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

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 realistic, functional 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 the observed trajectories and constraints 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

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

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, 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 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; Guillaume and Otto 2012), and thus may introduce constraints on evolutionary trajectories.

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 interactions 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 2012, 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 nested within the larger architecture of flavonoid biosynthesis, which produces an array of related compounds that act as pigments, sunscreens, and 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, respectively; 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. 1) (Streisfeld and Rausher 2011; Wessinger and Rausher 2012). These changes may involve structural mutations that alter enzyme function (Wessinger and Rausher 2014), regulatory changes that shift expression (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 give rise to evolutionary trends, such as the rarity of some flower colors (Ng and Smith 2018) and the highly asymmetric transition rates among phenotypes (e.g.(Ng et al. 2018; Tripp and Manos 2008)). Nevertheless, testing such a relationship requires a theoretical framework that unites the topology of the pathway with phenotypic evolution via mutations 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 structure 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? Simulating evolution toward a new optimum pigment distribution will allow us to quantify how control is distributed across the pathway (i.e., which enzymes determine flux) and how pathway structure shapes evolutionary

trajectories. 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 that these transitions may lead to trade-offs within and between branches of the pathway which are under selection. The results of these simulation analyses will not only lay the foundation for continuing 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 activity are generally 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 realistic starting state to learn about the relative importance of

different model parameters and the interplay between these parameters. We chose values for catalytic constants and Michaelis constants from a comprehensive published meta-analysis (Bar-Even et al. 2011) of enzyme properties performed by mining the BRENDA and KEGG databases (Kanehisa 2008; Schomburg et al. 2002). The model was initalized at a starting state with all parameters of a certain type set to be equal. Values for K_{cat}were set to 14 s^{-1}, the median of the distribution calculated in the Barr-Even et al. meta-analysis (Bar-Even et al. 2011). K_M values were set to 0.013 mM, one tenth the median value of the empirical distribution to insure that starting conditions of the model put the upstream substrate concentration well above K_M of the first enzyme and all subsequent enzymes. Enzyme concentrations were initialized to 0.0001 mM, a value within the range of naturally-occurring protein concentrations (Miranda et al. 2008). This model represents a naïve, un-optimized 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. The simulations were performed under the assumption of a single compartment, analagous to all pathway components being present together in a test tube. Tellurium numerically integrates the set of coupled differential equations describing the pathway model using the Roadrunner API (Somogyi et al. 2015). To confirm that steady state values calculated by Tellurium were consistent with other standard methods, the model was written out in SBML format ("supplemental-file-1.sbml") (Hucka et al. 2003) and the simulations were then re-run in COPASI (Hoops et al. 2006). The Tellurium and COPASI values for floating species steady state concentrations were identical.

Implementation of 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. The core Pathway class of *enzo* is a wrapper for Tellurium and Roadrunner model objects, which holds a model object as an attribute. The Pathway object possesses bound methods for conducting evolutionary operations, which allow it to access the built-in attributes of the Roadrunner model. A second PathwaySet class holds an ensemble of Pathway objects, allowing multiple simulations to be conducted while the data for each is collected separately. All Tellurium/Roadrunner model attributes and functions remain accessible to the user.

Evolutionary simulations proceed as follows: First, a PathwaySet object object is initialized with a Roadrunner model, built from a user-defined Antimony string. The number of simulations (replicate Pathway objects held as attributes of the Pathway set) is defined. The *evolve* method of the PathwaySet object is then used (with user-defined settings for the optimum, number of iterations, and other arguments) to evolve each Pathway object contained in the PathwaySet toward the new optimum using a simple evolutionary model, in which fixation probability is weighted by the mutational effect size on fitness. Each Pathway simulation stops either when it reaches a pre-defined maximum number of mutational steps (50,000) or reaches the new optimum within a user-defined fractional tolerance level (see Supplemental text for details of evolutionary algorithm). Steady state concentrations at each fixed step, selection coefficients for each mutation, fixed parameter mutation values, control coefficient and elasticity matrices for each evolved state, and a final optimum parameter set are retained for each simulation and held in memory as attributes of the Pathway objects. The entire PathwaySet object, containing the ensemble of simulated models and all of the corresponding data, can be stored as a serialized *pickle* file.

Given the structure of the model, mutations during these evolutionary simulations can alter a range of enzyme parameters. These include enzyme concentration (E_t), Michaelis constant (K_m, proportional to binding affinity), and catalytic rate (K_{cat}). Each enzyme has one concentration

parameter, but can have multiple K_m and K_{cat} values, depending on the number of substrates. The model allows any change to occur independently without correlated effects on other parameters. For example, DFR, an enzyme which can bind to multiple substrates, can undergo a mutation to improve its activity on one substrate without a change in activity on other substrates, consistent with the different Michaelis constants and catalytic rates 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 patterns of correlated changes, positive or negative. Although our model currently assumes independent changes in kinetic parameters, expanding to include correlated changes could be implemented in future (see Discussion).

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 model. These tolerance values are flexible arguments to the *evolve* function of the *Pathway* object in *enzo* and can be modified to suite the needs of the user.

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 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). 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 with 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. 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 9985 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 losing control over competing substrates (DHQ and DHK). 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 Suppl. 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 steady state 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).

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

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. 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 Geoffrey Herbert 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 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 2012). A growing suite of studies has examined evolutionary transitions between blue and red flowers, and these implicate two branching enzymes, F3'H and F3'5'H as the key players determining the type of anthocyanin produced (Marais and Rausher 2010; Smith and Rausher 2011; Streisfeld and Rausher 2009b;

Wessinger and Rausher 2015). Our evolutionary simulations from a red, mostly pelargonidin starting-state to a blue, mostly delphinidin ending-state recovered a similar targeting of changes to these loci (Fig. 5e; 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. 5e, Fig. S4). The action of these genes as hotspots for color evolution intuitively make 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 remarkable alignment across these empirical and theoretical studies points to a strong relationship between the structure of the pathway and loci that predictably underlie phenotypic evolution.

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 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 (reviewed in (Ng and Smith 2016a)). 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 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 sorts 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 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), 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 of the pathway and organismal phenotypes, with the ultimate goal of understanding how these molecular mechanisms interact with selection in shaping phenotypic diversity.

Acknowledgements

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 criticisms. Finally, we thank Herbet Sauro, Kiri Choi, and Kyle Medley from the University of Washington for their prompt assistance with using the Tellurium library.

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 datasetwill be uploaded to the Dryad data repository (https://datadryad.org/) upon publication. Additional details are available in the supplemental text.

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

Nick W. Albert, Kevin M. Davies, David H. Lewis, Huaibi Zhang, Mirco Montefiori, Cyril Brendolise, Murray R. Boase, Hanh Ngo, Paula E. Jameson, and Kathy E. Schwinn. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell*, 26(3):962–980, March 2014. ISSN 1040-4651, 1532-298X.

Barbara Arbeithuber, Andrea J. Betancourt, Thomas Ebner, and Irene Tiemann-Boege. Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *PNAS*, 112 (7):2109–2114, February 2015. ISSN 0027-8424, 1091-6490.

Arren Bar-Even, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S. Tawfik, and Ron Milo. The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. https://pubs.acs.org/doi/full/10.1021/bi2002289, May 2011.

Beale Geoffrey Herbert, Price James Robert, Sturgess V. C., and Robinson Robert. 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, August 1941.

Rogier Braakman, Michael J. Follows, and Sallie W. Chisholm. Metabolic evolution and the self-

organization of ecosystems. *Proc Natl Acad Sci U S A*, 114(15):E3091–E3100, April 2017. ISSN 0027-8424.

Hsi-Chuan Chen, Jina Song, Jack P. Wang, Ying-Chung Lin, Joel Ducoste, Christopher M. Shuford, Jie Liu, Quanzi Li, Rui Shi, Angelito Nepomuceno, Fikret Isik, David C. Muddiman, Cranos Williams, Ronald R. Sederoff, and Vincent L. Chiang. 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, March 2014. ISSN 1040-4651, 1532-298X.

Kevin N. Chesmore, Jacquelaine Bartlett, Chao Cheng, and Scott M. Williams. Complex patterns of association between pleiotropy and transcription factor evolution. *Genome Biol Evol*, 8(10): 3159–3170, October 2016.

Kiri Choi, Jesse Kyle Medley, Carroline Cannistra, Matthias Konig, Lucian Smith, Kaylene Stocking, and Herbert M. Sauro. Tellurium: A python based modeling and reproducibility platform for systems biology. *bioRxiv*, page 054601, June 2016.

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

Amy Marshall Colón, Neelanjan Sengupta, David Rhodes, Natalia Dudareva, and John Morgan. 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.

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

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

William A. Cresko, Angel Amores, Catherine Wilson, Joy Murphy, Mark Currey, Patrick Phillips, Michael A. Bell, Charles B. Kimmel, and John H. Postlethwait. Parallel genetic basis for repeated evolution of armor loss in alaskan threespine stickleback populations. *PNAS*, 101(16):6050–6055, April 2004. ISSN 0027-8424, 1091-6490.

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

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

Korinna Esfeld, Andrea E. Berardi, Michel Moser, Eligio Bossolini, Loreta Freitas, and Cris Kuhlemeier. Pseudogenization and resurrection of a speciation gene. *Current Biology*, 28(23):3776–3786.e7, December 2018. ISSN 0960-9822.

Thilo C Fischer, Heidrun Halbwirth, Barbara Meisel, Karl Stich, and Gert Forkmann. Molecular cloning, substrate specificity of the functionally expressed dihydroflavonol 4-reductases from malus domestica and pyrus communis cultivars and the consequences for flavonoid metabolism. *Archives of Biochemistry and Biophysics*, 412(2):223–230, April 2003. ISSN 0003-9861.

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

Nicolas Gompel, Benjamin Prud'homme, Patricia J. Wittkopp, Victoria A. Kassner, and Sean B. Carroll. Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature*, 433(7025):481–487, February 2005. ISSN 1476-4687.

Marian Groenenboom, Victoria Gomez-Roldan, Hans Stigter, Laura Astola, Raymond van Daelen, Jules Beekwilder, Arnaud Bovy, Robert Hall, and Jaap Molenaar. The flavonoid pathway in tomato seedlings: Transcript abundance and the modeling of metabolite dynamics. *PLOS ONE*, 8(7):e68960, July 2013. ISSN 1932-6203.

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

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

Heidrun Halbwirth, Stefan Kahl, Walter Jäger, Gottfried Reznicek, Gert Forkmann, and Karl Stich. 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, March 2006. ISSN 0168-9452.

David Heckmann, Stefanie Schulze, Alisandra Denton, Udo Gowik, Peter Westhoff, Andreas P. M. Weber, and Martin J. Lercher. 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.

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

Marcel H. N. Hoefnagel, Marjo J. C. Starrenburg, Dirk E. Martens, Jeroen Hugenholtz, Michiel Kleerebezem, Iris I. Van Swam, Roger Bongers, Hans V. Westerhoff, and Jacky L. Snoep. Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. *Microbiology*, 148(4):1003–1013, 2002.

Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Jürgen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes, and Ursula Kummer. Copasi—a complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, December 2006. ISSN 1367-4803.

M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro,

T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, March 2003. ISSN 1367-4803.

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

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

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

Kenjiro Katsu, Rintaro Suzuki, Wataru Tsuchiya, Noritoshi Inagaki, Toshimasa Yamazaki, Tomomi Hisano, Yasuo Yasui, Toshiyuki Komori, Motoyuki Koshio, Seiji Kubota, Amanda R. Walker, Kiyoshi Furukawa, and Katsuhiro Matsui. A new buckwheat dihydroflavonol 4-reductase (dfr), with a unique substrate binding structure, has altered substrate specificity. *BMC Plant Biology*, 17(1):239, December 2017. ISSN 1471-2229.

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

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

Meihua Christina Kuang, Jacek Kominek, William G. Alexander, Jan-Fang Cheng, Russell L. Wrobel, and Chris Todd Hittinger. Repeated cis-regulatory tuning of a metabolic bottleneck gene during evolution. *Mol Biol Evol*, 35(8):1968–1981, August 2018. ISSN 0737-4038.

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

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

Hugo Vicente Miranda, António E. N. Ferreira, Alexandre Quintas, Carlos Cordeiro, and Ana Ponces Freire. Measuring intracellular enzyme concentrations: Assessing the effect of oxidative stress on the amount of glyoxalase i. *Biochem. Mol. Biol. Educ.*, 36(2):135–138, March 2008. ISSN 14708175, 15393429.

Natsuki Miyagawa, Taira Miyahara, Mitsutoshi Okamoto, Yukio Hirose, Kimitoshi Sakaguchi, Shoji Hatano, and Yoshihiro Ozeki. 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.

Joseph Mol, Erich Grotewold, and Ronald Koes. How genes paint flowers and seeds. *Trends in Plant Science*, 3(6):212–217, June 1998. ISSN 1360-1385.

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

Erin S. Morrison and Alexander V. Badyaev. Structuring evolution: biochemical networks and metabolic diversification in birds. *BMC Evolutionary Biology*, 16(1):168, August 2016. ISSN 1471-2148.

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

Hojung Nam, Tom Conrad, and Nathan E. Lewis. Metabolic pathway evolution and cellular objectives. *Curr Opin Biotechnol*, 22(4):595–600, August 2011. ISSN 0958-1669.

Julienne Ng and Stacey D. Smith. How to make a red flower: the combinatorial effect of pigments. *AoB PLANTS*, 8, January 2016a.

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

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

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

Silvio Panettieri, Erisa Gjinaj, George John, and David J. Lohman. Different ommochrome pigment mixtures enable sexually dimorphic batesian mimicry in disjunct populations of the common palmfly butterfly, elymnias hypermnestra. *PLOS ONE*, 13(9):e0202465, September 2018. ISSN 1932-6203.

Elisa Petrussa, Enrico Braidot, Marco Zancani, Carlo Peresson, Alberto Bertolini, Sonia Patui, and Angelo Vianello. Plant flavonoids—biosynthesis, transport and involvement in stress responses. *Int J Mol Sci*, 14(7):14950–14973, July 2013. ISSN 1422-0067.

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

Frantisek Poustka, Niloufer G. Irani, Antje Feller, Yuhua Lu, Lucille Pourcel, Kenneth Frame, and Erich Grotewold. A trafficking pathway for anthocyanins overlaps with the endoplasmic reticulum-to-vacuole protein-sorting route in arabidopsis and contributes to the formation of vacuolar inclusions. *Plant Physiol*, 145(4):1323–1335, December 2007. ISSN 0032-0889.

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

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

Ken G. Ryan, Kenneth R. Markham, Stephen J. Bloor, J. Marie Bradley, Kevin A. Mitchell, and Brian R. Jordan. 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.

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

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

Sascha Schäuble, Anne Kristin Stavrum, Pål Puntervoll, Stefan Schuster, and Ines Heiland. Effect of substrate competition in kinetic models of metabolic networks. *FEBS Letters*, 587(17):2818–2824, September 2013. ISSN 0014-5793.

Ida Schomburg, Antje Chang, Oliver Hofmann, Christian Ebeling, Frank Ehrentreich, and Dietmar Schomburg. Brenda: a resource for enzyme data and metabolic information. *Trends in Biochemical Sciences*, 27(1):54–56, January 2002. ISSN 0968-0004.

Stefan Schuster, Thomas Dandekar, and David A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology*, 17(2):53–60, February 1999. ISSN 0167-7799.

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

Y. Shimada, M. Ohbayashi, R. Nakano-Shimada, Y. Okinaka, S. Kiyokawa, and Y. Kikuchi. 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.

Lucian P. Smith, Frank T. Bergmann, Deepak Chandran, and Herbert M. Sauro. Antimony: a modular model definition language. *Bioinformatics*, 25(18):2452–2454, September 2009. ISSN 1367-4803.

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

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

Endre T. Somogyi, Jean-Marie Bouteiller, James A. Glazier, Matthias König, J. Kyle Medley, Maciej H. Swat, and Herbert M. Sauro. libroadrunner: a high performance sbml simulation and analysis library. *Bioinformatics*, 31(20):3315–3321, October 2015. ISSN 1367-4803.

Gregory Stephanopoulos. Metabolic fluxes and metabolic engineering. *Metabolic Engineering*, 1(1): 1–11, January 1999. ISSN 1096-7176.

Matthew A. Streisfeld and Mark D. Rausher. 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, February 2009a. ISSN 0737-4038.

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

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

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

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

D. S. Treves, S. Manning, and J. Adams. 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.

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

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

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

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

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

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

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

Jennifer H. Wisecaver, Jason C. Slot, and Antonis Rokas. The evolution of fungal metabolic pathways. *PLOS Genetics*, 10(12):e1004816, December 2014. ISSN 1553-7404.

Patricia J. Wittkopp, Sean B. Carroll, and Artyom Kopp. Evolution in black and white: genetic control of pigment patterns in drosophila. *Trends in Genetics*, 19(9):495–504, September 2003. ISSN 0168-9525.

Robert Woods, Dominique Schneider, Cynthia L. Winkworth, Margaret A. Riley, and Richard E. Lenski. 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.

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

Yao-Wu Yuan, Alexandra B. Rebocho, Janelle M. Sagawa, Lauren E. Stanley, and Harvey D. Bradshaw. Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between mimulus species. *PNAS*, 113(9):2448–2453, March 2016. ISSN 0027-8424, 1091-6490.

Linlin Zhang, Arnaud Martin, Michael W. Perry, Karin R. L. van der Burg, Yuji Matsuoka, Antónia Monteiro, and Robert D. Reed. Genetic basis of melanin pigmentation in butterfly wings. *Genetics*, 205(4):1537–1550, April 2017. ISSN 0016-6731.

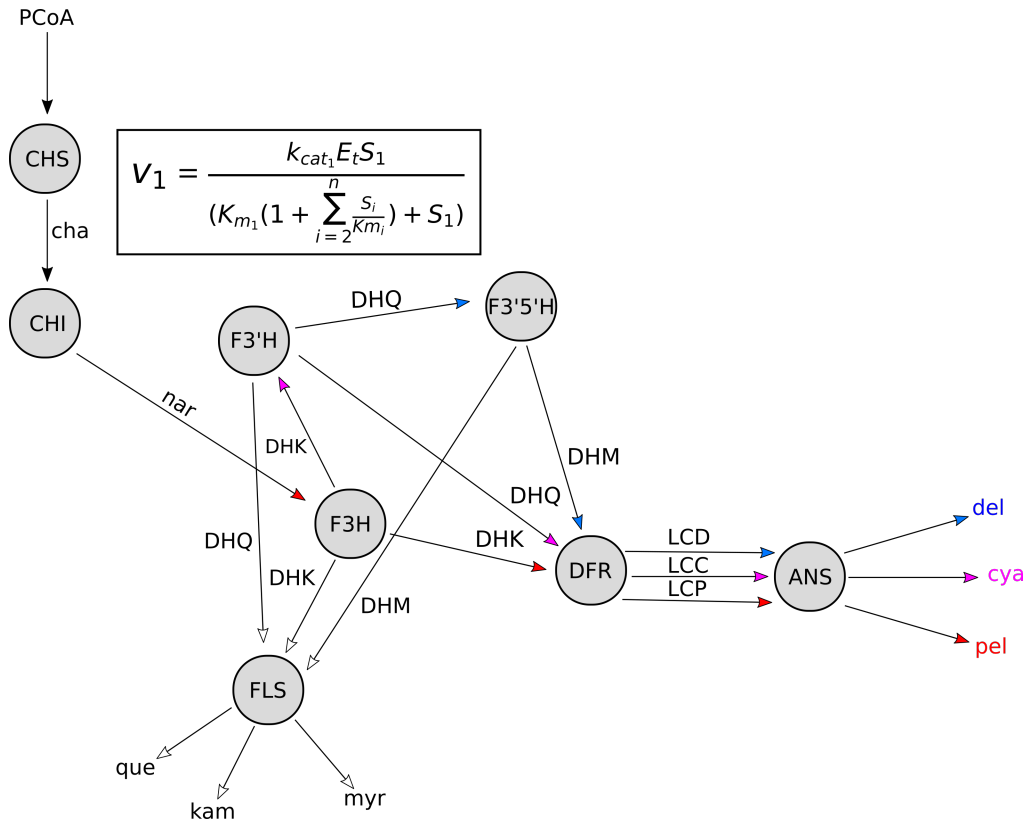

Fig 1. Diagram of the simplified anthocyanin pathway model. The simplified model is depicted as an enzyme-centric pathway diagram, including the key anthocyanin and flavonol branches (see Fig. S1 for classical substrate-centric diagram). Grey circles are enzymes. Arrows indicate influx and outflux of 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).

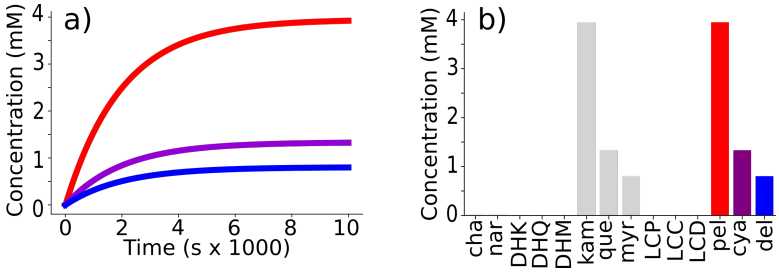

Fig 2. Kinetic behavior of the naïve starting state model. a) time-course simulation of this model showing anthocyanidin concentrations over time: pelargondin (red), cyanidin (purple), and delphinidin (blue). b) Steady state concentrations of each floating species in the naïve model. The intermediate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

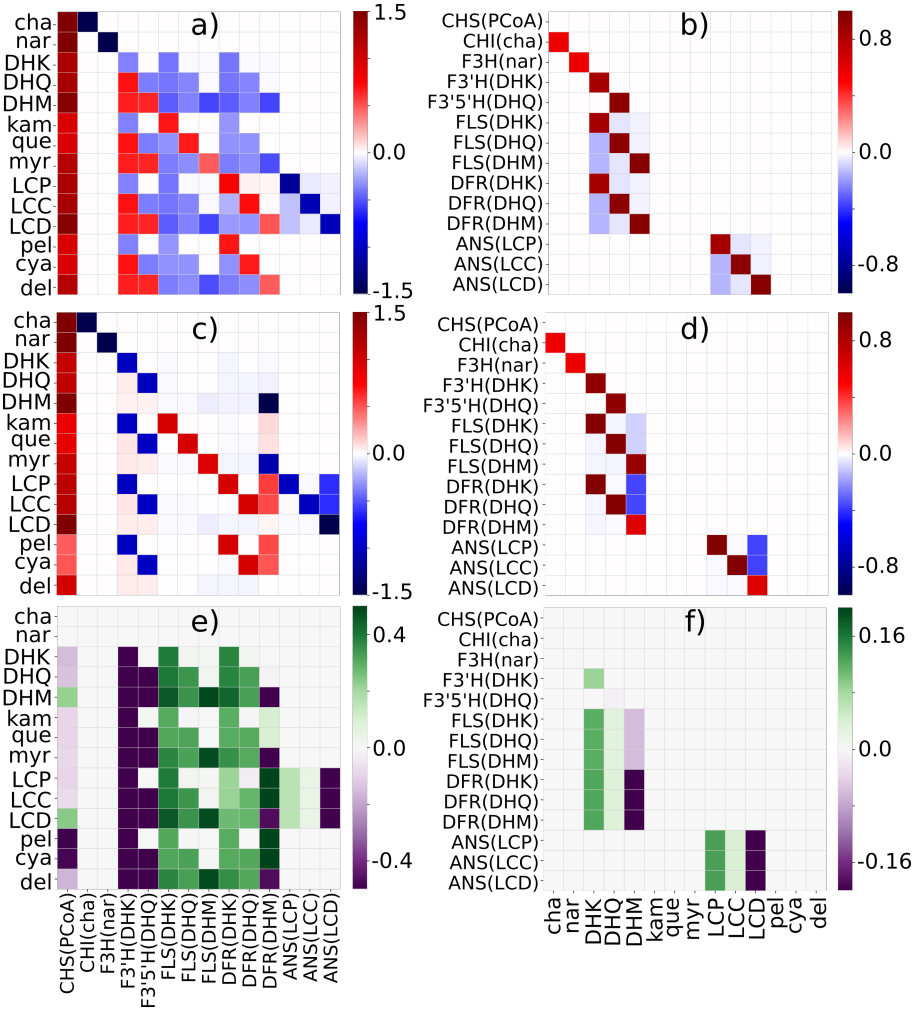

Fig 3. Pathway control shifts to acheive a bias toward the delphinidin branch. a) Heatmap of concentration control coefficients for all pathway reactions (except for "sinks" at boundary conditions) in the naïve starting state model. b) Heatmap of elasticity matrix for all pathway reactions (except for 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.

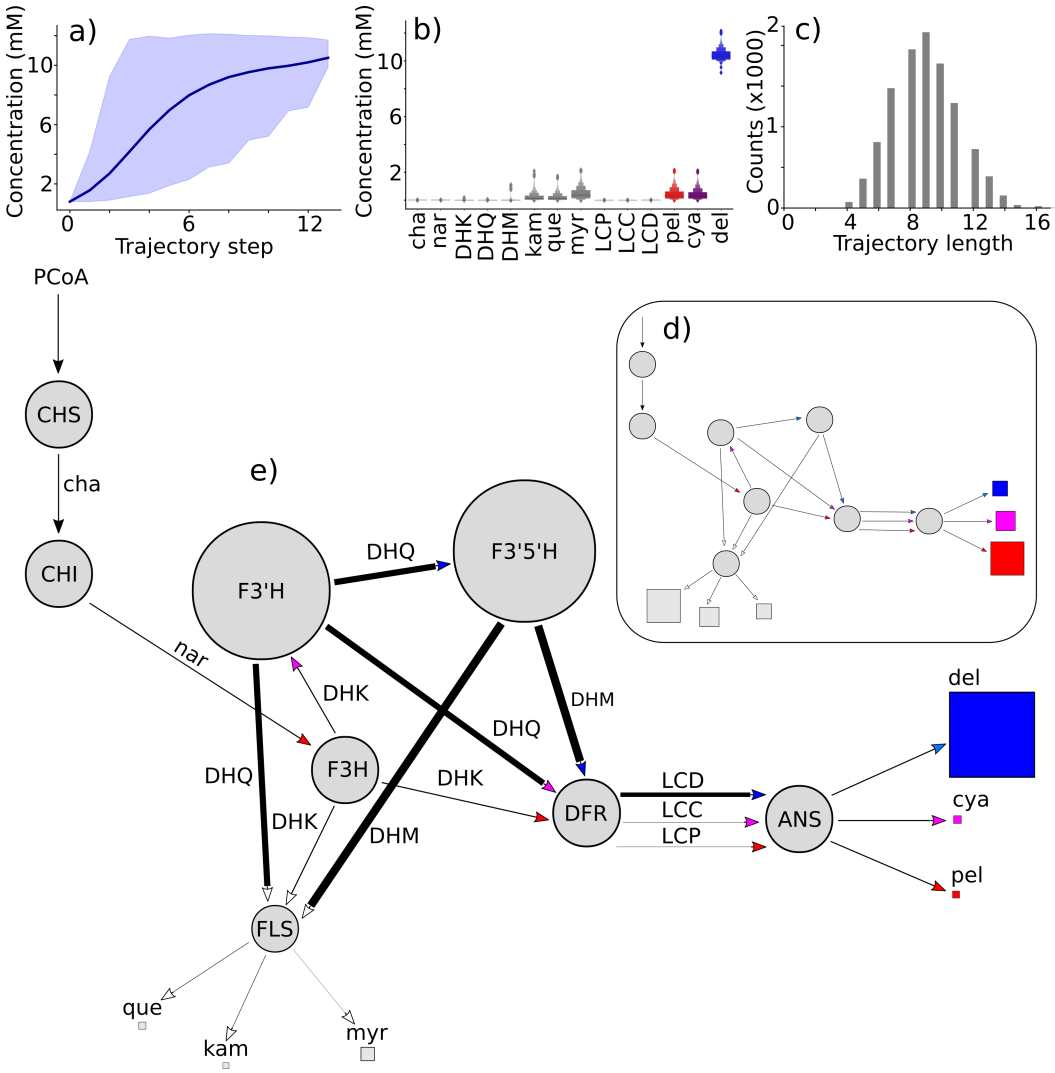

Fig 4. Simulated evolution of delphinidin production. a) Trajectories through delphinidin space. Dark blue line shows mean trajectory, averaged across all 9,985 simulations. Shaded blue area shows envelope containing 95% of trajectories. Top limit is maximum value at each step, bottom limit is minimum 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.

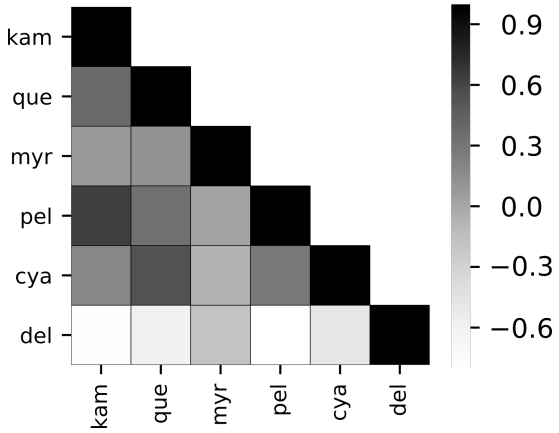

Fig 5. Pleiotropic trade-offs in pathway products. Heatmap of mean Spearman correlations between concentrations of anthocyanins (pel, del, and cya) and flavonols (kam, que, myr) across all 9,985 evolutionarytrajectories. A full correlation matrix for all floating species and correlation matrices for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