins (Table S2). Introduction of AN2 into *P. axillaris* causes the synthesis of petunidin and malvidin, indicating that the structural genes in the pathway are intact (Figure 3D). Induced mutagenesis experiments in *P. hybrida* have uncovered loss of color mutations in many structural and regulatory genes of the anthocyanin pathway [56]. This rules out that *AN2* is a preferred target for mutagenesis per se and suggests that the evolutionary preference for *AN2* mutations represents a case of optimal pleiotropy [57].

In *P. axillaris*, the synthesis of large quantities of the flavonol quercetin required the upregulation of *MYB-FL* [16]. In addition, our data indicate a severe *cis* reduction of the expression of the two *F3′5′H/HF* genes and a strong upregulation of *HT1*, presumably by MYB-FL. Genetic evidence from *P. hybrida* indicates that *F3′5′H/HF1* is epistatic over *F3′5′H/HF2* and *F3′H/HT1* [55]. Moreover, *Petunia* FLS has a low affinity for the trihydroxylated precursors but efficiently accepts the dihydroxylated precursors [54]. Thus, the hydroxylation status of the precursors is important for the relative abundance of anthocyanins versus flavonols. Therefore, the switch from anthocyanins to flavonols during the evolution of hawkmoth pollination in *Petunia* proceeded through

mutations in *AN2* and *MYB-FL* and most likely the *F3′5′H/HF* genes as well.

P. secreta presents an outstanding case of trait re-gain during reversal to bee pollination. At the same time, it is important to note that, although color and bee visitation have been restored, *P. secreta* retains the typical long-tube floral morphology, which limits bee visitors to collecting pollen. Pollinator observations found that *P. secreta* attracts small halictid bees (Apoidea and Halictidae; as opposed to short-tube species' solitary bee pollinators of the Andrenidae and Colletidae), who indeed collect pollen rather than nectar [58].

A transposon insertion in the *MYB-FL* gene reduces the competing flavonol biosynthesis (Figures 1, 6, and S5), but, most interestingly, we found a functional *AN2* in *P. secreta.* Based on comparative *AN2* sequence analysis, ancestral state reconstruction, and the structure of the *AN2* haplotype network, we conclude that the functional H11 haplotype evolved independently within the long-tube clade. Note that the reference accession *P. axillaris* N used here contains AN2 haplotype H7 (Figure 3B), which is not only inactive but also very lowly expressed (Figure 4A). The expression levels of the *P. axillaris AN2* haplotypes vary between accessions [13] and members of the non-functional H10 clade do express *AN2* at appreciable levels [13]. Therefore, the *P. secreta AN2* allele most likely derived from a non-functional but expressed H10 haplotype. In summary, the functional *P. secreta AN2* is a long-tube allele that arose through a compensatory 2-bp deletion, a rare case of pseudogene resurrection. The occurrence of this 2-bp deletion in the close vicinity of the 1-bp deletion may represent a rare event, but it is also biochemically simple and most likely happened in a single step. Similarly, the inactivation of *MYB-FL* by transposon insertion must have been a single event. Thus, the color reversal in *P. secreta* did not proceed through the gradual accumulation of multiple independent small-effect mutations in these two genes but through one to two mutations with large phenotypic effects on color. Even though flower color is often considered to be a "simple" trait, flavonoid biosynthesis is not simple because it involves substrate competition and specificities at its branching points. Anthocyanins and flavonols also serve as abiotic stress protectants, male fertility factors, and inhibitors of polar auxin transport [21]. Their synthesis is intertwined with the synthesis of lignins, phytoalexins, and benzenoid volatiles, compounds with essential functions in cell wall structure, defense, and pollinator attraction, respectively [21]. The complexity of such a biochemical network puts severe constraints on network flexibility and restricts the mutational potential via selection [59, 60]. Our data add to the accumulating evidence that major changes in complex traits can happen through few mutations of major effect. Such mutations are expected when the phenotype is displaced far from its adaptive optimum [61], a scenario that may apply particularly well during pollinator shifts or more generally when caused by biotic selection as compared to abiotic selection [62].

In 1893, the paleontologist Louis Dollo postulated that evolution is not easily reversed in that an organism cannot return to a previous state [5]. Only a few cases in support of this have been reported [63, 64], causing a systematic re-thinking of Dollo's laws and focusing instead on trait irreversibility [65]. Interpreting this second of Dollo's laws in modern terms [6], the resurrection of a functional *AN2* in *P. secreta* can be considered a return to a previous character state, as evidenced by the almost identical flavonoid gene expression patterns and color profiles of *P. inflata* and *P. secreta*. However, *P. secreta AN2* evolved in the long-tube clade from an AN2 pseudogene, and because the long-tube clade itself is young, Dollo's law becomes less strictly applicable. In fact, it has been suggested that, for groups undergoing recent adaptive radiations, seemingly lost traits may "flicker" on and off, resulting in a discordant distribution of character states across the phylogenetic tree, with reactivation events especially likely in an evolutionary timescale of 0.5–6 MYA [66]. The reacquisition of AN2 functionality in the *P. secreta* H11 allele works in this adjusted framework with the adaptive radiation of *Petunia* dated between 1.3 [60] and 2.85 MYA [61]. Consequently, the divergence of H11 and H10 alleles within the long-tube clade must have happened after the adaptive radiation and before the present. It follows that the focus of recent literature, and especially of the flower color literature, has shifted from permanent losses of traits to the degree of irreversibility [65, 67, 68]. As such, the pseudogenization of *AN2* and its subsequent resurrection is an example of trait and gene function reversibility.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Color and UV images
  - DNA/RNA manipulations
  - Transgenic complementation
  - Virus induced gene silencing
  - Genotyping
  - Quantitative RT-PCR
  - RNA sequencing
  - Read processing
  - Allele-specific expression analyses
  - P. secreta MYB-FL transposon
  - Reconstruction of species relationships
  - AN2 gene trees
  - AN2 haplotype ancestral state reconstruction
  - AN2 haplotype network analysis
  - UPLC-UV-MS quantification of flavonoids
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and seven tables and can be found with this article online at https://doi.org/10.1016/j.cub.2018.10.019.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation and the NCCR "Plant Survival" to C.K., the Brazilian National Council for Scientific and Technological Development (Science without Borders, MCTIC/CNPq) to

L.F., and the Swiss National Science Foundation "R'EQUIP" no. 157884. We thank Therese Mandel, Lea Jäggi, Theres Imhoff, Christopher Ball, Jasmin Sekulovski, Roman Köpfli, Peter von Ballmoos, and Avichai Amrad for technical support; Gina Cannarozzi and Rémy Bruggmann for bioinformatics advice; and Matthias Erb, Christelle Robert, and Jean Daniel Berset for biochemical support. Computationally demanding processes were performed on the HPC cluster of the University of Bern (http://www.id.unibe.ch/hpc).

AUTHOR CONTRIBUTIONS

K.E., A.E.B., M.M., and E.B. performed the experiments and analyzed the data, and L.F. and C.K. supervised the project. K.E., A.E.B., M.M., and C.K. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 31, 2018
Revised: September 5, 2018
Accepted: October 5, 2018
Published: November 21, 2018

REFERENCES

1. Manceau, M., Domingues, V.S., Linnen, C.R., Rosenblum, E.B., and Hoekstra, H.E. (2010). Convergence in pigmentation at multiple levels: mutations, genes and function. Philos. Trans. R. Soc. Lond. B Biol. Sci. *365*, 2439–2450.
2. Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Jr., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science *327*, 302–305.
3. Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. Nat. Rev. Genet. *14*, 807–820.
4. Rausher, M.D. (2008). Evolutionary transitions in floral color. Int. J. Plant Sci. *169*, 7–21.
5. Dollo, L. (1893). Les lois de l'évolution. Bulletin de la Société belge de Géologie, de Paléontologie et D'hydrologie *7*, 164–166.
6. Gould, S.J. (1970). Dollo on Dollo's law: irreversibility and the status of evolutionary laws. J. Hist. Biol. *3*, 189–212.
7. van der Niet, T., and Johnson, S.D. (2012). Phylogenetic evidence for pollinator-driven diversification of angiosperms. Trends Ecol. Evol. *27*, 353–361.
8. Whittall, J.B., Voelckel, C., Kliebenstein, D.J., and Hodges, S.A. (2006). Convergence, constraint and the role of gene expression during adaptive radiation: floral anthocyanins in *Aquilegia*. Mol. Ecol. *15*, 4645–4657.
9. Wessinger, C.A., and Rausher, M.D. (2014). Predictability and irreversibility of genetic changes associated with flower color evolution in Penstemon barbatus. Evolution *68*, 1058–1070.
10. Schemske, D.W., and Bradshaw, H.D., Jr. (1999). Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). Proc. Natl. Acad. Sci. USA *96*, 11910–11915.
11. Smith, S.D., Miller, R.E., Otto, S.P., Fitzjohn, R.G., and Rausher, M.D. (2010). The effects of flower color transitions on diversification rates in Morning Glories (Ipomoea subg. Quamoclit, Convolvulaceae). In Darwin's Heritage Today: Proceedings of the Darwin 200 Beijing International Conference, M. Long, H. Gu, and Z. Zhou, eds. (Higher Education Press), pp. 202–226.
12. Bradshaw, H.D., and Schemske, D.W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. Nature *426*, 176–178.
13. Hoballah, M.E., Gübitz, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., Dell'Olivo, A., Arnold, M., and Kuhlemeier, C. (2007). Single gene-mediated shift in pollinator attraction in *Petunia*. Plant Cell *19*, 779–790.
14. Hopkins, R., and Rausher, M.D. (2011). Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. Nature *469*, 411–414.
15. Klahre, U., Gurba, A., Hermann, K., Saxenhofer, M., Bossolini, E., Guerin, P.M., and Kuhlemeier, C. (2011). Pollinator choice in *Petunia* depends on two major genetic Loci for floral scent production. Curr. Biol. *21*, 730–739.
16. Sheehan, H., Moser, M., Klahre, U., Esfeld, K., Dell'Olivo, A., Mandel, T., Metzger, S., Vandenbussche, M., Freitas, L., and Kuhlemeier, C. (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.
17. Rodríguez-Gironés, M.A., and Santamaría, L. (2004). Why are so many bird flowers red? PLoS Biol. *2*, e350.
18. Fenster, C.B., Armbruster, W.S., Wilson, P., Dudash, M.R., and Thompson, J.D. (2004). Pollination syndromes and floral specialization. Annu. Rev. Ecol. Evol. Syst. *35*, 375–403.
19. Chittka, L., and Raine, N.E. (2006). Recognition of flowers by pollinators. Curr. Opin. Plant Biol. *9*, 428–435.
20. Sasaki, K., and Takahashi, T. (2002). A flavonoid from *Brassica rapa* flower as the UV-absorbing nectar guide. Phytochemistry *61*, 339–343.
21. Winkel, B.S.J. (2006). The biosynthesis of flavonoids. In The Science of Flavonoids, E. Grotewold, ed. (Springer New York), pp. 71–95.
22. Davies, K.M., Albert, N.W., and Schwinn, K.E. (2012). From landing lights to mimicry: the molecular regulation of flower colouration and mechanisms for pigmentation patterning. Funct. Plant Biol. *39*, 619–638.
23. Quattrocchio, F., Baudry, A., Lepiniec, L., and Grotewold, E. (2006). The regulation of flavonoid biosynthesis. In The Science of Flavonoids, E. Grotewold, ed. (Springer), pp. 97–122.
24. Stracke, R., Jahns, O., Keck, M., Tohge, T., Niehaus, K., Fernie, A.R., and Weisshaar, B. (2010). Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in *Arabidopsis thaliana* plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. New Phytol. *188*, 985–1000.
25. Albert, N.W., Davies, K.M., Lewis, D.H., Zhang, H., Montefiori, M., Brendolise, C., Boase, M.R., Ngo, H., Jameson, P.E., and Schwinn, K.E. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. Plant Cell *26*, 962–980.
26. Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J., and Koes, R. (1999). Molecular analysis of the *anthocyanin2* gene of *petunia* and its role in the evolution of flower color. Plant Cell *11*, 1433–1444.
27. Durbin, M.L., Lundy, K.E., Morrell, P.L., Torres-Martinez, C.L., and Clegg, M.T. (2003). Genes that determine flower color: the role of regulatory changes in the evolution of phenotypic adaptations. Mol. Phylogenet. Evol. *29*, 507–518.
28. Schwinn, K., Venail, J., Shang, Y., Mackay, S., Alm, V., Butelli, E., Oyama, R., Bailey, P., Davies, K., and Martin, C. (2006). A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. Plant Cell *18*, 831–851.
29. Streisfeld, M.A., and Rausher, M.D. (2009). Altered trans-regulatory control of gene expression in multiple anthocyanin genes contributes to adaptive flower color evolution in *Mimulus aurantiacus*. Mol. Biol. Evol. *26*, 433–444.
30. Yuan, Y.W., Sagawa, J.M., Young, R.C., Christensen, B.J., and Bradshaw, H.D., Jr. (2013). Genetic dissection of a major anthocyanin QTL contributing to pollinator-mediated reproductive isolation between sister species of *Mimulus*. Genetics *194*, 255–263.
31. Hoshino, A., Morita, Y., Choi, J.D., Saito, N., Toki, K., Tanaka, Y., and Iida, S. (2003). Spontaneous mutations of the flavonoid 3′-hydroxylase gene conferring reddish flowers in the three morning glory species. Plant Cell Physiol. *44*, 990–1001.

32. Zufall, R.A., and Rausher, M.D. (2004). Genetic changes associated with floral adaptation restrict future evolutionary potential. Nature *428*, 847–850.
33. Streisfeld, M.A., and Rausher, M.D. (2009). Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. New Phytol. *183*, 751–763.
34. Des Marais, D.L., and Rausher, M.D. (2010). Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in Ipomoea. Evolution *64*, 2044–2054.
35. Smith, S.D., and Rausher, M.D. (2011). Gene loss and parallel evolution contribute to species difference in flower color. Mol. Biol. Evol. *28*, 2799–2810.
36. Wessinger, C.A., and Rausher, M.D. (2012). Lessons from flower colour evolution on targets of selection. J. Exp. Bot. *63*, 5741–5749.
37. Lorenz-Lemke, A.P., Togni, P.D., Mäder, G., Kriedt, R.A., Stehmann, J.R., Salzano, F.M., Bonatto, S.L., and Freitas, L.B. (2010). Diversification of plant species in a subtropical region of eastern South American highlands: a phylogeographic perspective on native *Petunia* (Solanaceae). Mol. Ecol. *19*, 5240–5251.
38. Särkinen, T., Bohs, L., Olmstead, R.G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol. Biol. *13*, 214.
39. Reck-Kortmann, M., Silva-Arias, G.A., Segatto, A.L., Mäder, G., Bonatto, S.L., and de Freitas, L.B. (2014). Multilocus phylogeny reconstruction: new insights into the evolutionary history of the genus *Petunia*. Mol. Phylogenet. Evol. *81*, 19–28.
40. Segatto, A.L.A., Cazé, A.L., Turchetto, C., Klahre, U., Kuhlemeier, C., Bonatto, S.L., and Freitas, L.B. (2014). Nuclear and plastid markers reveal the persistence of genetic identity: a new perspective on the evolutionary history of *Petunia exserta*. Mol. Phylogenet. Evol. *70*, 504–512.
41. Turchetto, C., Fagundes, N.J., Segatto, A.L., Kuhlemeier, C., Solís Neffa, V.G., Speranza, P.R., Bonatto, S.L., and Freitas, L.B. (2014). Diversification in the South American Pampas: the genetic and morphological variation of the widespread *Petunia axillaris* complex (Solanaceae). Mol. Ecol. *23*, 374–389.
42. Dell'olivo, A., Hoballah, M.E., Gübitz, T., and Kuhlemeier, C. (2011). Isolation barriers between petunia axillaris and Petunia integrifolia (Solanaceae). Evolution *65*, 1979–1991.
43. Grotewold, E. (2006). The genetics and biochemistry of floral pigments. Annu. Rev. Plant Biol. *57*, 761–780.
44. Rausher, M.D., Miller, R.E., and Tiffin, P. (1999). Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol. Biol. Evol. *16*, 266–274.
45. Kimura, M. (1983). The Neutral Theory of Molecular Evolution (Cambridge University Press).
46. Ehrenreich, I.M., and Purugganan, M.D. (2006). The molecular genetic basis of plant adaptation. Am. J. Bot. *93*, 953–962.
47. Dittmar, E.L., Oakley, C.G., Conner, J.K., Gould, B.A., and Schemske, D.W. (2016). Factors influencing the effect size distribution of adaptive substitutions. Proc. Biol. Sci. *283*, 20153065.
48. Dean, A.M., and Thornton, J.W. (2007). Mechanistic approaches to the study of evolution: the functional synthesis. Nat. Rev. Genet. *8*, 675–688.
49. Stern, D.L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? Evolution *62*, 2155–2177.
50. Edwards, S.V., Potter, S., Schmitt, C.J., Bragg, J.G., and Moritz, C. (2016). Reticulation, divergence, and the phylogeography-phylogenetics continuum. Proc. Natl. Acad. Sci. USA *113*, 8025–8032.
51. Mallet, J., Besansky, N., and Hahn, M.W. (2016). How reticulated are species? BioEssays *38*, 140–149.
52. Snowden, K.C., Simkin, A.J., Janssen, B.J., Templeton, K.R., Loucas, H.M., Simons, J.L., Karunairetnam, S., Gleave, A.P., Clark, D.G., and Klee, H.J. (2005). The decreased apical dominance1/*Petunia hybrida* CAROTENOID CLEAVAGE DIOXYGENASE8 gene affects branch production and plays a role in leaf senescence, root growth, and flower development. Plant Cell *17*, 746–759.
53. Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. J. Exp. Bot. *62*, 2465–2483.
54. Holton, T.A., Brugliera, F., and Tanaka, Y. (1993). Cloning and expression of flavonol synthase from *Petunia hybrida*. Plant J. *4*, 1003–1010.
55. Wiering, H., de Vlaming, P., Schram, A.W., Jonsson, L.M.V., and Bennink, G.J.H. (1984). Inheritance and biochemistry of pigments. In Petunia, K.C. Sink, ed. (Springer Berlin Heidelberg), pp. 49–76.
56. van Houwelingen, A., Souer, E., Spelt, K., Kloos, D., Mol, J., and Koes, R. (1998). Analysis of flower pigmentation mutants generated by random transposon mutagenesis in *Petunia hybrida*. Plant J. *13*, 39–50.
57. Martin, A., and Orgogozo, V. (2013). The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution *67*, 1235–1250.
58. Rodrigues, D.M., Caballero-Villalobos, L., Turchetto, C., Jacques, R.A., Kuhlemeier, C., and Freitas, L.B. (2018). Do we truly understand pollination syndromes in Petunia as much as we suppose? AoB PLANTS *ply057*. Published online October 1, 2018. https://doi.org/10.1093/aobpla/ply057.
59. Wagner, A. (2009). Evolutionary constraints permeate large metabolic networks. BMC Evol. Biol. *9*, 231.
60. Rausher, M.D., Lu, Y., and Meyer, K. (2008). Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. J. Mol. Evol. *67*, 137–144.
61. Orr, H.A. (2005). The genetic theory of adaptation: a brief history. Nat. Rev. Genet. *6*, 119–127.
62. Louthan, A.M., and Kay, K.M. (2011). Comparing the adaptive landscape across trait types: larger QTL effect size in traits under biotic selection. BMC Evol. Biol. *11*, 60.
63. Bekpen, C., Marques-Bonet, T., Alkan, C., Antonacci, F., Leogrande, M.B., Ventura, M., Kidd, J.M., Siswara, P., Howard, J.C., and Eichler, E.E. (2009). Death and resurrection of the human IRGM gene. PLoS Genet. *5*, e1000403.
64. Kitano, T., Blancher, A., and Saitou, N. (2012). The functional A allele was resurrected via recombination in the human ABO blood group gene. Mol. Biol. Evol. *29*, 1791–1796.
65. Collin, R., and Miglietta, M.P. (2008). Reversing opinions on Dollo's Law. Trends Ecol. Evol. *23*, 602–609.
66. Marshall, C.R., Raff, E.C., and Raff, R.A. (1994). Dollo's law and the death and resurrection of genes. Proc. Natl. Acad. Sci. USA *91*, 12283–12287.
67. Wessinger, C.A., and Rausher, M.D. (2015). Ecological transition predictably associated with gene degeneration. Mol. Biol. Evol. *32*, 347–354.
68. Smith, S.D., and Goldberg, E.E. (2015). Tempo and mode of flower color evolution. Am. J. Bot. *102*, 1014–1025.
69. Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C.S., Bliek, M., Boersma, M.R., Borghi, L., Bruggmann, R., et al. (2016). Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. Nat. Plants *2*, 16074.
70. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 11.10.1–11.10.33.
71. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.
72. Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873–881.

73. Castel, S.E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. Genome Biol. *16*, 195.
74. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.
75. Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. *35*, 543–548.
76. Keller, O., Kollmar, M., Stanke, M., and Waack, S. (2011). A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics *27*, 757–763.
77. Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. *41*, e121.
78. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.
79. Sela, I., Ashkenazy, H., Katoh, K., and Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. *43* (W1), W7–W14.
80. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
81. Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics *30*, i541–i548.
82. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19* (*Suppl 6*), 153.
83. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods *9*, 772.
84. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *10*, e1003537.
85. Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. *29*, 1969–1973.
86. Drummond, A.J., Ho, S.Y.W., Phillips, M.J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. PLoS Biol. *4*, e88.
87. Rambaut, A., Suchard, M.A., Xie, D., and Drummond, A.J. (2013). Tracer v1.6.0. https://github.com/beast-dev/tracer/releases.
88. Rambaut, A. (2016). FigTree v1.4.3: Tree Figure Drawing Tool. http://tree.bio.ed.ac.uk/software/figtree/.
89. Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics *25*, 1451–1452.
90. Bandelt, H.-J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. *16*, 37–48.
91. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
92. Mayba, O., Gilbert, H.N., Liu, J., Haverty, P.M., Jhunjhunwala, S., Jiang, Z., Watanabe, C., and Zhang, Z. (2014). MBASED: allele-specific expression detection in cancer tissues and cell lines. Genome Biol. *15*, 405.
93. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289–290.
94. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223.
95. Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. *8*, 4321–4325.
96. Spitzer-Rimon, B., Cna'ani, A., and Vainstein, A. (2013). Virus-aided gene expression and silencing using TRV for functional analysis of floral scent-related genes. In Virus-Induced Gene Silencing: Methods and Protocols, A. Becker, ed. (Humana Press), pp. 139–148.
97. Tornielli, G., Koes, R., and Quattrocchio, F. (2009). The genetics of flower color. In Petunia: Evolutionary, Developmental and Physiological Genetics, Second Edition, T. Gerats, and J. Strommer, eds. (Springer-Verlag), pp. 269–299.
98. Mallona, I., Lischewski, S., Weiss, J., Hause, B., and Egea-Cortines, M. (2010). Validation of reference genes for quantitative real-time PCR during leaf and flower development in Petunia hybrida. BMC Plant Biol. *10*, 4.
99. Wang, Z., Ye, S., Li, J., Zheng, B., Bao, M., and Ning, G. (2011). Fusion primer and nested integrated PCR (FPNI-PCR): a new high-efficiency strategy for rapid chromosome walking or flanking sequence cloning. BMC Biotechnol. *11*, 109.
100. Scotto-Lavino, E., Du, G., and Frohman, M.A. (2006). 3′ end cDNA amplification using classic RACE. Nat. Protoc. *1*, 2742–2745.
101. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv, arXiv:1207.3907v2, https://arxiv.org/abs/1207.3907.
102. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.
103. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.
104. Heled, J., Bryant, D., and Drummond, A.J. (2013). Simulating gene trees under the multispecies coalescent and time-dependent migration. BMC Evol. Biol. *13*, 44.
105. Berardi, A.E., Hildreth, S.B., Helm, R.F., Winkel, B.S.J., and Smith, S.D. (2016). Evolutionary correlations in flavonoid production across flowers and leaves in the Iochrominae (Solanaceae). Phytochemistry *130*, 119–127.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
DH5-alpha competent *E. coli*	New England Biolabs	C2987I
Agrobacterium tumefaciens strain GV3101	Monsanto Company	N/A
Biological Samples		
Wild *Petunia* accessions are listed in Tables S1 and S2, including *P. secreta*	[16, 40, 41]	N/A
Petunia axillaris N	Rostock Botanical Garden (Germany)	http://www.garten.uni-rostock.de
P. inflata S6	R. Koes (University of Amsterdam, the Netherlands)	N/A
P. exserta	R.J. Griesbach (USDA, Beltsville, USA)	N/A
Chemicals, Peptides, and Recombinant Proteins		
Phusion High-Fidelity DNA polymerase	New England Biolabs	M0530L
Q5 High-Fidelity DNA polymerase	New England Biolabs	M0491L
DNase I	Sigma-Aldrich	AMP-D1
Standard Taq DNA polymerase	New England Biolabs	M0273L
GoTaq DNA polymerase	Promega	M7845
Delphinidin, cyanidin, petunidin, peonidin, malvidin, and pelargonidin chlorides	Extrasynthese	Catalog #, CAS # 0904S, 528-53-0 0909S, 528-58-5 0942, 1429-30-7 0906S, 134-01-0 0913S, 643-84-5 0912S, 134-04-3
Kaempferol, quercetin, myricetin	Extrasynthese	Catalog #, CAS # 1124S, 520-18-3 1135S, 6151-25-3 1127S, 529-44-2
LC-MS grade acetonitrile	Thermo Fisher Scientific	51101
LC-MS grade formic acid	Thermo Fisher Scientific	85178
Critical Commercial Assays		
RNeasy Plant Mini Kit	QIAGEN	74104
NucleoFast 96 PCR Filtration Kit	Macherey-Nagel	743100.10
Transcriptor Universal cDNA Master	Previously Roche, now Sigma	05893151001
KAPA SYBR® FAST qPCR Kit optimized for LightCycler® 480	KAPA Biosystems; Axon Lab	KK4611
Illumina TruSeq PE Cluster Kit v3	Illumina	PE-401-3001
RevertAid First Strand cDNA Synthesis Kit	Thermo Fisher Scientific	K1621
Deposited Data		
P. axillaris N genome v162	SolGenomics Network [69]	https://solgenomics.net/organism/Petunia_axillaris/genome
P. inflata S6 genome v102	SolGenomics Network [69]	https://solgenomics.net/organism/Petunia_inflata/genome
P. inflata S6 RNA-seq Ilumina reads	This paper	SRP159913: SRX4649699, SRX4649698, SRX4649692
P. axillaris N RNA-seq Illumina reads	This paper	SRP159913: SRX4649695, SRX4649694, SRX4649693
P. secreta RNA-seq Illumina reads	This paper	SRP155164
P. axillaris N x *P. inflata* S6 F1 hybrid RNA-seq Illumina reads	This paper	SRP159913: SRX4649697, SRX4649696, SRX4649691
P. axillaris N x *P. secreta* F1 hybrid RNA-seq Illumina reads	This paper	SRP155427: SRX4474307, SRX4474308, SRX4474309
Petunia species AN2 gene sequences (92 sequences)	This paper	MH732768 - MH732859

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
P. secreta MYB-FL gene sequence	This paper	MH732766
Calibrachoa ovalifolia AN2 gene sequence	This paper	MH732767
Oligonucleotides		
A list of oligonucleotides is given in Table S6		N/A
Recombinant DNA		
Tobacco Rattle Virus VIGS constructs	*Arabidopsis* Biological Resource Center at Ohio State University	ABRC stock #s:
pTRV1	*Arabidopsis* Biological Resource Center at Ohio State University	CD3-1039
pTRV2-MCS	*Arabidopsis* Biological Resource Center at Ohio State University	CD3-1040
pTRV2-*NbPDS*	*Arabidopsis* Biological Resource Center at Ohio State University	CD3-1045
pTRV2-*P. secreta AN2*	This paper	N/A
CaMV 35S:AN2$^{P.secreta}$	This paper	N/A
pGEM-T Easy	Promega	A1360
Software and Algorithms		
Geneious 9.0.5	Biomatters	N/A
LightCycler 480 Software (v.1.1.0.1320)	Roche	N/A
GATK	[70, 71]	N/A
GSNAP	[72]	N/A
Illumina Pipeline Software v.1.82	Illumina	N/A
ASEReadCounter	[73]	N/A
BUSCO (v3)	[74, 75]	N/A
Augustus (v3.2.3)	[76]	N/A
Hmmer (v3.1b2)	[77]	N/A
MAFFT (v7.397)	[78]	N/A
Guidance2	[79]	N/A
RAxML (v8.2.10)	[80]	N/A
ASTRAL (v5.6.2)	[81, 82]	N/A
jModelTest2 (v2.1.10)	[83]	N/A
BEAST2 (v2.4.3)	[84–86]	N/A
TRACER (v1.6.0)	[87]	N/A
FigTree (v.1.4.3)	[88]	N/A
DNASP (v5.10.01)	[89]	N/A
PopART (http://popart.otago.ac.nz)	[90]	N/A
Trimmomatic (v.0.36)	[91]	N/A
R (v3.3.3)	http://www.R-project.org/	N/A
R: MBASED	[92]	N/A
R: ape (v4.1)	[93]	N/A
R: phytools (v0.6-0)	[94]	N/A
Other		
Resource website for this publication	This paper	https://github.com/Kuhlemeier-lab/Pseudogenization_and_resurrection_CB

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Cris Kuhlemeier (cris.kuhlemeier@ips.unibe.ch).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Wild accessions have been previously described [16, 40, 41] and are listed in Tables S1 and S2. The reference accession *Petunia axillaris* N (hereafter referred to as *P. axillaris*) is from the Rostock Botanical Garden (Germany); *P. inflata* S6 was provided by R. Koes (University of Amsterdam, the Netherlands); *P. exserta* from R.J. Griesbach (USDA, Beltsville, USA). *P. secreta* was collected in its natural habitat (Tables S1 and S2); a voucher is deposited at the herbarium of the Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. Plants were grown in a growth chamber under a light:dark regime of 15:9 h at 22:17°C in commercial soil (70% Klasman substrate, 15% Seramis clay granules, 15% quartz sand) and fertilized once a week (Nitrogen-Phosphorous-Potassium and Iron).

METHOD DETAILS

Color and UV images

Color images were recorded using a Panasonic DMC-TZ10 camera. UV pictures were recorded using a Nikon 60 mm 2.8D microlens with a Nikon D7000 SLR camera converted to record UV light using a UV-specific filter (blocking visible and infrared light).

DNA/RNA manipulations

Genomic DNA extraction was performed with a modified CTAB method [95]; except for wild accessions, which was performed as described previously [40]. For RNA extractions, all tissue was homogenized in liquid nitrogen. Extraction was carried out with the RNeasy Plant Mini Kit (QIAGEN) replacing β-mercaptoethanol by dithiothreitol. DNA and RNA were quantified with a Nanodrop ND-1000 (Thermo Fisher); RNA was additionally analyzed on a Bioanalyzer 2100 (Agilent Technologies) prior to cDNA preparation.

Transgenic complementation

Transgenic complementation was performed as described [13] using the *AN2* gene of *P. secreta* with its introns, under the control of the *CaMV 35S* promoter. Three positive transgenic plants were obtained, that displayed similar pink floral phenotypes with *AN2* being expressed.

Virus induced gene silencing

VIGS was performed as described [96] using *Agrobacterium tumefaciens* strain GV3101 transformed with pTRV1 and pTRV2 derivatives (pTRV1 (CD3-1039), pTRV2 carrying a multiple cloning site (CD3-1040)). To generate pTRV2 carrying the *P. secreta AN2* gene, 334 bp of *AN2* were amplified via PCR of floral cDNA. This fragment of *AN2* is located at the 3′ end of the gene, which is more specific, to reduce off-target effects. BamHI and EcoRI sites added during PCR are shown in italics in the forward and reverse primers, respectively, and are listed in Table S6. Four plants were infiltrated with pTRV2-*P. secreta-AN2*; three infiltration buffer control plants and two PDS control plants (PDS: phytoene desaturase from *Nicotiana benthamiana*, which causes photobleaching and is used as a positive control for VIGS efficiency) were done. pTRV2-*P. secreta-AN2* infiltrated plants showed in at least 70% of the flowers the white spots on the otherwise purple flowers, whereas neither controls showed any flower-specific phenotypes.

Genotyping

AN2 sequences were amplified from genomic DNA either with Phusion polymerase (NEB) or Q5 polymerase (NEB) with primers designed on the 5′ and 3′ UTRs of the gene (listed in Table S6). PCR condition followed the instruction of the supplier with an annealing temperature of 59°C and 1:30 min elongation in 35 cycles. Likewise, primers used to amplify *Calibrachoa ovalifolia* as outgroup (GenBank: MH732767) are listed in Table S6. Full-length AN2 sequences for gene tree reconstruction were obtained for a subset of the accessions (Table S1) either by cloning (pGEM-T Easy kit, Promega) and Sanger sequencing (Microsynth, Balgach, Switzerland) using primers listed in Table S6 or by sequencing with both amplification primers without prior cloning (GATC, Cologne, Germany; Table S6). Since for most sequences we only wanted to verify the diagnostic position at nucleotide 380 (Table S2; see also Figure 3B) the majority of the amplification products were purified with the NucleoFast Kit (Machery-Nagel) and sequenced using only the reverse amplification primer (Table S6) to reveal this diagnostic position. All received sequences were manually inspected and aligned to existing AN2 sequences in Geneious 9.0.5 (Biomatters). Intron and exon borders were annotated by aligning genomic DNA to cDNA, the latter obtained from previous work [13]. For each sequence listed in Table S1, NCBI GenBank accessions are given.

Quantitative RT-PCR

In addition to the transcription factors *MYB-FL* and *AN2*, all known biosynthetic genes of the flavonoid pathway [97] were investigated for their expression level. Primer sequences, including the one for the reference gene *SAND* [98], are listed in Table S6 and were derived from aligned sequences of the three species using sequence information available on NCBI or solgenomics.net. Each time one bud (stage 4) was collected from each of three different plants of each species representing three biological replicates (n = 12); *P. axillaris* and *P. secreta* (bud length: 22-30 mm), *P. inflata* (9-13 mm). Samples were immediately frozen in liquid nitrogen and stored at −80°C until extraction. Corolla limb tissue was used for RNA extraction, extracted RNA treated with DNase I (Sigma-Aldrich), and subsequently quality measurements done on a 2100 Bioanalyzer (Agilent) discarding samples with a RIN value

below 8.5. First strand synthesis was performed with random hexamer primers using the Transcriptor Universal cDNA Master (Roche) kit according to the supplier. Quantitative RT-PCR experiments were performed using the LightCycler 96 Real-Time PCR System (Roche) with KAPA SYBR® FAST qPCR Kit optimized for LightCycler 480 (KAPA Biosystems), according to the manufacturer's recommendations. Reactions were run in triplicate for each data point, means of these technical replicates were taken as the value for the biological replicate. Cycle of quantification (Cq) thresholds and normalization calculations were determined by the LightCycler 480 Software (v.1.1.0.1320; Roche). Standard curves were used to determine the PCR efficiencies. The *SAND* reference gene and controls with no-reverse transcription were included for each sample. For each of the flavonoid genes of interest, pairwise comparisons of the mean relative gene expression values were made among the three species in R (v3.3.3, http://www.R-project.org/) using the TukeyHSD function.

RNA sequencing

To detect allele-specific expression (ASE) in *P. axillaris* x *P. inflata* F_1 and *P. axillaris* x *P. secreta* F_1 hybrids, three biological replicates were used for each parental line and the respective F_1 (SRP159913, SRP155164, SRP155427). Limb tissue was harvested at stage 4 buds (*P. axillaris*/*P. secreta*: 22-30 mm, *P. inflata*: 10-15 mm, *P. axillaris* x *P. inflata* F_1: 17.5-23.5 mm, and *P. axillaris* x *P. secreta* F_1: 22-30 mm) from plants grown under controlled conditions. RNA was sequenced in the Lausanne Genomic Technologies Facility (Lausanne, Switzerland). Quality of RNA was checked using a Fragment Analyzer (Advanced Analytical). RNA Quality Numbers (RQN) of the samples ranged from 7.7 to 9.6. Libraries were prepared using Illumina TruSeq PE Cluster Kit v3 and sequenced on a total of six lanes on Illumina HiSeq 2500 as single-end 100nt reads. Sequencing data was processed using the Illumina Pipeline Software v.1.82.

Read processing

All read data was processed with Trimmomatic v.0.36 [91] to remove Illumina adaptor sequences and trim low-quality regions. After trimming, reads shorter than 60 bp were discarded. These pre-processed reads were mapped against the draft reference genome of *P. axillaris* (v.1.6.2 [69],) using the SNP-tolerant and splice-aware aligner GSNAP (v.2016-11-07 [72, 73],). To allow for extended variation between reference and reads, the "–m" option was set to 0.04 for samples other than *P. axillaris*.

Allele-specific expression analyses

Variants were detected with GATK according to GATK Best Practices for RNA-seq data [70]. In short, after duplicate marking and splitting reads with N in their CIGAR string, local realignment around indels was undertaken and base quality scores were recalibrated, using a set of high quality SNPs determined by an initial run of the HaplotypeCaller [71]. Allelic coverage for variant positions was detected with ASEReadCounter implemented in GATK [71, 73]. Analyses of allelic imbalance were conducted in R (v3.3.3, http://www.R-project.org/) with the package MBASED (v.1.2.0) [92]. Parameters of read count over-dispersion were estimated with a custom R script provided by the author of the MBASED package. P values of ASE were corrected for multiple comparisons using the p.adjust function with the "BH" method in R base package "stats" (v3.3.3, http://www.R-project.org/).

P. secreta MYB-FL transposon

Standard PCR amplification of the 3′ end of the fourth exon of the *MYB-FL* gene in *P. secreta* using primers previously described in [16] initially failed, so chromosome walking via fusion primer and nested integrated PCR (FPNI-PCR) was used, described in [99]. The forward primers used in the three sequential PCR amplifications were B367, B404, and B406 (Table S6), PCR protocols were followed according to [99] with Standard *Taq* polymerase (NEB). Single-band products were Sanger sequenced (Microsynth, Balgach, Switzerland), and aligned to *P. axillaris, P. exserta,* and *P. inflata MYB-FL* sequences (GenBank: KT962949, KT962950, KT962951, respectively). Sequences were searched against NCBI GenBank. One product (696 bp) yielded 591 bp with high homology to exon 4 of *P. axillaris, P. exserta,* and *P. inflata* MYB-FL loci, but the remaining 135 bp in the product showed 96% homology to the *Dad-one* transposon found in *P. hybrida* (AY745251 [52]). To confirm the presence of this transposon via PCR, reverse primer B636 (Table S6) was designed inside the transposon to be amplified with *MYB-FL* forward primer (Table S6).

To determine whether the transposon is transcribed in mRNA, a modified 3′ RACE technique was used [100]. Limb tissue from two stage 4 buds (22-30 mm) was harvested from *P. secreta*. RNA was extracted and DNase I treated as previously described in the methods for Quantitative RT-PCR. cDNA was synthesized with RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) with the GeneRacer (Thermo Fisher Scientific) oligo-dT primer (Table S6). PCR was performed using forward primer B367 and a reverse primer complimentary to the oligo-dT, GRnested (Table S6). Nested PCR was performed on a 1:50 dilution of these products with forward primer B408 (Table S6), which binds 73 bp upstream of the functional stop codon of *P. axillaris* MYB-FL, and GRnested using GoTaq DNA polymerase (Promega). Conditions for both rounds of PCR followed the instructions of the supplier with an annealing temperature of 57.5°C, and 1:30 min elongation in 35 cycles. The resulting product was ~180 bp, and Sanger sequence yielded 132 bp of sequence that connected the exon 4 of MYB-FL to the *Dad-one* transposon. This configuration of at least a section of the *Dad-one* transposon insertion into the exon 4 of *P. secreta* MYB-FL is further confirmed by a matching contig containing the insertion from RNA-seq analysis of *P. secreta*, as well as from a draft genome assembly of *P. secreta* (Figure S5).

Reconstruction of species relationships

Genome-wide phylogenetic analysis based on SNPs was conducted from RNA-seq reads of open-flower tissue of *P. axillaris*, *P. exserta*, *P. inflata*, and *P. secreta* (SRA: SRP155164, PRJNA344560). Per species, RNA-seq reads from three biological replicates were concatenated and aligned against the *P. axillaris* genome and processed as described above in Read Processing. SNP detection was carried out using freebayes (v.1.1.0 [101],) with parameters "-no-mnps -no-complex -m 10 –min-coverage 6," obtaining a total of 931,089 variant sites.

Bi-allelic, homozygous sites were used in a second round of variant calling using HaplotypeCaller [71] in "EMIT_ALL_SITES" mode. Emitted base calls of all four species were filtered for missing data resulting in 399,673 concatenated sites. Tree construction was performed with 40,758 polymorphic sites using the maximum likelihood method in IQ-TREE v1.5.2 [102]. After modeltesting with ModelFinder [103], options "-m TEST+ASC+R -bb 1000 -alrt1000 -b 500" were used for tree construction.

To complement the SNP-based maximum likelihood approach, single copy orthologs were extracted using the BUSCO v3 pipeline ([74, 75]; NCBI BLAST 2.6.0+, Augustus v3.2.3 with starting species "tomato" [76], HMMER v.3.1b2 [77]) from the genomes of *P. axillaris* and *P. inflata* (both from [69], available from the Sol Genomics Network https://solgenomics.net/), draft genomes of *P. exserta* and *P. secreta,* and a draft transcriptome from *Calibrachoa pygmea* as an outgroup. Of the complete and single copy orthologs retrieved using BUSCO, 984 genes were shared among the five taxa. Amino acid sequences were aligned with MAFFT using the Guidance2 algorithm [78, 79]. Individual gene trees based on amino acid sequences were generated using a maximum likelihood approach, using the P-threads version of RAxML (raxmlHPC; v8.2.10 [80]) with automatic protein substitution model assignment algorithm PROTGAMMAAUTO and 100 rapid bootstraps. From these gene trees, a genome-scale coalescent-based species tree was estimated using ASTRAL (v5.6.2), which is suited to inferring a species tree from multiple gene trees while accounting for gene tree discordance [81, 82]. Multi-locus bootstrapping provided local posterior probabilities for ancestral nodes. The resulting tree is as follows: (PAX,((PEX,PSE)1:0.09024687784777805,(CPY,PIN)1:1.053454529509685));) with branch lengths in coalescent units. Approximately 62.2% of the induced quartet trees from the gene trees (3061 out of 4920) could be found in the species tree. As the ASTRAL coalescent tree agreed with the topology of the SNP-based method, we only displayed the latter in the text.

AN2 gene trees

AN2 gene trees were constructed using the genomic or derived coding sequences (cds) with *Calibrachoa ovalifolia* (GenBank: MH732767) as an outgroup. Phylogenetic reconstruction was done with the subset of the data that comprise the full-length sequences including several accessions from each haplotype (Table S1). Missing data were coded as question marks. Model selection for both alignments was performed with AIC, BIC, and DT selection criteria using jModelTest2 v2.1.10 [83]. All three methods selected the HKY85 model with a gamma model of rate heterogeneity (HKY85+G) as the optimal model. Phylogenetic trees for the genomic and cds alignments were constructed in BEAST2 v2.4.3 [84–86], specifically using the *BEAST multispecies coalescent model that estimates gene trees simultaneously with the species tree; we only used the resulting gene trees for our analyses [104]. *BEAST was run using the HKY substitution model with four gamma categories, a birth-death prior, 10^8 MCMC chains sampled every 1000 generations, and a strict clock, as described in [39]. Identical runs with different priors (yule, calibrated yule, coalescent models) all resolved the same general topology of haplotypes. Partitions were used for introns and exons in the genomic analyses. The first 20% of all trees were discarded as burn-in, as determined by examination of effective sample size values (ESS > 200) and likelihood plots compared to an identical run with 10^6 MCMC chains in the program TRACER 1.6.0 [87]. The treeannotator package provided in BEAST2 was used to generate a maximum clade credibility tree with mean node heights. Trees and posterior probabilities were drawn in FigTree v.1.4.3 [88].

AN2 haplotype ancestral state reconstruction

The evolution of haplotypes was estimated using R (v3.3.3) and the packages phytools (v0.6-0 [94];) and ape (v4.1 [93];), following the protocol described on the phytools website for discrete characters (http://www.phytools.org/eqg2015/asr.html). Since all short-tube clade species have one haplotype, they were collapsed into a single, representative tip. Next, the *AN2* gene tree described above was made ultrametric using the *chronopl* function from the ape package, using 2.85 MYA as the minimum age of the *Petunia* clade [38]. All haplotypes were coded as discrete characters. Three character transition models (equal rates, ER, 1 parameter; all-rates-different, ARD, 132 parameters; symmetric, SYM, 66 parameters) were compared using a log-likelihood ratio test, where the ER model was the best fit.

Marginal ancestral states of internal/ancestral nodes were estimated with the continuous-time Markov chain model, or the Mk model, with the ace function in the ape R package. Stochastic character mapping was then used to sample the posterior probabilities of ancestral nodes using the phytools package *make.simmap* function, with 1000 stochastic character map simulations, and were summarized using the *describe.simmap* function. Resulting ancestral state reconstructions from the two methods were highly correlated (mean pairwise Spearman correlation, r = 0.54). Therefore we present results from the stochastic character mapping, as it samples unambiguous histories for discrete character evolution.

AN2 haplotype network analysis

The AN2 sequence alignment of the coding sequence was used. Individual haplotypes were identified and phased using DNASP 5.10.01 [89]; sites with gaps were considered and invariable sites removed. The evolutionary relationships among haplotypes were estimated with PopART (http://popart.otago.ac.nz) using the median-joining network method and $\varepsilon = 0$ [90]. According to the median joining method, PopART assigns ambiguous sites (including indels) to the setting of the most common definite state of the alignment with ties being broken arbitrarily. This method does not explicitly handle indels, and so haplotypes distinguished by indels (e.g., H11) are not separated into distinct nodes unless alternate coding techniques are used. Thus, to incorporate indels all gaps were coded as Ts (nucleotide chosen arbitrarily) to force their inclusion, and re-ran the analysis as a companion haplotype network to the previous analysis.

UPLC-UV-MS quantification of flavonoids

Flavonoid extraction

Fresh limb tissue of three individuals per species (*P. inflata*, *P. axillaris, P. secreta, P. exserta*; total n = 12) was weighed and placed overnight in 1 mL of 2M HCl in a darkened 2 mL screw-cap tube. Flavonoid extraction was performed via acid hydrolysis and phase separation, resulting in two fractions per sample: one containing flavonols and the other containing anthocyanidins. The extraction procedure is previously described [105], with a modified acid hydrolysis time of 90 minutes. Final extracts were dried at 4°C and stored at −20°C. Samples were re-suspended in 75% MeOH with 1% formic acid (v/v) and the two fractions per sample were combined for LC-MS analysis, except for *P. axillaris* samples which were analyzed separately due to the disparate concentrations of flavonols and anthocyanidins.

LC-MS-UV detection

Waters Aquity I-Class UPLC system was used with a binary solvent manager, Sample Manager-FL autosampler, eLambda Detector 800 nm photodiode array, and QDa Detector (single quadrupole mass spectrometer with ESI). All data was visualized using MassLynx v4.1 SCN888 and peaks were integrated with the included TargetLynx. Samples were cooled at 4°C before injection in the autosampler, and 5 μL of each sample was injected onto an ACQUITY UPLC® BEH C18 1.7 μm column kept at 40°C with a flow rate of 0.2 mL/min. Solvent A = H_2O: acetonitrile: formic acid 98.9: 1: 0.1; Solvent B = Acetonitrile: formic acid 99.9: 0.1. A 23-minute gradient with linear and nonlinear ramps (exponential increases represented as Waters gradient curve numbers) was established with the following conditions: 0-2 min 99% A, linear; 2-10 min 85% A, curve 7; 10-13 min 80% A, curve 9; 13-19 min 40% A, linear; 19-21 min hold at 0% A, 21.1-23 min 99% A, linear. Peaks were detected using a coupled Waters Acquity PDA (model: UPLC eLambda 800 nm) at 520 nm (anthocyanidins) and 365 nm (flavonols), scanning from 190 to 790 nm at a resolution of 1.2 nm and a sampling rate of 20 points/sec. Mass spectrometry with a Waters QDa detector coupled to the UPLC system was used to verify compound identity, in addition to retention time and UV spectra. Flavonoids were detected in positive electrospray ionization mode (m/z 150-400) from 0-21 minutes, with data collection in centroid mode, scan time 0.195 s, capillary voltage 0.80 kV, cone voltage 10V, ion source temperature 120°C and probe temperature 600°C. As the pre-selected ionization parameters caused only little fragmentation, only the precursor ions were considered ([M+H]+ in ESI+ mode) for compound identification.

Compound identification and quantification

Standard compounds (delphinidin, cyanidin, petunidin, peonidin, malvidin, and pelargonidin chlorides, kaempferol, quercetin, and myricetin; Extrasynthese, Genay, France) were prepared into a dilution series for compound quantitation. Flavonoid compounds were identified by comparison to standard compound retention time, UV spectra, and mass (Table S7). Means and standard deviations per compound were calculated per species by averaging biological replicates. Additionally, total anthocyanidins and total flavonols were calculated per species by summing the amounts of each class of compound. For *P. exserta*, only total anthocyanidins and total flavonols values were calculated.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistics were calculated in R (v3.3.3). All measured data are presented as means ± SD or SE and specified along with sample sizes (n) in the methods and in figure legends. Comparisons between groups for the analysis of qRT-PCR was performed with Tukey's Honest Significance Differences, and significance levels are marked as: * $p < 0.05$, ** $p < 0.01$, **** $p < 0.001$. Gene-level measures of allele specific expression (ASE) were performed with the MBASED R-package [92], which used a one-sample analysis that tests whether a single haplotype is over- or under-represented relative to the other haplotype based on allelic counts at individual single nucleotide variations (described in greater detail in the MBASED package vignette). P values represent a significant deviation from a 50:50 haplotype representation over the gene, and were additionally corrected for multiple comparisons (number of comparisons = number of genes in dataset) using the Benjamini & Hochberg/False Discovery Rate of 5% method.

DATA AND SOFTWARE AVAILABILITY

The accession numbers for the sequenced genes reported in this paper are GenBank: MH732766–MH732859.

The accession numbers for the sequence reads reported in this paper are SRA: SRP159913, SRP155164, SRP155427.

Phylogenetic trees, AN2 gene alignments, and R-scripts with allele-specific expression analysis can be found on the GitHub repository (https://github.com/Kuhlemeier-lab/Pseudogenization_and_resurrection_CB).

Primers are listed in Table S6.

All software and algorithms are available from the references listed in the Key Resources Table (section ''Software and Algorithms'').