bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Computational modeling of anthocyanin pathway evolution: Biases, hotspots, and trade-offs

Lucas C. Wheeler,[1,*] Stacey D. Smith[1]

1.Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

* lucas.wheeler@colorado.edu

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Alteration of metabolic pathways is a common mechanism underlying the evolution of new phenotypes. Flower color is a striking example of the importance of metabolic evolution in a complex phenotype, wherein shifts in the activity of the underlying pathway lead to a wide range of pigments. Although experimental work has identified common classes of mutations responsible for transitions among colors, we lack a unifying model that relates pathway function and activity to the evolution of distinct pigment phenotypes. One challenge in creating such a model is the branching structure of pigment pathways, which may lead to evolutionary trade-offs due to competition for shared substrates. In order to predict the effects of shifts in enzyme function and activity on pigment production, we created a simple kinetic model of a major plant pigmentation pathway: the anthocyanin pathway. This model describes the production of the three classes of blue, purple and red anthocyanin pigments, and accordingly, includes multiple branches and substrate competition. We first studied the general behavior of this model using a naïve set of parameters. We then stochastically evolved the pathway toward a defined optimum and analyzed the patterns of fixed mutations. This approach allowed us to quantify the probability density of trajectories through pathway state space and identify the types and number of changes. Finally, we examined whether our simulated results qualitatively align with experimental observations, i.e., the predominance of mutations which change color by altering the function of branching genes in the pathway. These analyses provide a theoretical framework that can be used to predict the consequences of new mutations in terms of both pigment phenotypes and pleiotropic effects.

Keywords

flavonoids, enzymes, evolutionary trajectories, pleiotropy, simulations, metabolic control

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Many complex phenotypes evolve through changes to the activity of metabolic pathways (Morrison and Badyaev 2016; Nam et al. 2011; Pichersky and Gang 2000; Schuster et al. 2000; Strohman 2002). In addition to forming the basis for the extraction and transfer of energy within organisms, metabolism contributes to phenotypes by producing essential cellular products such as structural components, toxins, and pigments (Chen et al. 2014; Colón et al. 2010; Flowers et al. 2007; Keller et al. 2005; Morrison and Badyaev 2017). Evolution of metabolic pathways is known to be shaped by both the topological structure of pathways and the biochemical constraints of individual enzymes (Dykhuizen et al. 1987; Morrison and Badyaev 2017; Nam et al. 2011). Furthermore, the regulatory architecture governing pathway gene expression plays a vital role in the evolution of pathway activity and the resulting alterations in phenotype (Braakman et al. 2017; Gompel et al. 2005; Morrison and Badyaev 2017; Panettieri et al. 2018; Reed and Serfas 2004; True et al. 1999; Wittkopp et al. 2003; Zhang et al. 2017). These properties make metabolic pathways excellent systems to study the fundamental principles that underlie the genotype-phenotype map for complex phenotypes.

A large body of theoretical work has been devoted to understanding the control of metabolic pathways. Metabolic control analysis (MCA) arose from classical enzyme kinetics and solved the problem of analyzing entire metabolic systems simultaneously, rather than focusing on individual components in isolation (Cornish-Bowden 1995; Heinrich and Schuster 1996; Kacser et al. 1995). These methods have been widely used to understand the behavior of metabolic pathways in both basic and applied research (Chen et al. 2014; Colón et al. 2010; Hoefnagel et al. 2002; Schuster et al. 1999, 2000; Stephanopoulos 1999). Still, few studies have used MCA to investigate the evolution of metabolic pathways and the resulting phenotypes, and these have focused on simple pathways with little or no branching (Heckmann et al. 2013; Rausher 2013; Wright and Rausher 2010). While the results point to predictable evolutionary patterns, such as the concentration of fixed mutations in enzymes with the most control over pathway flux, it remains unclear whether these 'rules' apply to the large and highly branched metabolic pathways common in biological

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

systems. Such complex branched topologies are likely to generate trade-offs due to competition for substrates and enzymes (Dekel and Alon 2005; Dykhuizen et al. 1987), and thus may introduce constraints on evolutionary trajectories (Guillaume and Otto 2012).

Although theory has not extensively examined the evolution of phenotypes arising from branching pathways, empirical studies can provide insight into the types of mutations that contribute to evolutionary change. Evolutionary studies on experimental systems, such as carotenoid metabolism in birds (Morrison and Badyaev 2016, 2018), various metabolic pathways in yeast (Wisecaver et al. 2014) and Drosophila (Flowers et al. 2007), and anthocyanin biosynthesis in flowering plants (Smith and Rausher 2011; Streisfeld and Rausher 2009a,b), have demonstrated that causal mutations can be drawn from several broad classes, such as regulatory vs. biochemical (structural) and upstream vs. downstream. Depending on the genetic architecture, these classes tend to be differentially favored during evolutionary trajectories, leading to detectable genomic "hotspots" for phenotypic evolution (Cooper et al. 2003; Cresko et al. 2004; Esfeld et al. 2018; Martin and Orgogozo 2013; Treves et al. 1998; Woods et al. 2006). Pleiotropic effects have commonly been suggested as an explanation for these patterns (Gompel et al. 2005; Streisfeld and Rausher 2009b; Wessinger and Rausher 2014) although other factors, such as target size and chromosomal location, may also play a role (Arbeithuber et al. 2015). Determining how pathway structure gives rise to these predictable patterns is critical for understanding the evolution of the phenotypes that result from metabolic activity (Kuang et al. 2018; Wessinger and Rausher 2012b, 2014).

Flower color has proven to be a particularly useful model for studying the origin of novel phenotypes through metabolic evolution. Coloration in plants is primarily determined by the presence of various pigment compounds, the most common of which are produced by the highly conserved anthocyanin pathway (Fig. 1, S1). This pathway is part of the larger flavonoid biosynthesis pathway, which produces an array of related flavonoid compounds, that includes the anthocyanins and flavonols. These compounds play roles as pigments, sunscreens, stress response molecules, and act as substrates for downstream reactions (Winkel-Shirley 2001). The network of enzymatic reactions of the anthocyanin pathway gives rise to three principal classes of red, purple, and blue pigments (the monohydroxylated pelargonidin, dihydroxylated cyanidin and trihydroxylated delphinidin, re-

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

spectively; Fig. 1, S1). Further downstream modifications of these basic building blocks result in the diversity of colorful anthocyanin pigments found across the flowering plants.

Building on existing knowledge of this pathway, experimental work spanning several decades has identified genetic hotspots for flower color evolution. For example, shifts among the three types of anthocyanins consistently target two branching enzymes (F3'H and F3'5'H), whose activity determines the relative production of three anthocyanin precursors (DHM, DHQ, DHK, Fig.) (Streisfeld and Rausher 2011; Wessinger and Rausher 2012b). These changes may involve structural mutations that alter enzyme function (Wessinger and Rausher 2014), regulatory changes that shift expression (Des Marais and Rausher 2010), or both (Smith and Rausher 2011). The observation of repeated fixation of mutations at certain loci suggests a level of predictability in flower color evolution that may be explained by the underlying pathway structure. At a broad phylogenetic scale, this genetic architecture may lead to evolutionary trends, such as the rarity of some flower colors (Ng and Smith 2018) and the highly asymmetric transition rates among phenotypes (Tripp and Manos 2008). Nevertheless, testing such a relationship requires a theoretical framework that unites the topology of the pathway with phenotypic evolution via changes in the function or expression of pathway enzymes.

The present study takes the first step toward such a unifying framework by creating a mathematical model of the anthocyanin pathway based on classic enzyme kinetics along with a computational pipeline for simulating evolution under directional selection. We use this framework to address four key evolutionary questions: 1) What is the predicted steady-state activity of the anthocyanin pathway based on its topology? 2) How does the pathway control (i.e., which enzymes determine flux) change during evolution toward a new phenotypic optimum? 3) Are certain loci and types of mutations predictably involved in phenotypic evolution? 4) Do trade-offs arising from pathway structure constrain the evolution of pathway components? We hypothesize that the evolution of a new color phenotype will require shifts in pathway control, accomplished by increasing or decreasing the activity of pathway reactions. Accordingly, mutations at loci with the greatest control are predicted to be the greatest contributors to phenotypic transitions (Rausher 2013; Wright and Rausher 2010). However, because of reticulations in the pathway created by the branching enzymes, we also expect

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that these transitions may lead to trade-offs within and between branches of the pathway that are under selection. The results of these simulation analyses will not only lay the foundation for continued theoretical advances, but will also generate concrete, quantitative predictions that can be tested in empirical systems.

Methods

Development and computational implementation of the mathematical pathway model

We used the generalized rate law formulation of (Chou and Talaly 1977; Schäuble et al. 2013) to specify rate laws for each enzymatic reaction in the pathway model (Supplemental text). This formulation scales the Michaelis constant (K_M, which relates to binding affinity) for each enzyme-substrate pair by those of competing enzymes to explicitly incorporate substrate competition. In the absence of substrate competition, the rate law reduces to classic Michaels-Menten kinetics (Chou and Talaly 1977; Schäuble et al. 2013). We chose to use irreversible Michaelis-Menten kinetics, because *in vivo* measurements of flavonoid pathway dynamics have been described well by an irreversible model (Groenenboom et al. 2013), but see (Halbwirth 2010; Halbwirth et al. 2006). Moreover, the irreversible form of the model reduces the required number of parameters two-fold relative to the completely reversible equivalent (Schäuble et al. 2013).

Given that detailed kinetic studies are lacking for most anthocyanin pathway enzymes, we designed a naïve model as a starting state to learn about the relative importance of different model parameters and the interplay between them. We chose values for catalytic constants and Michaelis constants from a meta-analysis of enzyme properties in the BRENDA and KEGG databases (Bar-Even et al. 2011; Kanehisa 2008; Schomburg et al. 2002). The model was initialized at a starting state with all parameters of a certain type set to be equal. Values for K_{cat} (the catalytic rate constant) were set to 14 s^{-1}, the median of the distribution calculated in Bar-Even et al. 2011. K_M values were set to 0.013 mM, one tenth the median value of the empirical distribution to

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ensure that starting conditions of the model put the upstream substrate concentration well above the K_M of the first enzyme and all subsequent enzymes. Enzyme concentrations were initialized to 0.001 mM, a value within the range of naturally-occurring protein concentrations (Miranda et al. 2008). This model represents a naïve starting state, with no preference for any substrates, allowing the "null" unbiased behavior of the pathway topology to be studied. Boundary conditions of the model were imposed by using an upstream source that flows into the pathway at a constant concentration (0.01 mM) and sinks at the end of all branches. The sinks are encoded as simple mass action processes with a single rate constant k_{sink} (0.0005 M^{-1}) that is the same for all sinks. This process represents the diffusion of the pathway products away from the volume in which they are synthesized, which is consistent with the physiological transport of anthocyanin pigments to the vacuole (Petrussa et al. 2013; Poustka et al. 2007). We implemented this mathematical model of the anthocyanin pathway in Python 3.6 using the Tellurium library (Choi et al. 2016) and the Antimony (Smith et al. 2009) markdown language (see detailed description in supplemental text).

Evolutionary simulations between defined phenotypic states

We developed a custom Python library, called *enzo* (https://github.com/lcwheeler/enzo), to conduct stochastic evolutionary simulations of the pathway (see Supplemental text for details). Evolutionary simulations are performed using the core PathwaySet object of *enzo*, which acts as a wrapper for the Tellurium pathway model. Within the simulation framework, we randomly select a single parameter from the model at each step and introduce a mutation with effect size randomly sampled from a gamma distribution. Fixation events are determined based on a formula that weights mutations by fitness effect size (see Supplemental text).

Given the structure of the pathway, mutations during these evolutionary simulations can alter any of the enzyme parameters (K_M, K_{cat}, and enzyme concentration E_t). While these simulated mutations are not modeled at the level of DNA substitutions, they have a close connection with the kinds of variation found in natural systems. The K_M and K_{cat} for each enzyme are determined by its biophysical properties, which in turn are tied to its protein sequence (Bar-Even et al. 2011; Kaltenbach and Tokuriki 2014; Khersonsky and Tawfik 2010). By contrast, enzyme concentration

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is controlled by regulation of gene expression and protein degradation (Cooper 2000), and in the case of the anthocyanin pathway, regulation takes place largely at the level of transcription (Quattrocchio et al. 2006). Thus, changes in K_M and K_{cat} of anthocyanin pathway enzymes likely occur through mutations in enzyme coding sequences (Johnson et al. 2001; Shimada et al. 2001) while changes in enzyme concentration evolve through mutations in cis-regulatory regions or trans-regulatory loci (Sobel and Streisfeld 2013).

In our simulations, each enzyme has a single E_t parameter (reflecting the shared volume of all pathway components in simulations), but can have multiple K_M and K_{cat} values, depending on the number of substrates. Changes in these parameters occur without correlated effects on other parameters. For example, DFR (Fig. 1) can experience a mutation that improves its activity on one substrate without changing activity on other substrates, consistent with the different K_M and K_{cat} values for different substrates that have been measured experimentally (Fischer et al. 2003; Katsu et al. 2017; Miyagawa et al. 2015). While it might be biologically reasonable to assume some cost of improvement on one substrate for activity on other substrates (Kaltenbach and Tokuriki 2014; Khersonsky and Tawfik 2010), we surveyed the kinetic data for all multi-substrate anthocyanin pathway enzymes in the BRENDA database (Schomburg et al. 2002) and were unable to detect any consistent patterns of correlated changes, positive or negative. Nevertheless, our model could be expanded to include correlated changes in future.

A total of 9,985 independent evolutionary simulations were run between the naïve starting state and a new optimum, which we selected to be a blue-flowered state. We set the optimum as 90% of the total steady state concentration comprising delphinidin, with a 10% tolerance. All other concentrations were allowed to drift, subject only to the global constraint placed on total steady state concentration of all species, which was held to within a 10% tolerance of that from the initial starting state. These tolerance values are flexible arguments to the *evolve* function in *enzo* and can be modified to suit the needs of the user.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Analysis of the simulated data

We designed a series of analyses to address the four questions outlined in the introduction. First, in order to characterize the intrinsic behavior of the pathway, we used metabolic control analysis (MCA), which was developed specifically for analyzing the control of flux through metabolic pathways (Cornish-Bowden 1995; Heinrich and Schuster 1996; Kacser et al. 1995). The standard approach in MCA is to calculate two properties for each pathway reaction: 1) the flux and/or concentration control coefficients, which are the partial derivatives of each flux/steady state concentration with respect to enzyme activity, and 2) elasticities, which are sensitivity coefficients calculated by taking the partial derivative of each reaction rate with respect to each substrate concentration. Together, these measures allow us to determine how pathway enzymes differ in their control over concentrations of floating species (substrates and products) and how reactions vary in their sensitivity to substrate concentration. We used the MCA tools in Tellurium (Choi et al. 2016) to calculate control coefficients and elasticities for our naïve model to assess how the structure of the pathway relates to the steady-state pigment production. We applied the same tools in addressing our second question regarding the changes in pathway control during evolution to a new optimum. In this case, we repeated the MCA analyses on each evolved model and subtracted starting-point control and elasticity matrices from the median-optimal matrices in order to quantify the shift in control.

We next used the simulation data to identify the loci and the types of changes that contributed to evolution toward the delphinidin optimum. For each of our simulations, we counted the number of fixation events involving each model parameter (e.g., enzyme concentration, E_t, for DFR) during the trajectories. Our *enzo* package also logged the selection coefficients for each mutation introduced during the simulations, and these were summarized across all trajectories to estimate the distribution of fitness effects (DFE) for each model parameter.

Finally, we used the simulations to assess potential evolutionary trade-offs arising from the pathway structure. We first calculated mean Spearman correlation coefficients (ρ) between all pairs of floating species as well as all pairs of model parameters (E_t, K_M, K_{cat} values) across all of

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the simulations. We also compared the distribution of parameter values after the evolutionary simulation with the starting state in order to visualize these correlated changes in pathway activity.

Results

Mathematical model reveals behavior of anthocyanin pathway topology

Our naïve starting model, with all enzymes at equal concentrations and with equal kinetic parameters, yielded stable time-course dynamics (Fig. 2a) and resulted in a steady state with species concentrations reaching into the low mM range (Fig. 2b). The steady state is characterized by a bias toward products requiring the fewest biochemical reactions (kaempferol and pelargonidin), resulting from the greater accumulation of these products early in the time-course (Fig. 2a). The colored pigments pelargonidin, cyanidin, and delphinidin comprise 32.7% , 10.9% , and 6.6% of the total steady state concentration, respectively (Fig. 2b). This chemical profile would likely result in red or pink floral coloration (Ng and Smith 2016a; Smith and Rausher 2011).

Metabolic control analysis of this naïve model indicated that pathway control is concentrated in the upstream step of the pathway and in the branching enzymes that interact with multiple floating species. The first pathway enzyme, CHS, exerts strong positive control over all downstream reactions and steady state concentrations (Fig. 3a), consistent with previous theoretical studies (Rausher 2013; Wright and Rausher 2010). The control of the branching enzymes varies across products in a manner reflecting their position in the pathway. For example, F3'H has positive control over cyanidin and delphinidin production, which both require its activity, and negative control over pelargonidin, which does not require it (Fig. 3a). Similar patterns of contrasting positive and negative control are seen for other branching enzymes, including F3'5'H, DFR, and FLS. The final enzyme in the pathway (ANS) exhibits little control over any floating species except its substrates, a pattern consistent with its downstream position. The elasticity patterns were similarly intuitive based on pathway structure. For example, CHI and F3H are highly sensitive to the concentrations of their substrates. We also observed negative values due to competition, where, for instance, the reactions involving substrates that compete with DHK (DHQ, DHM, Fig. 1; Fig.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S1) are negatively sensitive to DHK concentration (Fig. 3b).

Evolution towards delphinidin shifts pathway control

Our evolutionary simulations revealed that the evolutionary transition to delphinidin production required the gain of pathway control by branches leading to or competing with delphinidin. Based on MCA analyses of the 9,985 trajectories that succeeded in reaching 90% delphinidin without error (see Supplemental Results), we observed that F3'H and F3'5'H, both required for delphinidin production, sharply increased in their control coefficients over all downstream substrates and products while the competing enzyme that leads to flavonol production (FLS) lost control over pathway flux (Fig. 3c,e). DFR shifted its activity to gain control over the flux of DHM (the delphinidin precursor), while DFR reactions involving competing substrates (DHQ and DHK) lost control over DHM flux. We found concordant changes in elasticities, with these enzymes becoming more sensitive to the concentration of delphinidin precursors (DHM, LCD) and less sensitive to the competing substrates (Fig. 3d,f). These patterns suggest that while the most upstream gene controls overall flux, the downstream branching enzymes control the type of pigment that is made.

Loci with greater pathway control are mutational targets

As predicted, we found that approximately 99% of mutations fixed during the simulations involved F3'H, F3'5'H, DFR, or FLS (Table S1), the loci with strong control over delphinidin production. These mutations are spread across all classes of parameters (K_M, K_{cat}, and E_t), (Fig. S4, Table S1), but for some loci, are concentrated on particular parameters (Table S1 and Supplemental text). Overall, mutations affecting all parameters for the branching enzymes show similar distributions of fitness effects (DFE), from highly negative ($s = -0.55$) to highly positive ($s = 0.45$) (Fig. S3). By contrast, the other pathway enzymes show narrow DFEs skewed towards negative fitness effects (Fig. S3), reflecting that these loci have little control on delphinidin production and that mutations away from the original parameter values are largely detrimental. Across the 9,985 simulations, the median selection coefficient for fixed mutations was 0.044 (range, $s = 2.4x10^{-6}$ to 0.43), with 90% delphinidin production achieved with a median of 9 mutational steps ($range = 3$ to 80; Fig. 4c).

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Enzyme and substrate competition result in pleiotropic tradeoffs

Evolution toward the delphinidin optimum resulted in sharp and predictable trade-offs in pathway function and activity. Increased delphinidin production came at the cost of decreased production of other pathway products, including the other two anthocyanidins and the three flavonols (Fig. 4d,e, Fig. S2). These effects varied by the number of shared steps, such that compounds requiring the fewest shared steps (namely pelargonidin and kaempferol) showed the strongest decreases. We also observed trade-offs across multiple enzymatic parameters for the loci under selection (F3'H, F3'5'H, DFR, FLS; Fig. S5). For example, the K_{cat} values for DFR acting on the blue precursor DHM were substantially improved while those for competing substrates (DHQ, DHK) were drastically weakened, resulting in negative correlations (Fig. S5d).

Beyond these antagonistic relationships, we found cases of positive correlations in line with selection toward delphinidin. For example, parameters affecting the same reactions (e.g. K_{cat} and K_M for DFR acting on DHK) are often positively correlated, whether they are coordinately increasing (because they contribute to delphidinin production) or coordinately decreasing (because they move flux way from delphinidin) (Fig. S5). Similarly, we observed positive correlations between delphinidin production and its two immediate precursors (DHM and LCD) (Fig. S5). Such positive correlations were also seen for other products (pelargonidin, cyanidin) and their precursors (DHK and LCP; DHQ and LCC, respectively). Overall, these patterns reveal how selection on a single pathway product has reverberating effects across all products and precursors due to the changes in amount and activity of their shared enzymes.

Discussion

In this study, we used a computational approach to probe the fundamental rules governing evolution of floral pigmentation. We constructed a mathematical model of the anthocyanin biosynthesis pathway, which produces a suite of red, purple, and blue pigments. Evolving this model from an unbiased starting state toward a blue-pigmented state revealed the pathway components that are capable of shifting flux between pigment types. These components, in particular the concentration

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and activity of branching enzymes, were not only the primary mutational targets in our simulations, but also constitute the evolutionary hotspots for transitions among flower color phenotypes in natural systems. We also observed strong trade-offs inherent to the branched pathway topology, which result in predictable shifts in pathway control that are likely to constrain evolution of floral pigmentation. As discussed below, these results have direct implications for connecting color phenotypes with metabolic pathway function and for understanding the evolution of flower color more broadly.

Intrinsic bias toward red coloration

Our naïve starting state model uncovered an inherent bias toward the production of red pelargonidin pigments, an interesting finding given that they are not the predominant pigments in natural systems. Since this naïve model was based on equal concentrations and equal kinetic parameters for all enzymes, we infer that the tendency towards pelargonidin production (at nearly two-fold the amount of the other two pigments) relates to the topology of the pathway. Specifically, the production of pelargonidin requires five enzymatic steps while the others, cyanidin and delphinidin, require six and seven steps, respectively. This topological bias results in the early accumulation of pelargonidin in the time-course of the model. Consistent with this notion, a similar pattern is observed in the flavonols, where the mono-hydroxylated counterpart of pelargonidin (kaempferol) is produced in greater quantities than the other two flavonols (Fig. 4d,e).

Although we lack an angiosperm-wide dataset on floral pigments, multiple surveys indicate that pelargonidin pigments are generally rare compared to the other classes of anthocyanins. For example, in their survey of 530 species, Beale et al. (Beale et al. 1941) estimated the proportions of taxa containing pelargonidin, cyanidin, and delphinidin-based floral pigments to be 19%, 40%, and 50%, respectively. In a more extreme example, studies in the tomato family suggest that as few as 0.3% of the roughly 2800 extant species produce pelargonidin (Ng and Smith 2016b; Ng et al. 2018). This pattern is not easily explained by genetic mechanisms, as the loss-of-function mutations which often lead to red flowers (Smith and Rausher 2011; Wessinger and Rausher 2015) would instead favor an accumulation of species producing pelargonidin. A more viable explanation

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is selection driving species away from the pelargonidin-pigmented state, possibly due to pleiotropic effects on the production of the dihydroxylated flavonol quercetin, which is important for UVB protection (Ryan et al. 1998, 2002). By developing a quantitative model linking pathway function to both classes of flavonoids, this study will provide new avenues for testing such hypotheses in future.

Hotspots of color evolution

Shifts in anthocyanin production, and in turn, flower color, are commonly observed across angiosperms, making these transitions a valuable system for studying the predictability of evolution at the molecular level (Streisfeld et al. 2011; Wessinger and Rausher 2012b). A growing suite of studies has examined evolutionary transitions between flower colorphenotypes, and these implicate two branching enzymes, F3'H and F3'5'H, as the key players determining the type of anthocyanin produced (Des Marais and Rausher 2010; Smith and Rausher 2011; Streisfeld and Rausher 2009b; Wessinger and Rausher 2015). For example, the downregulation of F3'h (with 30-fold to 1000-fold lower expression) was responsible for a switch from blue to red pigmentation in at least four evolutionary cases (Wessinger and Rausher 2012b, 2014). Indeed, a survey of the literature to-date (Table S3) reveals that all fifteen natural examples of transitions between pigment typesinvolved F3'H, F3'5'H or both, while two of these involved mutations in DFR as well. Likewise fifteen of the seventeen examples of engineered pigmentation shifts involved F3'H and/orF3'5'H (Table S3), making thesethe major genetic targets for artificially modifying flower hue(Tanaka Yoshikazu and Brugliera Filippa 2013).Our evolutionary simulations from a red, mostly pelargonidin starting-state to a blue, mostly delphinidin ending-state recovered a similar targeting of changes to F3'H and F3'5'H (Fig. 4e; Table S1; Fig. S4). Mutations at these loci ranged from those altering kinetic properties (enzyme function) to those altering enzyme concentration (analogous to expression) (Fig. 4e, Fig. S4). We observed a mean 5.3-fold increase in F3'H concentration, which combined with a mean 5.3-fold increase in the K_{cat} for DHK and a mean 1.04-fold tightening of the K_M for DHK, yielded an overall 29-fold mean increase in F3'H activity (Fig. 4; Fig. S4). Changes in F3'5'H activity followed a similar pattern (Fig. 4; Fig. S4). The action of these genes as hotspots for color

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

evolution in nature and in our simulationsmakes sense because, while the upstream loci control flux across the entire pathway, F3'H and F3'5'H control the relative production of the three pigment precursors (DHK, DHQ, and DHM). The preferential fixation of mutations at F3'H and F3'5'H is also in line with theoretical work showing disproportionate occurrence of large-effect beneficial mutations at enzymes with the most control (Fig. 3, (Rausher 2013)). Overall, the striking similarity between these empirical and theoretical studies points to a predictable relationship between the structure of the pathway and loci that underlie phenotypic evolution (Wessinger and Rausher 2012a).

In addition to these major shifts in F3'H and F3'5'H expression and function, we found fixed changes at two other loci, DFR and FLS, which have been implicated less often in natural systems (Table S3). In accordance with selection for increased flux toward delphinidin, DFR improved its activity on the precursor (DHM) roughly 4-fold (on average) during the simulations. Such a matching of flux through the pathway with enzyme activity has been documented in the *Iochroma* system, where DFR from the red species prefers the red precursor DHK while DFR from the blue species prefers DHM (Smith et al. 2013). In this case, the evolution of DFR to track the primary precursor (as determined by F3'H and F3'5'H activity) was interpreted as selection for optimal pathway flux (Smith and Rausher 2011; Smith et al. 2013). Our simulations also recovered fixed mutations that reduced the activity of FLS and directed flux away from flavonol production. Although this gene has been shown to influence floral anthocyanin production in one empirical case (Yuan et al. 2016), its general importance as a locus of color evolution remains poorly understood.

Evolutionary trade-offs in color production

One of the primary motivations for this study was to understand how the structure of the anthocyanin pathway relates to the phenotypic space for flower color and the mechanisms for moving through that space. Previous surveys have revealed gaps in this color space, i.e. phenotypes that are theoretically possible but not observed in nature (Ng et al. 2018). For example, most taxa produce primarily a single type of anthocyanin, and of those that produce two types, none include the two extremes of the hydroxylation spectrum (pelargonidin plus delphinidin). One explanation for these

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

complex patterns is the presence of trade-offs, which could constrain the course of evolution (Chesmore et al. 2016; Wagner and Lynch 2008). Our study took the first step toward modeling these trade-offs by tracking the changes in pathway function and output during an adaptive walk. These simulations uncovered significant 'ripple effects' of selection for delphinidin, which were closely tied to pathway structure. For example, evolution towards the blue state negatively affected production of the other anthocyanins, but the decrease was stronger for the red pelargonidin pigments. This pattern is predicted based on the fact that flux toward delphinidin required higher activity of F3'H and F3'5'H, sapping the DHK precursor for pelargonidin. These types of trade-offs could account for the focusing of production onto one or two anthocyanins as well as the inaccessibility of a phenotype comprising the two extremes (pelargonidin plus delphinidin). Nonetheless, in reaching 90% delphinidin production, we did not observe the complete loss of pelargonidin as is seen in nature (Ng et al. 2018). This discord may reflect several factors that are not incorporated into our simple model, such as the co-regulation of pathway enzymes by shared transcription factors (Albert et al. 2014; Mol et al. 1998).

Conclusions

This study provides independent support for influence of pathway topology on the distribution of functional mutations for evolving anthocyanin pigmentation. We envision that our model will be valuable for basic research aimed at understanding how selection acts on complex traits as well as applied work focused on flower pigmentation (Sasaki and Nakayama 2015; Shimada et al. 2001) or other phenotypes controlled by metabolic pathways. We note that the basic computational approach developed here could be applied to any enzymatic pathway, opening the door to a broader exploration of how the elements of pathway topology (e.g., length, degree of branching) influence the evolutionary process. It is also straightforward to expand the model to include reversible reactions, multiple 'containers' (cellular compartments), pleiotropic mutations, complex selective optima and other features common in biological systems. In the case of the anthocyanin pathway, our ongoing and future research aims to connect this computational model with *in vivo* activity

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of the pathway and organismal phenotypes, with the ultimate goal of understanding how these molecular mechanisms interact with selection in shaping phenotypic diversity.

Acknowledgments

We thank members of the Smith lab for useful conversations regarding implementation and interpretation of our computational model. We thank Jacob Stanley and Rutendo Sigauke in the Dowell group at CU Boulder for helpful conversations regarding development of the simulation framework and appropriate implementation of the subsequent analyses. We also thank Boswell Wing, Sebastian Kopf, and the other members of the CU Boulder Geobiology Super Group for their helpful feedback. Finally, we thank Herbet Sauro, Kiri Choi, and Kyle Medley from the University of Washington for their prompt assistance with using the Tellurium library. This work utilized the RMACC Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University.

Funding

This work was funded by NSF-DEB 1553114. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The python library *enzo*, is available on github (https://github.com/lcwheeler/enzo). The raw simulated dataset ("wheeler-smith-simulated-dataset.p") can be found along with other materials on the Zenodo online repository (https://zenodo.org/record/2611739#.XJvOwN-YU8o). Additional details are available in the supplemental text.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author contributions

LCW and SDS conceived the study and outlined the computational approach. LCW constructed the mathematical model, wrote the code, performed the simulations, conducted the data analysis, and generated figures. LCW and SDS wrote the manuscript. SDS secured funding for the work. All authors have read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

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. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell*, 26(3):962–980, Mar. 2014. ISSN 1040-4651, 1532-298X.

Arbeithuber, B., Betancourt, A. J., Ebner, T., and Tiemann-Boege, I. Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *PNAS*, 112(7):2109–2114, Feb. 2015. ISSN 0027-8424, 1091-6490.

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21):4402–4410, 2011.

Beale, G. H., Price, J. R., Sturgess, V. C., and Robinson, R. A survey of anthocyanins vii. the natural selection of flower colour. *Proceedings of the Royal Society of London. Series B - Biological Sciences*, 130(858):113–126, Aug. 1941.

Braakman, R., Follows, M. J., and Chisholm, S. W. Metabolic evolution and the self-organization of ecosystems. *Proc. Natl. Acad. Sci. U.S.A.*, 114(15):E3091–E3100, Apr. 2017. ISSN 0027-8424.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chen, H.-C., Song, J., Wang, J. P., Lin, Y.-C., Ducoste, J., Shuford, C. M., Liu, J., Li, Q., Shi, R., Nepomuceno, A., Isik, F., Muddiman, D. C., Williams, C., Sederoff, R. R., and Chiang, V. L. Systems biology of lignin biosynthesis in populus trichocarpa: Heteromeric 4-coumaric acid:coenzyme a ligase protein complex formation, regulation, and numerical modeling. *Plant Cell*, 26(3):876–893, Mar. 2014. ISSN 1040-4651, 1532-298X.

Chesmore, K. N., Bartlett, J., Cheng, C., and Williams, S. M. Complex patterns of association between pleiotropy and transcription factor evolution. *Genome Biol Evol*, 8(10):3159–3170, Oct. 2016.

Choi, K., Medley, J. K., Cannistra, C., Konig, M., Smith, L., Stocking, K., and Sauro, H. M. Tellurium: A python based modeling and reproducibility platform for systems biology. *bioRxiv*, page 054601, June 2016.

Chou, T. C. and Talaly, P. A simple generalized equation for the analysis of multiple inhibitions of michaelis-menten kinetic systems. *J. Biol. Chem.*, 252(18):6438–6442, Sept. 1977. ISSN 0021-9258, 1083-351X.

Colón, A. M., Sengupta, N., Rhodes, D., Dudareva, N., and Morgan, J. A kinetic model describes metabolic response to perturbations and distribution of flux control in the benzenoid network of petunia hybrida. *Plant J.*, 62(1):64–76, 2010. ISSN 1365-313X.

Cooper, G. M. Protein synthesis, processing, and regulation. *Cell Mol. Approach 2nd Ed.*, 2000.

Cooper, T. F., Rozen, D. E., and Lenski, R. E. Parallel changes in gene expression after 20,000 generations of evolution in escherichia coli. *PNAS*, 100(3):1072–1077, Feb. 2003. ISSN 0027-8424, 1091-6490.

Cornish-Bowden, A. Metabolic control analysis in theory and practice. In Bittar, E. E., editor, *Advances in Molecular and Cell Biology*, volume 11 of *Enzymology in Vivo*, pages 21–64. Elsevier, Jan. 1995.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M. A., Kimmel, C. B., and Postlethwait, J. H. Parallel genetic basis for repeated evolution of armor loss in alaskan threespine stickleback populations. *PNAS*, 101(16):6050–6055, Apr. 2004. ISSN 0027-8424, 1091-6490.

Dekel, E. and Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050):588–592, July 2005. ISSN 1476-4687.

Des Marais, D. L. and Rausher, M. D. Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in ipomoea. *Evolution*, 64(7):2044–2054, 2010. ISSN 1558-5646.

Dykhuizen, D. E., Dean, A. M., and Hartl, D. L. Metabolic flux and fitness. *Genetics*, 115(1): 25–31, Jan. 1987. ISSN 0016-6731, 1943-2631.

Esfeld, K., Berardi, A. E., Moser, M., Bossolini, E., Freitas, L., and Kuhlemeier, C. Pseudogenization and resurrection of a speciation gene. *Current Biology*, 28(23):3776–3786.e7, Dec. 2018. ISSN 0960-9822.

Fischer, T. C., Halbwirth, H., Meisel, B., Stich, K., and Forkmann, G. Molecular cloning, substrate specificity of the functionally expressed dihydroflavonol 4-reductases from malus domestica and pyrus communis cultivars and the consequences for flavonoid metabolism. *Arch. Biochem. Biophys.*, 412(2):223–230, Apr. 2003. ISSN 0003-9861.

Flowers, J. M., Sezgin, E., Kumagai, S., Duvernell, D. D., Matzkin, L. M., Schmidt, P. S., and Eanes, W. F. Adaptive evolution of metabolic pathways in drosophila. *Mol. Biol. Evol.*, 24(6): 1347–1354, June 2007. ISSN 0737-4038.

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., and Carroll, S. B. Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature*, 433(7025):481–487, Feb. 2005. ISSN 1476-4687.

Groenenboom, M., Gomez-Roldan, V., Stigter, H., Astola, L., van Daelen, R., Beekwilder, J., Bovy,

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A., Hall, R., and Molenaar, J. The flavonoid pathway in tomato seedlings: Transcript abundance and the modeling of metabolite dynamics. *PLoS One*, 8(7):e68960, July 2013. ISSN 1932-6203.

Guillaume, F. and Otto, S. P. Gene functional trade-offs and the evolution of pleiotropy. *Genetics*, 192(4):1389–1409, Dec. 2012. ISSN 0016-6731, 1943-2631.

Halbwirth, H. The creation and physiological relevance of divergent hydroxylation patterns in the flavonoid pathway. *Int. J. Mol. Sci.*, 11(2):595–621, Feb. 2010.

Halbwirth, H., Kahl, S., Jäger, W., Reznicek, G., Forkmann, G., and Stich, K. Synthesis of (14c)-labeled 5-deoxyflavonoids and their application in the study of dihydroflavonol/leucoanthocyanidin interconversion by dihydroflavonol 4-reductase. *Plant Science*, 170(3):587–595, Mar. 2006. ISSN 0168-9452.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P. M., and Lercher, M. J. Predicting c4 photosynthesis evolution: Modular, individually adaptive steps on a mount fuji fitness landscape. *Cell*, 153(7):1579–1588, June 2013. ISSN 0092-8674, 1097-4172.

Heinrich, R. and Schuster, S. The regulation of cellular systems, 1996.

Hoefnagel, M. H. N., Starrenburg, M. J. C., Martens, D. E., Hugenholtz, J., Kleerebezem, M., Van Swam, I. I., Bongers, R., Westerhoff, H. V., and Snoep, J. L. Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. *Microbiology*, 148(4):1003–1013, 2002.

Johnson, E. T., Ryu, S., Yi, H., Shin, B., Cheong, H., and Choi, G. Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. *Plant J.*, 25(3):325–333, Feb. 2001. ISSN 1365-313X.

Kacser, H., Burns, J. A., Kacser, H., and Fell, D. A. The control of flux. *Biochem. Soc. Trans.*, 23 (2):341–366, May 1995. ISSN 0300-5127, 1470-8752.

Kaltenbach, M. and Tokuriki, N. Dynamics and constraints of enzyme evolution. *J. Exp. Zool. (Mol. Dev. Evol.)*, 322(7):468–487, Nov. 2014. ISSN 1552-5015.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kanehisa, M. The kegg database. In *'In Silico' Simulation of Biological Processes*, pages 91–103. John Wiley & Sons, Ltd, 2008. ISBN 978-0-470-85789-2.

Katsu, K., Suzuki, R., Tsuchiya, W., Inagaki, N., Yamazaki, T., Hisano, T., Yasui, Y., Komori, T., Koshio, M., Kubota, S., Walker, A. R., Furukawa, K., and Matsui, K. A new buckwheat dihydroflavonol 4-reductase (dfr), with a unique substrate binding structure, has altered substrate specificity. *BMC Plant Biol.*, 17(1):239, Dec. 2017. ISSN 1471-2229.

Keller, N. P., Turner, G., and Bennett, J. W. Fungal secondary metabolism — from biochemistry to genomics. *Nat. Rev. Microbiol.*, 3(12):937–947, Dec. 2005. ISSN 1740-1534.

Khersonsky, O. and Tawfik, D. S. Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu. Rev. Biochem.*, 79(1):471–505, 2010.

Kuang, M. C., Kominek, J., Alexander, W. G., Cheng, J.-F., Wrobel, R. L., and Hittinger, C. T. Repeated cis-regulatory tuning of a metabolic bottleneck gene during evolution. *Mol. Biol. Evol.*, 35(8):1968–1981, Aug. 2018. ISSN 0737-4038.

Martin, A. and Orgogozo, V. The loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. *Evolution*, 67(5):1235–1250, May 2013. ISSN 1558-5646.

Miranda, H. V., Ferreira, A. E. N., Quintas, A., Cordeiro, C., and Freire, A. P. Measuring intracellular enzyme concentrations: Assessing the effect of oxidative stress on the amount of glyoxalase i. *Biochem. Mol. Biol. Educ.*, 36(2):135–138, Mar. 2008. ISSN 14708175, 15393429.

Miyagawa, N., Miyahara, T., Okamoto, M., Hirose, Y., Sakaguchi, K., Hatano, S., and Ozeki, Y. Dihydroflavonol 4-reductase activity is associated with the intensity of flower colors in delphinium. *Plant Biotechnol.*, 32(3):249–255, 2015. ISSN 1342-4580, 1347-6114.

Mol, J., Grotewold, E., and Koes, R. How genes paint flowers and seeds. *Trends Plant Sci.*, 3(6): 212–217, June 1998. ISSN 1360-1385.

Morrison, E. S. and Badyaev, A. V. Structuring evolution: biochemical networks and metabolic diversification in birds. *BMC Evol. Biol.*, 16(1):168, Aug. 2016. ISSN 1471-2148.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Morrison, E. S. and Badyaev, A. V. Beyond topology: coevolution of structure and flux in metabolic networks. *J. Evol. Biol.*, 30(10):1796–1809, 2017. ISSN 1420-9101.

Morrison, E. S. and Badyaev, A. V. Structure versus time in the evolutionary diversification of avian carotenoid metabolic networks. *J. Evol. Biol.*, 31(5):764–772, May 2018. ISSN 1420-9101.

Nam, H., Conrad, T., and Lewis, N. E. Metabolic pathway evolution and cellular objectives. *Curr. Opin. Biotechnol.*, 22(4):595–600, Aug. 2011. ISSN 0958-1669.

Ng, J. and Smith, S. D. How to make a red flower: the combinatorial effect of pigments. *AoB PLANTS*, 8, Jan. 2016a.

Ng, J. and Smith, S. D. Widespread flower color convergence in solanaceae via alternate biochemical pathways. *New Phytol.*, 209(1):407–417, 2016b. ISSN 1469-8137.

Ng, J. and Smith, S. D. Why are red flowers so rare? testing the macroevolutionary causes of tippiness. *J. Evol. Biol.*, 31(12):1863–1875, 2018. ISSN 1420-9101.

Ng, J., Freitas, L. B., and Smith, S. D. Stepwise evolution of floral pigmentation predicted by biochemical pathway structure. *Evolution*, 72(12):2792–2802, 2018. ISSN 1558-5646.

Panettieri, S., Gjinaj, E., John, G., and Lohman, D. J. Different ommochrome pigment mixtures enable sexually dimorphic batesian mimicry in disjunct populations of the common palmfly butterfly, elymnias hypermnestra. *PLoS One*, 13(9):e0202465, Sept. 2018. ISSN 1932-6203.

Petrussa, E., Braidot, E., Zancani, M., Peresson, C., Bertolini, A., Patui, S., and Vianello, A. Plant flavonoids—biosynthesis, transport and involvement in stress responses. *Int. J. Mol. Sci.*, 14(7): 14950–14973, July 2013. ISSN 1422-0067.

Pichersky, E. and Gang, D. R. Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.*, 5(10):439–445, Oct. 2000. ISSN 1360-1385.

Poustka, F., Irani, N. G., Feller, A., Lu, Y., Pourcel, L., Frame, K., and Grotewold, E. A trafficking pathway for anthocyanins overlaps with the endoplasmic reticulum-to-vacuole protein-sorting

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

route in arabidopsis and contributes to the formation of vacuolar inclusions. *Plant Physiol.*, 145 (4):1323–1335, Dec. 2007. ISSN 0032-0889.

Quattrocchio, F., Baudry, A., Lepiniec, L., and Grotewold, E. *The Regulation of Flavonoid Biosynthesis*, pages 97–122. Springer New York, New York, NY, 2006. ISBN 978-0-387-28822-2.

Rausher, M. D. The evolution of genes in branched metabolic pathways: Evolution in branched pathways. *Evolution*, 67(1):34–48, Jan. 2013. ISSN 00143820.

Reed, R. D. and Serfas, M. S. Butterfly wing pattern evolution is associated with changes in a notch/distal-less temporal pattern formation process. *Curr. Biol.*, 14(13):1159–1166, July 2004. ISSN 0960-9822.

Ryan, K. G., Markham, K. R., Bloor, S. J., Bradley, J. M., Mitchell, K. A., and Jordan, B. R. Uvb radiation induced increase in quercetin: Kaempferol ratio in wild-type and transgenic lines of petunia. *Photochem. Photobiol.*, 68(3):323–330, 1998. ISSN 1751-1097.

Ryan, K. G., Swinny, E. E., Markham, K. R., and Winefield, C. Flavonoid gene expression and uv photoprotection in transgenic and mutant petunia leaves. *Phytochemistry*, 59(1):23–32, Jan. 2002. ISSN 0031-9422.

Sasaki, N. and Nakayama, T. Achievements and perspectives in biochemistry concerning anthocyanin modification for blue flower coloration. *Plant Cell Physiol.*, 56(1):28–40, Jan. 2015. ISSN 0032-0781.

Schäuble, S., Stavrum, A. K., Puntervoll, P., Schuster, S., and Heiland, I. Effect of substrate competition in kinetic models of metabolic networks. *FEBS Lett.*, 587(17):2818–2824, Sept. 2013. ISSN 0014-5793.

Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F., and Schomburg, D. Brenda: a resource for enzyme data and metabolic information. *Trends Biochem. Sci.*, 27(1):54–56, Jan. 2002. ISSN 0968-0004.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Schuster, S., Dandekar, T., and Fell, D. A. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.*, 17(2):53–60, Feb. 1999. ISSN 0167-7799.

Schuster, S., Fell, D. A., and Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.*, 18(3): 326–332, Mar. 2000. ISSN 1546-1696.

Shimada, Y., Ohbayashi, M., Nakano-Shimada, R., Okinaka, Y., Kiyokawa, S., and Kikuchi, Y. Genetic engineering of the anthocyanin biosynthetic pathway with flavonoid-3′,5′-hydroxylase: specific switching of the pathway in petunia. *Plant Cell Rep.*, 20(5):456–462, July 2001. ISSN 1432-203X.

Smith, L. P., Bergmann, F. T., Chandran, D., and Sauro, H. M. Antimony: a modular model definition language. *Bioinformatics*, 25(18):2452–2454, Sept. 2009. ISSN 1367-4803.

Smith, S. D. and Rausher, M. D. Gene loss and parallel evolution contribute to species difference in flower color. *Mol Biol Evol*, 28(10):2799–2810, Oct. 2011. ISSN 0737-4038.

Smith, S. D., Wang, S., and Rausher, M. D. Functional evolution of an anthocyanin pathway enzyme during a flower color transition. *Mol. Biol. Evol.*, 30(3):602–612, Mar. 2013. ISSN 0737-4038.

Sobel, J. and Streisfeld, M. Flower color as a model system for studies of plant evo-devo. *Frontiers in Plant Science*, 4:321, 2013. ISSN 1664-462X.

Stephanopoulos, G. Metabolic fluxes and metabolic engineering. *Metab. Eng.*, 1(1):1–11, Jan. 1999. ISSN 1096-7176.

Streisfeld, M. A. and Rausher, M. D. Altered trans-regulatory control of gene expression in multiple anthocyanin genes contributes to adaptive flower color evolution in mimulus aurantiacus. *Mol. Biol. Evol.*, 26(2):433–444, Feb. 2009a. ISSN 0737-4038.

Streisfeld, M. A. and Rausher, M. D. Genetic changes contributing to the parallel evolution of red floral pigmentation among ipomoea species. *New Phytol.*, 183(3):751–763, 2009b. ISSN 1469-8137.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Streisfeld, M. A. and Rausher, M. D. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution*, 65(3):629–642, 2011. ISSN 1558-5646.

Streisfeld, M. A., Liu, D., and Rausher, M. D. Predictable patterns of constraint among anthocyanin-regulating transcription factors in ipomoea. *New Phytol.*, 191(1):264–274, July 2011. ISSN 0028646X.

Strohman, R. Maneuvering in the complex path from genotype to phenotype. *Science*, 296(5568): 701–703, Apr. 2002. ISSN 0036-8075, 1095-9203.

Tanaka Yoshikazu and Brugliera Filippa. Flower colour and cytochromes p450†. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1612):20120432, Feb. 2013.

Treves, D. S., Manning, S., and Adams, J. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of escherichia coli. *Mol. Biol. Evol.*, 15(7):789–797, July 1998. ISSN 0737-4038.

Tripp, E. A. and Manos, P. S. Is floral specialization an evolutionary dead-end? pollination system transitions in ruellia (acanthaceae). *Evolution*, 62(7):1712–1737, 2008. ISSN 1558-5646.

True, J. R., Edwards, K. A., Yamamoto, D., and Carroll, S. B. Drosophila wing melanin patterns form by vein-dependent elaboration of enzymatic prepatterns. *Curr. Biol.*, 9(23):1382–1391, Dec. 1999. ISSN 0960-9822.

Wagner, G. P. and Lynch, V. J. The gene regulatory logic of transcription factor evolution. *Trends Ecol. Evol. (Amst.)*, 23(7):377–385, July 2008. ISSN 0169-5347.

Wessinger, C. A. and Rausher, M. D. Lessons from flower colour evolution on targets of selection. *J Exp Bot*, 63(16):5741–5749, Oct. 2012a. ISSN 0022-0957.

Wessinger, C. A. and Rausher, M. D. Lessons from flower colour evolution on targets of selection. *J. Exp. Bot.*, 63(16):5741–5749, Oct. 2012b. ISSN 0022-0957.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wessinger, C. A. and Rausher, M. D. Predictability and irreversibility of genetic changes associated with flower color evolution in penstemon barbatus. *Evolution*, 68(4):1058–1070, 2014. ISSN 1558-5646.

Wessinger, C. A. and Rausher, M. D. Ecological transition predictably associated with gene degeneration. *Mol Biol Evol*, 32(2):347–354, Feb. 2015. ISSN 0737-4038.

Winkel-Shirley, B. Flavonoid biosynthesis. a colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.*, 126(2):485–493, June 2001. ISSN 0032-0889.

Wisecaver, J. H., Slot, J. C., and Rokas, A. The evolution of fungal metabolic pathways. *PLos Genet.*, 10(12):e1004816, Dec. 2014. ISSN 1553-7404.

Wittkopp, P. J., Carroll, S. B., and Kopp, A. Evolution in black and white: genetic control of pigment patterns in drosophila. *Trends Genet.*, 19(9):495–504, Sept. 2003. ISSN 0168-9525.

Woods, R., Schneider, D., Winkworth, C. L., Riley, M. A., and Lenski, R. E. Tests of parallel molecular evolution in a long-term experiment with escherichia coli. *PNAS*, 103(24):9107–9112, June 2006. ISSN 0027-8424, 1091-6490.

Wright, K. M. and Rausher, M. D. The evolution of control and distribution of adaptive mutations in a metabolic pathway. *Genetics*, 184(2):483–502, Feb. 2010. ISSN 0016-6731, 1943-2631.

Yuan, Y.-W., Rebocho, A. B., Sagawa, J. M., Stanley, L. E., and Bradshaw, H. D. Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between mimulus species. *PNAS*, 113(9):2448–2453, Mar. 2016. ISSN 0027-8424, 1091-6490.

Zhang, L., Martin, A., Perry, M. W., van der Burg, K. R. L., Matsuoka, Y., Monteiro, A., and Reed, R. D. Genetic basis of melanin pigmentation in butterfly wings. *Genetics*, 205(4):1537–1550, Apr. 2017. ISSN 0016-6731.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

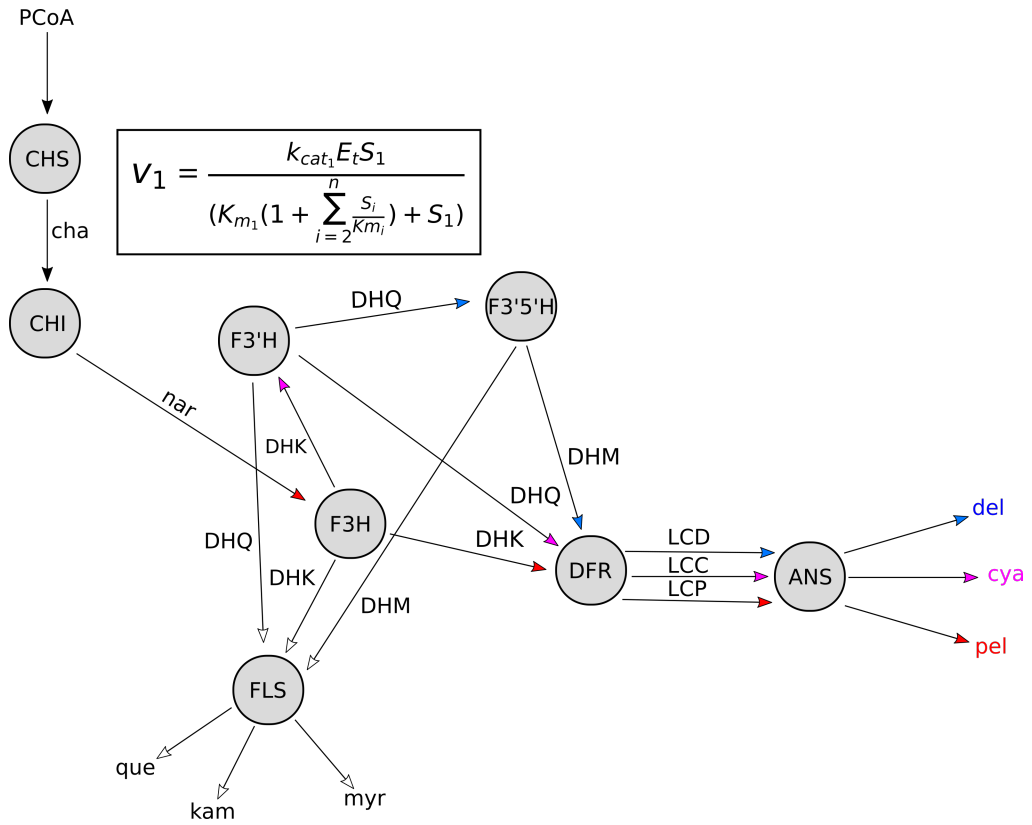

Fig 1. Diagram of the simplified anthocyanin pathway model. The simplified model is depicted as an enzyme-centric pathway diagram, including the key anthocyanin and flavonol branches (see Fig. S1 for classical substrate-centric diagram). Grey circles are enzymes. Arrows indicate influx and outflux of floating species. Arrow placement indicates the linkage: e.g. DHM to LCD to del. Blue arrows lead to delphinidin, red arrows lead to pelargonidin, purple arrows lead to cyanidin. Inset shows general form of the irreversible rate law (accounting for substrate competition): S_1 is concentration of substrate 1, $K_{cat,1}$ is the catalytic constant (turnover rate) for substrate 1, $K_{M,1}$ is the Michaelis constant for substrate 1, the sum is over all other K_M values and concentrations for competing substrates (Schäuble et al. 2013). A complete set of rate laws for all pathway reactions are available in the supplemental text. Floating species abbreviations are: PCoA (P-Coumaroyl-CoA), cha (chalcone), nar (naringenin), DHK (dihydrokampferol), DHQ (dihydroquercetin), DHM (dihydromyricetin), que (quercetin), kam (kampferol), myr (myricetin), LCD (leucopelargonidin), LCC (leucocyanidin), LCD (leucodelphinidin), pel (pelargonidin), cya (cyanidin), del (delphinidin). Del, cya, and pel are the anthocyanidins, which are glycosylated to form the various anthocyanins. Kam, que, and myr are the flavonols. Enzyme abbreviations are: CHS (chalcone synthase), CHI (chalcone isomerase), F3H (flavanone-3-hydroxylase), F3'H (flavonol-3'hydroxylase), F3'5'H (flavonoid-3'5'hydroxylase), DFR (dihyroflavonol-4-reductase), FLS (flavonol synthase), ANS (anthocyanidin synthase). The suite of glycosylating and methylating enyzmes (e.g. UF3GT, RT, OMT, (Winkel-Shirley 2001)) that convert the core three anthocyanidins (shown) into the wide range of anthocyanins are not included, for simplicity.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

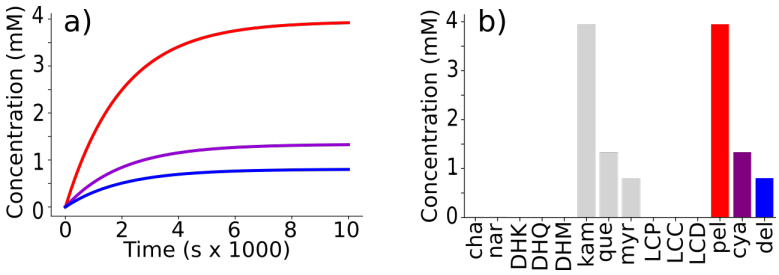

Fig 2. Kinetic behavior of the naïve starting state model. a) time-course simulation of the model showing anthocyanidin concentrations over time: pelargondin (red; highest saturation curve), cyanidin (purple; middle saturation curve), and delphinidin (blue; lowest saturation curve). b) Steady state concentrations of each floating species in the naïve model. The intermediate floating species (e.g. nar and DHM) have a negligible concentration at steady state.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

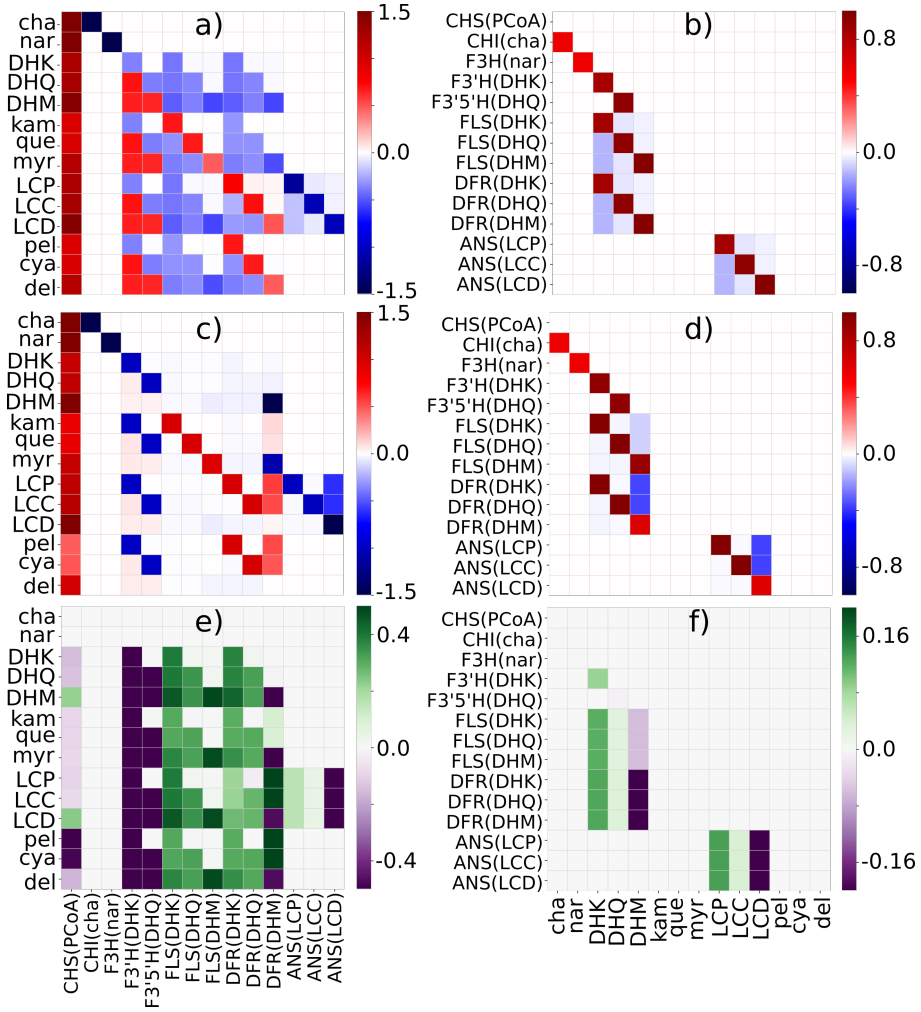

Fig 3. Pathway control shifts to achieve a bias toward the delphinidin branch. a) Heatmap of concentration control coefficients for all pathway reactions (except for "sinks" at boundary conditions) in the naïve starting state model. b) Heatmap of elasticity matrix for all pathway reactions (except for diffusion "sinks" at boundaries) in the naïves starting state model. c) Median optimal concentration control coefficient matrix, taken over all 9,985 simulations. d) Heatmap of median optimal elasticity coefficient matrix, taken over all 9,985 simulations. e) Difference between the two control coefficient matrices (median optimal end-state matrixminus starting-state matrix). f) Difference between the two elasticiity matrices (median optimal end-state matrix minus starting state matrix). In heatmaps, abbreviated floating species identifiers are used for each unique floating species in the model (see Fig. 1). Each unique reaction in the pathway model is denoted using the format "Enzyme(floating species)", for example ANS(LCD) represents the reaction resulting from ANS acting on LCD as a subtrate. For difference matrices: purple indicates a negative shift in control and green indicates a positive shift.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

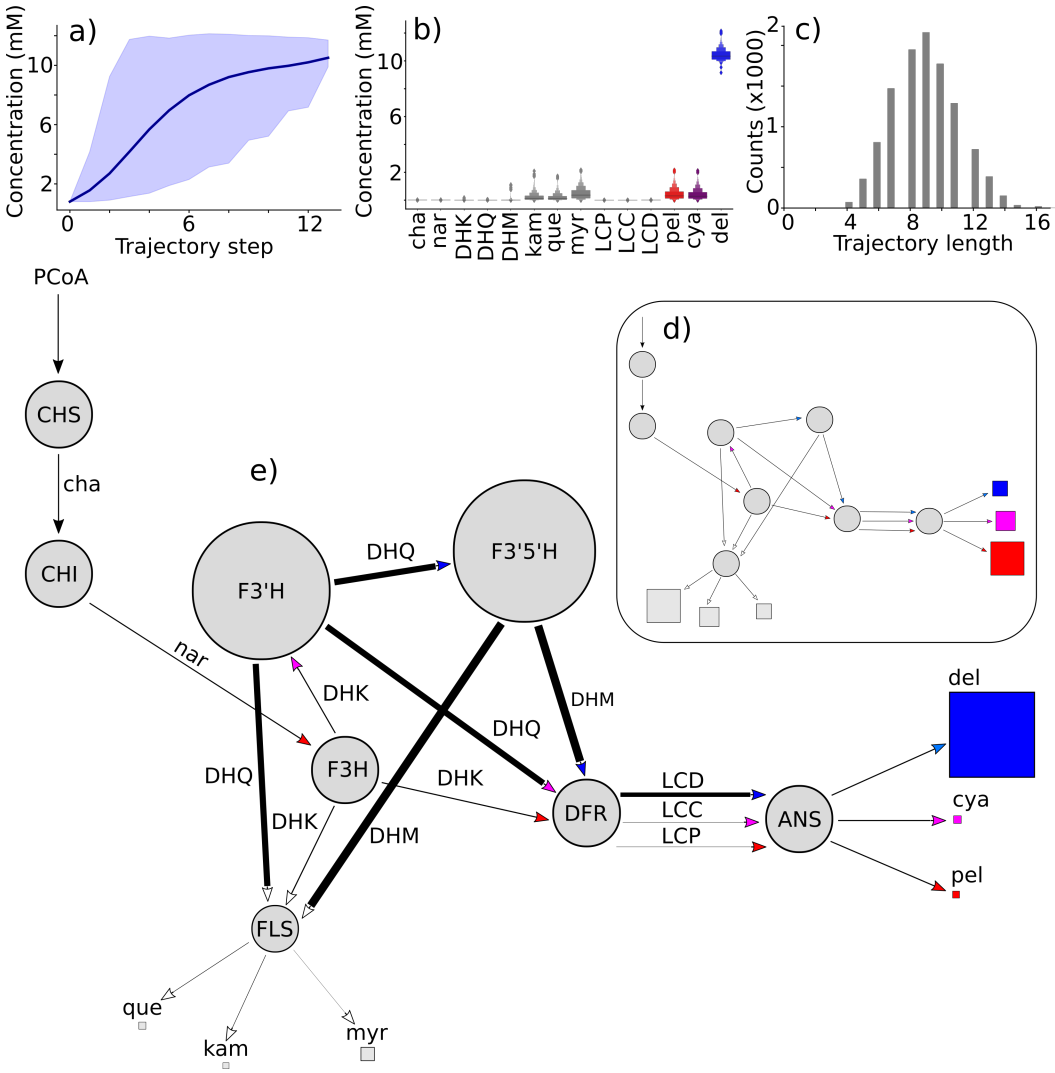

Fig 4. Simulated evolution of delphinidin production. a) Trajectories through delphinidin space. Dark blue line shows mean trajectory, averaged across all 9,985 simulations. Shaded blue area shows envelope containing 95% of trajectories. Top limit is maximum value at each step, bottom limit is minimum value. b) Distributions of optimal (end-point) steady state concentrations for each floating species in the model shown as a boxenplot. c) Distribution of simulated trajectory lengths (median length is 9 steps). d) Inset shows outline of the starting state model, in which all parameters of a given type are equal (see methods). e) The mean values for the model parameters from all delphinidin-optimized end points are shown on the pathway diagram. Enzyme circle size is proportional to concentration. Arrow line thickness is proportional to mean K_{cat}/K_M. The area of squares at the end of branches are proportional to steady state concentrations of the indicated floating species produced by a model initialized with mean parameter values.

bioRxiv preprint doi: https://doi.org/10.1101/511089; this version posted May 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

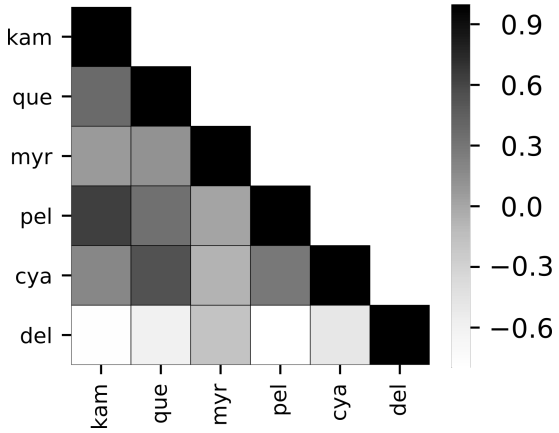

Fig 5. Pleiotropic trade-offs in pathway products. Heatmap of mean Spearman correlations between concentrations of anthocyanins (pel, del, and cya) and flavonols (kam, que, myr) across all 9,985 evolutionary trajectories. Abbreviations are the same as in Fig. 1. A full correlation matrix for all floating species and correlation matrices for model parameters are shown in Fig. S5.