Computational modeling of anthocyanin pathway evolution

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 key component of 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 pigmentaion 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 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 examine whether the observed trajectories and constraints help to explain 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 which 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 acitivity of metabolic pathways [1–5]. 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 [6–10]. Evolution of metabolic pathways is known to be shaped by both the topological structure of pathways and the biochemical constraints of the individual enzymatic components [4, 10, 11]. Furthermore, regulatory architecture governing pathway gene expression plays a vital role in the evolution of pathway activity and the resulting alterations in phenotype [10, 12–18]. These properties make metabolic pathways excellent systems to study the fundamental principles that underly the genotype-phenotype map of complex phenotypes.

A large body of theoretical work has been devoted to understanding the control of metabolic pathways by topology and enzyme properties. Metabolic control analysis arose from classical enzyme kinetics and solved the problem of analyzing entire metabolic systems simultaneously, rather than focusing on individual components in isolation [19–21]. These methods have been widely applied to the analysis of empirical data and to problems in metabolic engineering [2, 22–24]. However, application of these theoretical approaches to the study of metabolic pathway evolution has been sparse. A few studies within the last decade have used simulation-based approaches to study the basic principles governing the evolution of extremely simple pathway models [25, 26]. These studies revealed important patterns, such as the disproportionate occurrence of large-effect beneficial mutations at enzymes with the most control over pathway flux, and the focus of metabolic flux control at upstream and branching enzymes. However, it remains unclear if these patterns extend into more complex and biologically-realistic pathways or if greater complexity introduces additional constraints on pathway evolution. Highly branched pathways may be subject to strong trade-offs arising from competition for substrates and enzymes [11, 27, 28]. Several studies have probed the behavior of more biologically-realistic pathway models using similar modelling approaches [8, 9], but only recently have studies began to probe the evolutionary dynamics of pathways during phenotypic transitions [29], and these have dealt with simple linear pathways. The interplay between various aspects of pathway structure may result in complex mutational fitness landscapes characterized by inaccessible regions of phenotype space and therefore greater predictability in evolutionary trajectories.

Although theory has not 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 [5, 30], various metabolic pathways in yeast [31] and Drosophila [7], and anthocyanin biosynthesis in flowering plants [32–34], have yielded insights into the molecular mechanisms and systems-level constraints on evolving metabolic pathways. For example, causative mutations can be drawn from several broad classes such as regulatory vs. biochemical (structural) and upstream vs. downstream [32–37]. These classes tend to be differentially favored during evolutionary trajectories and can result in the emergence of detectable "hotspot" loci that recur in recapitulated instances of the same phenotypic transitions [38–43]. Pleiotropic interactions have commonly been suggested as an explanation for these patterns [15, 33, 37]. However, other factors such as target size and chromosomal location, may also play a role [44]. Determining how pathway structure gives rise to these types of predictable patterns is critical for understanding the evolution of the phenotypes that result from metabolic activity [35, 37, 45].

Flower color has proven to be a particularly useful model for studying the origin of novel phenotypes through metabolic evolution. Color is determined predominantly by the presence of various pigment compounds, the most common of which are produced by the highly conserved anthocyanin biosynthesis pathway (Fig. 1). 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 [46]. The anthocyanin pathway consists of a series of branching enzymes that give rise to three principal classes of red, purple, and blue pigments (Fig. 1). Further downstream modifications of these basic building blocks result in the diversity of colorful pigments found across the flowering plants. Building on existing knowledge of this pathway, experimental work that spans several decades has shown that changes in floral anthocyanin pigmentation result from both biochemical and regulatory modifications of pathway components, often occurring at branching enzymes [33, 47]. Based on the available data, these loci appear to be evolutionary hotspots, repeatedly targeted by mutations independently in multiple lineages. In particular, regulatory mutations at branching loci have been responsible for many large-effect mutations observed in nature [33–35, 37, 43, 48]. Biochemical modifications of downstream pathway enzymes have also been observed, apparently occurring to optimize the transitioning phenotypes [34, 36]. The observation of repeated fixation of mutations at certain pathway genes suggests a certain level of predictability in flower color evolution that could potentially be explained by the pathway structure.

Larger macro-evolutionary studies have also uncovered various patterns in the evolution of anthocyanin flower coloration. Certain phenotypic states, such as the joint production of red and blue pigments, are not commonly observed in nature [49–51]. Transition rates between differently-pigmented states, such as blue to red, display a high degree of asymmetry [51]. The complex branching structure of the anthocyanin pathway makes it difficult to intuitively rationalize some of these experimental observations. The possibility of evolutionary trade-offs due to competition for shared substrates, inaccessible regions in the state space of the pathway, and non-linear relationships between mutational effects at different loci, make it even more difficult to quantitatively predict the impact of mutations in pathway genes. We lack a unifying model that relates pathway function and activity to the evolution of distinct pigment phenotypes, which makes it impossible to construct a comprehensive picture of the mutational landscape underlying evolution of the pathway. Experimental observations appear to be consistent with theoretical predictions from simple models, such as the apparent prevalence of fixed mutations at branch-point enzymes [26]. However, neither the available empirical data nor the currently existing mathematical models are sufficient to draw broad conclusions or inform our expectations for complex biological pathways.

In the present study, we aimed to use a realistic model of a biological metabolic pathway to predict the predominant mutations that contribute to shifting flux in branching pathways. We hypothesized that enzymes capable of exerting differential control over pathway flux would be the dominant contributors to evolution between phenotypes and that strong trade-offs would arise due to competition between these effectors. In order to test these hypotheses and predict the effects of shifts in enzyme function on pigment production, we construct a simplified kinetic model that mirrors the structure and properties of the anthocyanin biosynthesis pathway. We then stochastically evolve this pathway model from an unbiased starting point toward a defined optimum to address four key questions related to general principles of metabolic pathway evolution: 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?* We hypothesized that the control structure of the pathway would shift during evolution to facilitate the new phenotypic optimum. We further hypothesized that certain enzymes in the pathway would represent evolutionary "hotspots", while others would tend not to contribute to evolution of pigment composition. We predicted that these primary targets of selection would be enzymes with the greatest ability to exert differential control over flux down pathway branches. Here, we quantify the inherent behavior of the anthocyanin pathway topology, identify systemic constraints on parameter modifications and pigment concentration, and quantify relative contributions of individual pathway components to evolution between phenotypes. Our results indicate that there are many parameter configurations resulting in stable kinetic behavior of the pathway model and a large number of accessible trajectories connecting phenotypic optima. Our analysis of trajectories between defined state space optima reveal key mutational targets, which are the dominant contributors to transitions between pigment phenotypes. This simulation approach informs our expectations for the evolution of the anthocyanin pathway, highlights the key properties that defined pathway evolution and lays the groundwork for understanding other enzymatic systems with branching topologies.

Methods

Development and computational implementation of the mathematical pathway model

We used the generalized rate law formulation specified in [52, 53] to specify rate laws for each enzymatic reaction in the pathway model (Supplemental text). This rate law form scales the Michaelis constant (K_M, which is related to the 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 [52, 53]. We chose to use irreversible Michaelis-Menten kinetics, because *in vivo* measurments flavonoid pathway activity are fit well by an irreversible model [54] (but, see [55, 56]). Moreover, the irreversible form of the model reduces the required number of parameters two-fold relative to the completely reversible equivalent [53].

Given that detailed kinetic studies are lacking for most anthocyanin pathway enzymes, we instead designed a naive 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 meta-analysis [57] of enzyme properties performed by mining the BRENDA and KEGG databases [57–59]. The model was initalized at a starting state with all parameters of a certain type set to be equal. Values for K_{cat} ($14\ s^{-1}$) were set to the median of the distribution calculated in the Barr-Even et al. meta-analysis [57]. K_M values ($0.013mM$) were set to 10 times 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. Enzyme concentrations were initialized to $0.0001mM$, a value reasonably in the range of naturally-occurring protein concentrations [60]. This model represents a naive, 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 (0.01mM) 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 synthsesized, which is consistent with the physiological transport of anthocyanin pigments to the vacuole [61, 62].

We implemented this mathematical model of the anthocyanin pathway in Python 3.6 using the Tellurium library [63] and the Antimony [64] 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 [65]. To confirm that time-course and steady state values calculated by Tellurium were consistent with other standard methods, the model was written out in SBML ("supplemental-file-1.sbml") [66] and the simulations were then run in COPASI to confirm results [67]. 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. For example, the Roadrunner method used to run a time-course simulation can be run directly on the model attribute of the Pathway object.

Evolutionary simulations proceed as follows: 1) A Roadrunner model object is initialized from a user-defined Antimony string and held as an attribute of the Pathway object. 2) Steady state concentrations for all floating species in the initial model are calculated. 3) A relative fitness for the current model is calculated compared to the user-defined optimum state. 4) The model enters a loop of user-defined length. 5) For each loop iteration, a single randomly-drawn parameter (from a specified list of available parameters) in the model is randomly mutated by multiplying the current parameter value by a value drawn from a gamma distribution with $\alpha = 0.8$ and $\beta = 3$, selected to make the total number of negative mutations approximately equal to the total number of positive mutations. Any mutation driving total steady state concentration of all floating species outside of a defined tolerance on the starting value (an argument to the *evolve* function) are discarded as non-functional, along with any mutants resulting in lack of a steady state solution. 6) Steady state concentrations for the mutant model are computed and used to calculate fitness [25, 26, 68]. This fitness value is used to calculate a selection coefficient (s) as the difference between the current and previous states. 7) For beneficial mutations ($s > 0$), the selection coefficient is used to calculate a fixation probability via a simplified expression ($1 - e^{-s}$), in which population size is ignored for simplicity. Fixation of the mutation is then determined randomly according the calculated fixation probability. Neutral ($s = 0$) and deleterious ($s < 0$) are discarded, because the fixation probabilities are neglible. This strategy speeds up the simulations by removing an extra step for a large number of mutations, which are more commonly deleterious than advantageous.

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 enzyme 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 change in activity on other substrates. While it might be biologically reasonable to assume some cost of improvement on one substrate for activity on other substrates [69, 70], we examined the kinetic data for all multi-substrate anthocyanin pathway enzymes in the BRENDA database [58] 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).

The iterative simulation process continues until either the specified number of iterations have occurred or the user-defined optimum is reached within a specified tolerance (default is 1%). Steady state concentrations at each simulation step, selection coefficients for each mutation (normalized to mutational size), 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 compressed *pickle* file.

A total of 9,985 independent evolutionary simulations were run between the *naive* starting state and a defined optimum representing a blue-flowered state. The optimum was defined as a delphinidin concentration of 90% of total steady state concentration with a 10% tolerance. All other values 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 in the *Pathway* object of *enzo* and can be modified to suite the needs of the user.

Analysis of the simulated data

The major goals of our analyses were 1) to characterize the intrinsic behavior of the anthocyanin pathway topology, 2) to determine how the pathway control structure shifts to facilitate movement toward the optimum phenotype, 3) to quantify which loci contribute the most to evolution between phenotypes, and 4) to examine whether competitive relationships between mutations result in pleiotropic trade-offs during pathway evolution.

To determine the intrinsic behavior of the anthocyanin pathway topology we first used the built-in methods of Tellurium to run time-course and steady state simulations of the naive starting state model. We then used metabolic control analysis (MCA) to assess how both the intrinsic control structure of the naive model and how this structure changed in the course of the evolutionary simulations. MCA is a body of theory developed specifically for analyzing the control of flux through metabolic pathways [19–21]. The standard approach is to calculate two key properties of each reaction in the pathway: 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 (also known as reaction order), which are sensitivity coefficients calculated by taking the partial derivative of each reaction rate with respect to each substrate concentration. Application of these methods allowed us to assess how the topology of the anthocyanin pathway determines the relative steady state concentrations of end products. To examine the patterns of control over floating species concentrations and the sensitivity of reactions to substrate concentration changes, we used the built-in MCA methods in Tellurium [63] to calculate matrices of concentration control coefficients (because we are interested in selection on steady state concentration of pigment compounds) and elasticities corresponding to all pathway reactions in the naive kinetic pathway model. After running the evolutionary simulations, we repeated this analysis on each evolved model (9,985 total) and calculated the median control and elasticity matrices. By subtracting these median matrices from those of the starting state, we calculated the shift in pathway control structure that occurred by evolving toward to the 90% delphinidin optimum.

To directly determine which loci contribute the most to evolution of the pathway toward the delphinidin optimum, we first directly counted the number of fixation events involving each model parameter across all 9,985 simulated trajectories. We then estimated the overall distributions of fitness effects (DFE) for each model parameter across all 9,985 replicate simulations by combining the dictionaries of selection coefficients logged by each *enzo* Pathway object.

To assess potential trade-offs arising from the pathway structure, we first quantified the relationships between the steady state concentrations of each floating species over the course of the simulations. We calculated Spearman correlation coefficients (ρ) between all pairs of floating species for each of the 9,985 simulated trajectories and then calculated the mean coefficient for each relationship. We next quantified the relationships between mutations in commonly fixed parameters by determining the directionality of shifts in each parameter at delphinidin-optimized states. We examined the mean behavior and distributions of evolved end-point parameter values across all 9,985 simulations. We then initialized a Roadrunner model with the mean-optimal parameter values to characterize the mean behavior of the evolved parameter sets, re-running the time-course and steady state calculations using Tellurium. Finally, we quantified the constraints that parameters place on one another during evolution, by calculating all Spearman correlation coefficients (ρ) for each pair of evolvable model parameters (excluding the static "*sink*" rate parameters) across all 9,985 simulated trajectories, and then computing the mean ρ for each relationship.

Results

Mathematical model reveals behavior of anthocyanin pathway topology

To assess the behavior of the anthocyanin pathway topology we first constructed a simplified kinetic model, represented as a series of linked, irreversible Michaelis-Menten reactions, initialized with a biologically-realistic set of parameters (see methods). Our naive starting state model, with all enzymes at equal concentrations 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). This steady state is characterized by a bias toward products of the first branches (leading to kampferol and pelargonidin); with pelargonidin, cyanidin, and delphinidin composing 32.7%, 10.9%, and 6.6%of the total steady state concentration respectively (Fig. 2b). This chemical concentration profile would likely result in red or pink floral coloration [34, 49], which is rare in nature compared to blue and purple coloration [50, 51]. This result suggests that evolution has directed pathway activity away from the outcome predicted from topology alone, i. e., the production of the pigment pelargonidin, which requires the fewest biochemical steps. The evolution of flowers with higher concentrations of cyanidin and delphinidin can presumably be acheived by a variety of modifications such as alteration of enzymes kinetics or changes in enzyme expression level. These mechanisms for modulating pathway flux are studied in detail in subsequent sections.

Metabolic control analysis of the naive starting state model indicated that concentration control is highly variable across the pathway (Fig. 2c). The first step in the pathway, catalyzing conversion of P-Coumaroyl-CoA to chalcone, has strong positive control over all downstream reactions and steady state concentrations. Control over downstream steady state concentrations is also largely focused in the core branching enzymes F3'H and F3'5'H (Fig. 1), which exhibit positive control over some products and negative control over others (Fig. 2c). The secondary branching enzyme DFR also exerts substantial control over the downstream anthocyanin and flavonol species. In contrast, the final enzyme in the pathway (ANS) exhibits little control over any floating species except its substrates. This observation is consistent with previously observed behavior of steps at the end of simple linear and branched pathways [e.g. [25, 26]], in which control is shifted to upstream nodes. The elasticities of the pathway also exhibit substantial variability. Most reactions are insenstitive to the concentrations of most floating species. However, there are hotspots that are consistent with intuitive expectations based on examining pathway topology. For example, reactions utilizing dihydrokampferol (DHK) as a substrate are highly positively sensitive to DHK concentration. Conversely, the reactions involving substrates that compete with DHK are negatively sensitive to DHK concentration (Fig. 2d). Because this branch leads to the red pelargonidin pigment, this control structure is consistent with our observation of bias down this branch in the steady state of the naive model. The topology of the pathway results in the "red" branch exerting a dampening effect on the branches with which it competes for upstream source material.

Replicated evolutionary simulations yield an envelope of viable trajectories

Our evolutionary simulations from the steady state in the naive model to the delphinidin production optimum (90% of total steady state concentration) revealed a high degree of variability in the intermediate phenotypes. We repeated 9,999 replicate evolutionary simulations under the same conditions to yield 9,999 trajectories from the naive starting state to the delphinidin-optimized end-point. Of these trajectories, twenty-four experienced some kind of numerical error or failed to reach an optimum in the alotted number of mutational attempts. These errors were detected by the built-in exception-catching scheme of *enzo* and subsequently discarded from the analysis to yield a final set of 9,985 good trajectories. These trajectories lie within a defined envelope when plotted in delphinidin space, but are nonethless highly variable (Fig. 3a). Variability in trajectories through anthocyanin space is also evident in the distributions of simulation end-point steady state concentrations (Fig. 3b) and the trajectories mapped in the pigment spaces of other floating species (Fig. S1). Because of the 10% tolerance level imposed on the 50% delphindin optimum, there is an small, expected degree of variation around delphinidin concentration. However, the other anthocyanins (pelargonidin and cyanidin) and the flavonol compounds (kampferol, quercetin, and myricetin) have very broad distributions of end-point steady state concentrations, ranging from 0% to approximately 10% of total steady state concentration (Fig. 3b). Of the 9,985 end-points 5.6% contain greater than or equal to 10% pelgardonidin and 4.4% contain greater than or equal to 10% cyanidin suggesting that the pathway topology is capable of acheiving a diversity of states within the imposed evolutionary constraints.

Pathway control shifts to bias activity toward the delphindin branch

Our comparison of control coefficients and elasticites between the naive starting state and the median of the evolved models revealed the patterns that facilitate evolution to the new delphinidin optimum. As predicted, pathway control and elasticity shifted away from the starting model. The control coefficients of the first step in the pathway (CHS conversion of P-Coumaroyl-CoA to chalcone) remained positive for all downstream reactions (Fig. 4a). However, these coefficients tended to become slightly more positive for reactions leading toward delphinidin production and slightly less positive for those directing flux toward other branches. The components of the delphinidin branch gained additional control over the reactions that involved competing substrates or lie on competing branches (Fig. 4a). Similarly, the elasiticities of these competing reactions became more negatively sensitive to the concentration of delphinidin precursors, particularly the branching precursor DHM (Fig. 4b). Large shifts were focused on reactions involving the branching enzymes F3'H, F3'5'H, as well as the downstream enzyme DFR, which directly acts on the branching precursors leading to anthocyanidin synthesis. The observed lack of change in the upstream steps relative to branching steps (and those experiencing substrate competition) was consistent with our prediction that enzymes capable of exerting differential control over flux down pathway branches would the predominant drivers of the phenotypic transition. Overall, the patterns of control and elasticity remained largely similar for most of the pathway, reflecting the ability of pathway flux to be modulated by quantitative changes occurring at certain loci.

Evolution of the pathway between optima reveals key mutational targets

As predicted, mutations selected during the evolutionary simulations were concentrated in a few 'hotspot' loci, indicating a certain degree of predictability in the evolution of the pathway. These changes were focused in enzymes capable of exerting differential control of pathway flux, such as the branching enzymes (F3'H, F3'5'H) and those subject to substrate competition (i.e. DFR). We counted the number of fixation events involving each model parameter across all 9,985 simulated trajectories, finding that greater than 95% of fixed mutations occur in the parameters of F3'H, F3'5'H, DFR, or FLS (Fig. S2a), which is consistent with our predictions and with the observed alterations to the pathway control structure.

We combined distributions of fitness effects (DFE) (selection coefficients for each mutation) for each model parameter across all 9,985 replicate simulations to estimate the overall DFE for each of these parameters (Fig. S2a). There are striking differences between the DFE for each parameter, with broad similarities observed between parameters of the same pathway enzymes (Fig. S2b). Across the data, selection coefficients were more commonly negative (deleterious) and thus most mutations were rejected during the evolutionary simulations. For certain parameters, such as those pertaining to the final step enzyme ANS, the distributions were almost entirely concentrated around zero. This effect is likely due to the limited pathway control exhibited by ANS (Fig. 2d), resulting in an extreme rarity of mutations with substantial fitness effect size in either direction. Upstream enzyme mutations, which generally lack the ability to exert substantial differential control over flux down the various branches, were similarly concentrated around zero. In contrast, for the parameters belonging to the frequently-fixed enzymes F3'H, F3'5'H, DFR, and FLS there exist long positive tails in the DFE (Fig. S2b), from which mutations with reasonably high selection coefficients could be fixed. The median selection coefficient for fixed mutations across all 9,985 trajectories was 0.044, with the smallest fixed mutation having $s = 2.4x10^{-6}$ and the largest having $s = 0.43$.

Enzyme and substrate competition result in pleiotropic tradeoffs

Evolution toward the delphinin optimum resulted in sharp and predictable trade-offs in the production of other pathway products. Shifts in the values of parameters for each enzyme were largely consistent in their directionality across all 9,985 simulations (Fig. 6; Fig. S2c,d,e). Large shifts occurred in the values of parameters with positive DFE tails (Fig. S2b), which act as major contributors during pathway evolution (Fig. S2). For example, the K_{cat} and K_M values for DFR acting on the blue precursor DHM were substantially improved in almost all simulations, whereas the K_{cat} and K_M values for competing subtrates of DFR tended to be drastically weakened, reflecting an evolved preference of the enzyme for the blue precursor. Activity of the first branching enzymes F3'H and F3'5'H, were also strongly increased relative to the starting point, reflecting the concerted re-direction of flux toward the blue delphinidin branch(Fig. 6; Fig. S2c,d,e). Meanwhile, the activity of FLS, which competes for pigment precursor substrate pools with DFR, was consistently reduced (Fig. 6; Fig. S2c,d,e), consistent with redirection of flux away from the flavonol branch and toward the anthocyanin branch. In contrast, there was virtually no shift away from the starting state model in parameters belonging to enzymes with DFEs concentrated around zero, such as the final enzyme ANS (Fig. 6; Fig. S2c,d,e). In contrast, there was virtually no shift away from the starting state model in parameters belonging to enzymes with DFEs concentrated around zero, such as the final enzyme ANS (Fig. 6; Fig. S2c,d,e). When we initialized a model with mean-optimal parameter values, taken over all trajectories, it yielded a steady state wherein delphinidin composed 95% of the total concentration, slightly overshooting the defined optimum (Fig. 6).

Mean Spearman correlation coefficients (ρ, see methods) revealed the relationships between the steady state concentrations of floating species (Fig. 6a). There are many tight correlations between steady state concentrations of floating species during evolution between states. Expected relationships, such as the positive correlation between precursor production and the production of their resulting reaction products, were reflected in the data (Fig. 6a). For example, the blue precursor dihydromyricetin (DHM) exhibits positive correlations with both downstream products that are derived from it: myricetin and delphinidin (Fig.6a). On average, the red precursor dihydrokampferol (DHK) is negatively correlated with myricetin, delphinidin, and the DHM precursor (Fig.6a), despite the fact that DHK is converted to DHQ by F3'H and then to the blue precursor DHM by F3'5'H (Fig. 1). This results suggests that the use of DHK as a precursor to the pelardonidin branch products (Fig.1; Fig. 6a) imposes a sufficiently large constraint on production of DHM to result in a detectable trade-off.

The interplay between the production of anthocyanin species and flavonol species was also detected in the correlation analysis. For example, the steady state concentration of delphinidin is negatively correlated with those of kampferol, quercetin, and myricetin (Fig.6a). This result demonstrates that there is a trade-off between production of the flavonols and optimization for production of delphinidin, with the strongest constraint on the concentration of the first-branch product, kampferol. We observed similar relationships within the anthocyanin group. For example, delphinidin production is negatively correlated with both pelargonidin (Fig. 6a) and cyanidin production (Fig. 6a). In contrast, pelargonidin and cyanidin are less strongly coupled and postively correlated (Fig. 6a), which demonstrates the presence of asymmetric relationships under selection for increased production along a single branch. Asymmetric constraints may help to explain the distribution of pigment phenotypes seen in nature, such as the rarity of states that produce both pelargonidin and delphinidin [50]. These results highlight the trade-offs between floating species that are inherent to the topology of the model and serve to shape evolution between phenotypic states when total flux is constrained within a certain tolerance.

The parameters of the pathway model place tight constraints on one another during evolution between phenotypic states, which is reflected in the above analysis of directional shifts. Mean Spearman correlation coefficients (ρ, see methods) for each pair of evolvable parameters (Fig. 6b) provide a more detailed look at the clear pattern of positive and negative relationships amongst the parameters of the model as they evolve (Fig. 6b). These patterns expose the inherent tradeoffs present in the pathway as it evolves. For example, the K_{cat} value of F3'H for DHK is negatively correlated with the K_{cat} values of FLS for all three prescursors (Fig. 6b) and an almost identical pattern is seen for the relationship between the K_{cat} of F3'5'H for DHK with these same FLS K_{cat}

values (Fig. 6b). In general, this antagonistic relationship between FLS activity and delphinidin production is reflected in correlations across the parameter set. For example, the K_{cat} value of DFR for DHM is also negatively correlated with the FLS precursor K_{cat} values (Fig. 6b; DHK). These relationships are intuitive upon inspection of the pathway topology, because both the F3'H and F3'5'H reactions are essential for directing pathway flux toward the blue branch, while the FLS reactions draws flux away (Fig. 1). Likewise, DFR activity on the DHM precursor is essential for the effective production of delphinidin by allowing the blue branch to compete with FLS for precursor material. Other antagonistic relationships between parameters are detectable, such as the negative correlation between the K_{cat} value of DFR for DHM and the K_{cat} values of DFR for the red and purple precursors (Fig. 6b), which reflects the inherent competition between the three precursors for binding to DFR (Fig. 1). However, synergistic relationships between parameters are also detected in this analysis. For example, the K_{cat} value of DFR for DHM is positively correlated with the K_{cat} values of F3'H for DHK and F3'5'h for DHQ (Fig. 6b). A similar pattern is seen for the K_M value of DFR for DHM, showing that as the activity of F3'H and F3'5'H increases the K_M decreases (binding becomes tighter). These observations indicate that as activity of these first branching enzymes increases to direct flux toward the blue branch, the activity of the downstream DFR on the blue precursor DHM also increases to keep pace.

Discussion and Conclusions

A computational model of the anthocyanin pathway provides a framework to study evolution

Transitions between flower color phenotypes have happened repeatedly during angiosperm evolution [51, 71, 72]] and these have often hinged on alterations to genes of the anthocyanin pigment pathway [32, 35, 37, 47, 48]. Here we developed a mathematical model descriding the dynamics of anthocyanin pathway activity and used it to probe the intrinsic behavior of the pathway. Using realistic values drawn from empirical distributions of enzyme kinetic parameters, we parameterized a “naive” starting state model that is biased only by the topology of the pathway. We found that this naive model exhibited stable dynamics, yielding a steady state solution that was characterized by inherent bias toward the red-pigment branch. This observation is not surprising, based on the pathway topology, because pelardonidin (and its flavonol counterpart; kampferol) is produced at the first branch point in the pathway (Fig. 1). However, in natural systems red flowers are comparatively rare [50], which indicates that there are factors beyond pathway structure that shape the distribution of flower colors across taxa. Indeed, it has previously been hypothesized that red flowers are an “evolutionary dead-end”, a state that is difficult to evolve out of once it has been entered, due the nature of mutations that lead to red floral pigmentation [50, 51]. Furthermore, there is likely to be selection on other products of the flavonoid pathway, such as the flavonols quercetin, kampferol, and myricetin. These compounds are known to act as sunscreens that protect plant tissues from UV damage [73–75]. Since there is a clear coupling between the anthocyanidins and the flavonols due to competition for shared substrates, simultaneous selection on both classes of compounds might impose some of the additional constraints necessary to acheive the observed distribution of flower colors in nature.

Evolution of the pathway model reveals major mutational targets and alterations to pathway control

To develop an intuition for the evolutionary properties of the anthocyanin pathway, we computationally evolved our mathematical model from the naive starting state toward an optimum phenotypic state defined by the proportion of delphinidin pigment produced at steady state (90%). While the starting point represents a red-flowered state in a real biological system (producing predominantly pelargonidin), the pre-defined optimum would result instead in blue flowers [34, 49, 76, 77]. Transitions between blue and red flower color have been repeatedly observed in nature, and are are known to involve both loss-of-function and function-swiching mutations [32–35, 37, 47, 48, 78], making this type of transition a rich example for understanding the mechanisms by which the phenotype can be altered.

Our replicated evolutionary simulations demonstrated the wide breadth of possible trajectories through pigment space that can be taken to reach the pre-defined delphinidin optimum. The other anthocyanins (pelargonidin and cyanidin) as well as the flavonols (kampferol, quercetin, and myricetin) are able to wander substantially when selecting on delphindin concentration. This is reflected in the broad distributions of end state concentrations as well as the trajectories through these pigment dimensions (Fig. S1). This result suggests that evolutionary trajectories in real biological systems could also potentially be highly variable. Experiments using directed evolution of the anthocyanin pathway in a model organism might help to reveal the biological evolutionary landscape.

One of the key observations in this study was the shifting of pathway control during evolution of the pathway model. Through a series of subtle changes the blue branch (leading toward delphindin) evolved to exert greater control over both competing branches and competing reactions on the same branch. Meanwhile, the reactions leading the delphiniding became less senstitive to the actions of competing reactions and branches. Strikingly, these processes resulted in a quantitative weighting of pathway activity, which achieved bias toward delphinidin production, without any dramatic qualitative changes to the pathway control structure. This result suggests that the pathway activity can be tuned dynamically by evolution to result in phenotypic changes.

In our replicate simulations we observed key genes that are the dominant contributors to evolution of the optimum phenotype. As we hypothesized, the branching enzymes (F3'H and F3'5'H) and enzymes exhibiting competition for the substrate pools (DFR and FLS) were the main targets for fixed mutations, composing the vast majority of fixed mutations. While mutations in the branching genes and the DFR precursor reaction leading to delphinidin tended to increase activity, the activity of competing reactions was dampened in almost all simulations. These observations intuitively make sense, because these enzymes can have different Michaelis constants and catalytic rates for different substrates [79–81]. Thus, they are poised to exert differential control over pathway flux down the branches, whereas changes in the upstream enzymes will tend to increase or decrease flux globally across all downstream reactions, limiting their ability to play a role in differential tuning of activity.

Simulated evolution informs expectations for biological evolution

Phenotype shifts in most of our simulations were achieved via a small number of mutations in the four key genes, with the median trajectory length being nine steps. This observation has broad implications for evolution of the anthocyanin pathway in biological systems, suggesting that transitioning between pigment phenotypes is relatively easy given mutations at the right loci. In fact, this prediction is consistent with the real systems that have been studied, wherein these same key genes are repeatedly observed as causative mutations [32, 33, 35, 48]. For example, in the genus Iochroma there was a transition from a blue-fowered ancestral state to a red-flowered derived state [36]. This change involved the down-regulation of F3'h, deletion of F3'5'h, and optimization of DFR for the red precursor dihydrokampferol (DHK). These are three of the key mutational targets seen in our simulations. The activity of FLS was not measured, but it is conceivable that it's activity was also tuned during this transition. Interestingly, the causative mutations in this instance pointed in the opposite direction from the changes we see here that shift pathway activity *toward* the blue branch. This type of symmetry is precisely what we would predict from our results for a blue-to-red (rather than red-to-blue) transition. It is also consistent with extensive pleiotropic tradeoffs in pathway activity that we detected in the simulated trajectories (see results). Competition between enzymes for substrates (i.e. DFR and FLS for the DHK, DHQ, and DHM precursors) results in trade-offs that give rise to a tug-of-war between branches. In contrast, the competition between substrates for a single enzyme (i.e. DHK, DHQ, and DHM for DFR) result in trade-offs within a branch. Thus, simultaneous increase in binding affinity or catalytic activity for all substrates is not advantageous when selecting on the steady state concentration of the product from a single branch, and the directionality of advantageous mutations is largely predictable when the phenotype under selection is known.

More experimental studies need to be conducted on anthocyanin pathway evolution to determine if the results from our simulations are generalizable to real systems. However, the apparent agreement of our results with the available empirical data supports the idea that there is a reasonable degree of predictability in evolution of the pathway, at least at the level of the genes that are likely to be involved and how we might expect them to change. These results thus have implications both for the evolution of the anthocyanin pathway in real biological systems and for engineering of the pathway for a desired activity. In fact, anthocyanin engineering applications are an active area of research, with induced production of blue flower coloration being a target [76, 77]. By using a model like the one presented here, parameterized based on measurements from a specific system of interest, a desired optimum phenotype can be defined and putative target mutations can be identified by evolving the model toward that phenotype. These mutations could then be directly introduced into a living system and functionally characterized, potentially avoiding arduous and expensive genetic screens.

The applicability of our results also extends beyond the scope of the anthocyanin pathway, having broader ramifications for understanding the evolution of branched pathways in general. For example, the dominant contributions of branch-point genes to shifts in pigment phenotype is largely due to the ability of these components to exert differential control over the flux down the connected branches. Thus, we predict that this property of branching enzymes should be largely unvisersal across systems. In fact, at least one other computational evolution study found a similar result in a much simpler pathway structure containing a single branch [26]. This property also differs from that expectations for linear pathways [25, 68], wherein the upstream enzymes tend to be the dominant components that control pathway dynamics during evolution. Furthermore, the general outline of the approach taken in this study consists of a few simple steps: 1) write realistic rate laws describing the pathway, 2) choose a simulation API such as Tellurium [63] that is capable of numerically integrating the coupled differential equations to simulate pathway dynamics, and 3) use a evolutionary framework similar to that we have implemented in *enzo* to evolve the pathway model between states. Despite it's simplicity, this approach is a powerful tool for developing expectations of evolutionary behavior and it should be widely applicable to any enzymatic pathway that can be modelled mathematically.

Limitations of the work and future directions

Our mathematical model of the anthocyanin pathway has proven to be very useful for informing expectations in the evolution of pathway activity and for helping to understand the interplay between pathway components during evolution. However, the model is by necessity a simplification of the real system and there are various potential complexities that could be added to future versions. Here we discuss several specific limitations imposed by our simplification and how future work might improve upon the current iteration of our model.

A potential limitation of our model is the use of irreversible Michaelis-Menten kinetics. It is possible that in some circumstances the reversibility of reactions could play an important role in constraining or facilitating evolution of pathway activity. For example, a reversible model might tend to accumulate more intermediate compounds, altering the time-course and steady state behaviors of the system [26, 53]. Such an effect could potentially result in alteration of evolutionary trajectories, requring more or different mutations to reach the optimum. There is evidence that some reactions in the anthocyanin pathway are in fact measureably reversible [55, 56]. Thus, future computational studies of the pathway should include a comparison of the effects of reversible and irreversible rate laws on evolutionary trajectories and outcomes. Reversibility could either be experimentally determined via kinetic measurements with products [82], esimated from thermodynamics [83, 84], or encoded in a naive way similar to our starting point model.

As noted in the methods, we allow all enzyme kinetic parameters to evolve independently in our model, because we were unable to detect a pattern in the relationships between these parameters. Likewise, although there is evidence of co-expression relationships between anthocyain enzyme genes [32, 33, 48, 75, 85], there is insufficient emperical data to build a quantitave model of the co-expression landscape for the anthocyanin pathway. Nonetheless, these relationships may add additional constraints to pathwat evolution and possibly even facilitate certain transitions more [86]. For example, coupling of catalytic constants for multiple substrates might render some hypothetical mutations impossible. Likewise, the ability to simultaneously increase or decrease the expression of multiple enzymes with a single mutation may either help or hinder movement toward another phenotype. Future work should use empirical data, such as enzyme assays and expression measurements, to construct quantitative models of mutational correlations that could be superimposed on the kinetic model to constrain mutational sampling during evolution.

Inclusion of these added biological complexities might help to explain certain properties of real systems, such as the lack of naturally-occurring states that produce both delphindin and pelargonidin [50]. In contrast, states producing delphinidin/cyanidin and pelargonidin/cyanidin combinations are relatively common [50]. Our model was able to capture inherent trade-offs between the two pigments and the relative decoupling of both from cyanidin. However, the extreme discretization of the pathway state space in nature implies that there may be additional constraints that are not accounted for in our simple model. Biochemical limitations on enzyme evolution, co-expression relationships, or simply selection against certain states might contribute to reducing accessible floral pigmentation phenotypes in real systems.

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*, written to perform evolutionary simulations of the pathway model, is available on github [https://github.com/lcwheeler/enzo]. The scripts used to run the entire simulation procedure ("pathway_sims_final.py" and "run_sims.sh") and a Jupyter notebook containing the subsequent analyses ("complete-analysis.ipynb") are available in a separate github repository [https://github.com/lcwheeler/antho-comp-evo-materials]. Files containing the published results in compressed format and an SMBL formatted version of the starting state model ("supplemental-file-1.sbml") are also available in the same repository. The supplemental text contains the rate laws used to construct the model and several supplemental figures referenced in the main 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

[1] Pichersky E, Gang DR. Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends in Plant Science. 2000 Oct;5(10):439–445.

[2] Schuster S, Fell DA, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnology. 2000 Mar;18(3):326–332.

[3] Strohman R. Maneuvering in the Complex Path from Genotype to Phenotype. Science. 2002 Apr;296(5568):701–703.

[4] Nam H, Conrad T, Lewis NE. Metabolic pathway evolution and cellular objectives. Current opinion in biotechnology. 2011 Aug;22(4):595–600.

[5] Morrison ES, Badyaev AV. Structuring evolution: biochemical networks and metabolic diversification in birds. BMC Evolutionary Biology. 2016 Aug;16(1):168.

[6] Keller NP, Turner G, Bennett JW. Fungal secondary metabolism — from biochemistry to genomics. Nature Reviews Microbiology. 2005 Dec;3(12):937–947.

[7] Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt PS, et al. Adaptive Evolution of Metabolic Pathways in Drosophila. Molecular Biology and Evolution. 2007 Jun;24(6):1347–1354.

[8] Colón AM, Sengupta N, Rhodes D, Dudareva N, Morgan J. A kinetic model describes metabolic response to perturbations and distribution of flux control in the benzenoid network of Petunia hybrida. The Plant Journal. 2010;62(1):64–76.

[9] Chen HC, Song J, Wang JP, Lin YC, Ducoste J, Shuford CM, et al. Systems Biology of Lignin Biosynthesis in Populus trichocarpa: Heteromeric 4-Coumaric Acid:Coenzyme A Ligase Protein Complex Formation, Regulation, and Numerical Modeling. The Plant Cell. 2014 Mar;26(3):876–893.

[10] Morrison ES, Badyaev AV. Beyond topology: coevolution of structure and flux in metabolic networks. Journal of Evolutionary Biology. 2017;30(10):1796–1809.

[11] Dykhuizen DE, Dean AM, Hartl DL. Metabolic Flux and Fitness. Genetics. 1987 Jan;115(1):25–31.

[12] True JR, Edwards KA, Yamamoto D, Carroll SB. Drosophila wing melanin patterns form by vein-dependent elaboration of enzymatic prepatterns. Current biology: CB. 1999 Dec;9(23):1382–1391.

[13] Wittkopp PJ, Carroll SB, Kopp A. Evolution in black and white: genetic control of pigment patterns in Drosophila. Trends in Genetics. 2003 Sep;19(9):495–504.

[14] Reed RD, Serfas MS. Butterfly Wing Pattern Evolution Is Associated with Changes in a Notch/Distal-less Temporal Pattern Formation Process. Current Biology. 2004 Jul;14(13):1159–1166.

[15] Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. Nature. 2005 Feb;433(7025):481–487.

[16] Braakman R, Follows MJ, Chisholm SW. Metabolic evolution and the self-organization of ecosystems. Proceedings of the National Academy of Sciences of the United States of America. 2017 Apr;114(15):E3091–E3100.

[17] Zhang L, Martin A, Perry MW, van der Burg KRL, Matsuoka Y, Monteiro A, et al. Genetic Basis of Melanin Pigmentation in Butterfly Wings. Genetics. 2017 Apr;205(4):1537–1550.

[18] Panettieri S, Gjinaj E, John G, Lohman DJ. Different ommochrome pigment mixtures enable sexually dimorphic Batesian mimicry in disjunct populations of the common palmfly butterfly, Elymnias hypermnestra. PLOS ONE. 2018 Sep;13(9):e0202465.

[19] Cornish-Bowden A. Metabolic Control Analysis in Theory and Practice. In: Bittar EE, editor. Advances in Molecular and Cell Biology. vol. 11 of Enzymology in Vivo. Elsevier; 1995. p. 21–64.

[20] Kacser H, Burns JA, Kacser H, Fell DA. The control of flux. Biochemical Society Transactions. 1995 May;23(2):341–366.

[21] The Regulation of Cellular Systems | Reinhart Heinrich | Springer;.

[22] Schuster S, Dandekar T, Fell DA. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends in Biotechnology. 1999 Feb;17(2):53–60.

[23] Stephanopoulos G. Metabolic Fluxes and Metabolic Engineering. Metabolic Engineering. 1999 Jan;1(1):1–11.

[24] Hoefnagel MHN, Starrenburg MJC, Martens DE, Hugenholtz J, Kleerebezem M, Van Swam II, et al. Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. Microbiology. 2002;148(4):1003–1013.

[25] Wright KM, Rausher MD. The Evolution of Control and Distribution of Adaptive Mutations in a Metabolic Pathway. Genetics. 2010 Feb;184(2):483–502.

[26] Rausher MD. THE EVOLUTION OF GENES IN BRANCHED METABOLIC PATHWAYS: EVOLUTION IN BRANCHED PATHWAYS. Evolution. 2013 Jan;67(1):34–48.

[27] Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. Nature. 2005 Jul;436(7050):588–592.

[28] Guillaume F, Otto SP. Gene Functional Trade-Offs and the Evolution of Pleiotropy. Genetics. 2012 Dec;192(4):1389–1409.

[29] Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, et al. Predicting C4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. Cell. 2013 Jun;153(7):1579–1588.

[30] Morrison ES, Badyaev AV. Structure versus time in the evolutionary diversification of avian carotenoid metabolic networks. Journal of Evolutionary Biology. 2018 May;31(5):764–772.

[31] Wisecaver JH, Slot JC, Rokas A. The Evolution of Fungal Metabolic Pathways. PLOS Genetics. 2014 Dec;10(12):e1004816.

[32] Streisfeld MA, Rausher MD. Altered trans-Regulatory Control of Gene Expression in Multiple Anthocyanin Genes Contributes to Adaptive Flower Color Evolution in Mimulus aurantiacus. Molecular Biology and Evolution. 2009 Feb;26(2):433–444.

[33] Streisfeld MA, Rausher MD. Genetic changes contributing to the parallel evolution of red floral pigmentation among Ipomoea species. New Phytologist. 2009 Aug;183(3):751–763.

[34] Smith SD, Rausher MD. Gene Loss and Parallel Evolution Contribute to Species Difference in Flower Color. Molecular Biology and Evolution. 2011 Oct;28(10):2799–2810.

[35] Wessinger CA, Rausher MD. Lessons from flower colour evolution on targets of selection. Journal of Experimental Botany. 2012 Oct;63(16):5741–5749.

[36] Smith SD, Wang S, Rausher MD. Functional Evolution of an Anthocyanin Pathway Enzyme during a Flower Color Transition. Molecular Biology and Evolution. 2013 Mar;30(3):602–612.

[37] Wessinger CA, Rausher MD. Predictability and Irreversibility of Genetic Changes Associated with Flower Color Evolution in Penstemon Barbatus. Evolution. 2014;68(4):1058–1070.

[38] Treves DS, Manning S, Adams J. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of Escherichia coli. Molecular Biology and Evolution. 1998 Jul;15(7):789–797.

[39] Cooper TF, Rozen DE, Lenski RE. Parallel changes in gene expression after 20,000 generations of evolution in Escherichia coli. Proceedings of the National Academy of Sciences. 2003 Feb;100(3):1072–1077.

[40] Cresko WA, Amores A, Wilson C, Murphy J, Currey M, Phillips P, et al. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proceedings of the National Academy of Sciences. 2004 Apr;101(16):6050–6055.

[41] Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. Proceedings of the National Academy of Sciences. 2006 Jun;103(24):9107–9112.

[42] Martin A, Orgogozo V. The Loci of Repeated Evolution: A Catalog of Genetic Hotspots of Phenotypic Variation. Evolution. 2013 May;67(5):1235–1250.

[43] Esfeld K, Berardi AE, Moser M, Bossolini E, Freitas L, Kuhlemeier C. Pseudogenization and Resurrection of a Speciation Gene. Current Biology. 2018 Dec;28(23):3776–3786.e7.

[44] Arbeithuber B, Betancourt AJ, Ebner T, Tiemann-Boege I. Crossovers are associated with mutation and biased gene conversion at recombination hotspots. Proceedings of the National Academy of Sciences. 2015 Feb;112(7):2109–2114.

[45] Kuang MC, Kominek J, Alexander WG, Cheng JF, Wrobel RL, Hittinger CT. Repeated Cis-Regulatory Tuning of a Metabolic Bottleneck Gene during Evolution. Molecular Biology and Evolution. 2018 Aug;35(8):1968–1981.

[46] Winkel-Shirley B. Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiology. 2001 Jun;126(2):485–493.

[47] Streisfeld MA, Liu D, Rausher MD. Predictable patterns of constraint among anthocyanin-regulating transcription factors in Ipomoea. New Phytologist. 2011 Jul;191(1):264–274.

[48] Streisfeld MA, Rausher MD. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. Evolution; International Journal of Organic Evolution. 2011 Mar;65(3):629–642.

[49] Ng J, Smith SD. How to make a red flower: the combinatorial effect of pigments. AoB PLANTS. 2016 Jan;8.

[50] Ng J, Smith SD. Why are red flowers so rare? Testing the macroevolutionary causes of tippiness. Journal of Evolutionary Biology. 2018;31(12):1863–1875.

[51] Ng J, Freitas LB, Smith SD. Stepwise evolution of floral pigmentation predicted by biochemical pathway structure. Evolution. 2018;72(12):2792–2802.

[52] Chou TC, Talaly P. A simple generalized equation for the analysis of multiple inhibitions of Michaelis-Menten kinetic systems. Journal of Biological Chemistry. 1977 Sep;252(18):6438–6442.

[53] Schäuble S, Stavrum AK, Puntervoll P, Schuster S, Heiland I. Effect of substrate competition in kinetic models of metabolic networks. FEBS Letters. 2013 Sep;587(17):2818–2824.

[54] Groenenboom M, Gomez-Roldan V, Stigter H, Astola L, Daelen Rv, Beekwilder J, et al. The Flavonoid Pathway in Tomato Seedlings: Transcript Abundance and the Modeling of Metabolite Dynamics. PLOS ONE. 2013 Jul;8(7):e68960.

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

[56] Halbwirth H. The Creation and Physiological Relevance of Divergent Hydroxylation Patterns in the Flavonoid Pathway. International Journal of Molecular Sciences. 2010 Feb;11(2):595–621.

[57] Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, et al.. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters [review-article]; 2011.

[58] Schomburg I, Chang A, Hofmann O, Ebeling C, Ehrentreich F, Schomburg D. BRENDA: a resource for enzyme data and metabolic information. Trends in Biochemical Sciences. 2002 Jan;27(1):54–56.

[59] Kanehisa M. The KEGG Database. In: ‘In Silico’ Simulation of Biological Processes. John Wiley & Sons, Ltd; 2008. p. 91–103.

[60] Miranda HV, Ferreira AEN, Quintas A, Cordeiro C, Freire AP. Measuring intracellular enzyme concentrations: Assessing the effect of oxidative stress on the amount of glyoxalase I. Biochemistry and Molecular Biology Education. 2008 Mar;36(2):135–138.

[61] Poustka F, Irani NG, Feller A, Lu Y, Pourcel L, Frame K, et al. 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 Physiology. 2007 Dec;145(4):1323–1335.

[62] Petrussa E, Braidot E, Zancani M, Peresson C, Bertolini A, Patui S, et al. Plant Flavonoids—Biosynthesis, Transport and Involvement in Stress Responses. International Journal of Molecular Sciences. 2013 Jul;14(7):14950–14973.

[63] Choi K, Medley JK, Cannistra C, Konig M, Smith L, Stocking K, et al. Tellurium: A Python Based Modeling and Reproducibility Platform for Systems Biology. bioRxiv. 2016 Jun;p. 054601.

[64] Smith LP, Bergmann FT, Chandran D, Sauro HM. Antimony: a modular model definition language. Bioinformatics. 2009 Sep;25(18):2452–2454.

[65] Somogyi ET, Bouteiller JM, Glazier JA, König M, Medley JK, Swat MH, et al. libRoadRunner: a high performance SBML simulation and analysis library. Bioinformatics. 2015 Oct;31(20):3315–3321.

[66] Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 2003 Mar;19(4):524–531.

[67] Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, et al. COPASI—a COmplex PAthway SImulator. Bioinformatics. 2006 Dec;22(24):3067–3074.

[68] Clark AG. Mutation-selection balance and metabolic control theory. Genetics. 1991 Nov;129(3):909–923.

[69] Khersonsky O, Tawfik DS. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. Annual Review of Biochemistry. 2010;79(1):471–505.

[70] Kaltenbach M, Tokuriki N. Dynamics and constraints of enzyme evolution. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 2014 Nov;322(7):468–487.

[71] Rausher MD. The Evolution of Flavonoids and Their Genes. In: Grotewold E, editor. The Science of Flavonoids. New York, NY: Springer New York; 2006. p. 175–211.

[72] Ng J, Smith SD. Widespread flower color convergence in Solanaceae via alternate biochemical pathways. New Phytologist. 2016;209(1):407–417.

[73] Landry LG, Chapple C, Last RL. Arabidopsis Mutants Lacking Phenolic Sunscreens Exhibit Enhanced Ultraviolet-B Injury and Oxidative Damage. Plant Physiology. 1995 Dec;109(4):1159–1166.

[74] Ryan KG, Markham KR, Bloor SJ, Bradley JM, Mitchell KA, Jordan BR. UVB Radiation Induced Increase in Quercetin: Kaempferol Ratio in Wild-Type and Transgenic Lines of Petunia. Photochemistry and Photobiology. 1998;68(3):323–330.

[75] Fornalé S, Lopez E, Salazar-Henao JE, Fernández-Nohales P, Rigau J, Caparros-Ruiz D. AtMYB7, a New Player in the Regulation of UV-Sunscreens in Arabidopsis thaliana. Plant and Cell Physiology. 2014 Mar;55(3):507–516.

[76] Shimada Y, Ohbayashi M, Nakano-Shimada R, Okinaka Y, Kiyokawa S, Kikuchi Y. Genetic engineering of the anthocyanin biosynthetic pathway with flavonoid-3′,5′-hydroxylase: specific switching of the pathway in petunia. Plant Cell Reports. 2001 Jul;20(5):456–462.

[77] Sasaki N, Nakayama T. Achievements and Perspectives in Biochemistry Concerning Anthocyanin Modification for Blue Flower Coloration. Plant and Cell Physiology. 2015 Jan;56(1):28–40.

[78] Gates DJ, Olson BJSC, Clemente TE, Smith SD. A novel R3 MYB transcriptional repressor associated with the loss of floral pigmentation in Iochroma. New Phytologist. 2018 Feb;217(3):1346–1356.

[79] Fischer TC, Halbwirth H, Meisel B, Stich K, 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. Archives of Biochemistry and Biophysics. 2003 Apr;412(2):223–230.

[80] Miyagawa N, Miyahara T, Okamoto M, Hirose Y, Sakaguchi K, Hatano S, et al. Dihydroflavonol 4-reductase activity is associated with the intensity of flower colors in delphinium. Plant Biotechnology. 2015;32(3):249–255.

[81] Katsu K, Suzuki R, Tsuchiya W, Inagaki N, Yamazaki T, Hisano T, et al. A new buckwheat dihydroflavonol 4-reductase (DFR), with a unique substrate binding structure, has altered substrate specificity. BMC Plant Biology. 2017 Dec;17(1):239.

[82] Bisswanger H. Enzyme assays. Perspectives in Science. 2014 May;1(1):41–55.

[83] Flamholz A, Noor E, Bar-Even A, Milo R. eQuilibrator–the biochemical thermodynamics calculator. Nucleic Acids Research. 2012 Jan;40(D1):D770–D775.

[84] Noor E, Flamholz A, Liebermeister W, Bar-Even A, Milo R. A note on the kinetics of enzyme action: A decomposition that highlights thermodynamic effects. FEBS Letters. 2013 Sep;587(17):2772–2777.

[85] Zhou H, Peng Q, Zhao J, Owiti A, Ren F, Liao L, et al. Multiple R2R3-MYB Transcription Factors Involved in the Regulation of Anthocyanin Accumulation in Peach Flower. Frontiers in Plant Science. 2016;7.

[86] Smith SD. Pleiotropy and the evolution of floral integration. New Phytologist. 2016;209(1):80–85.

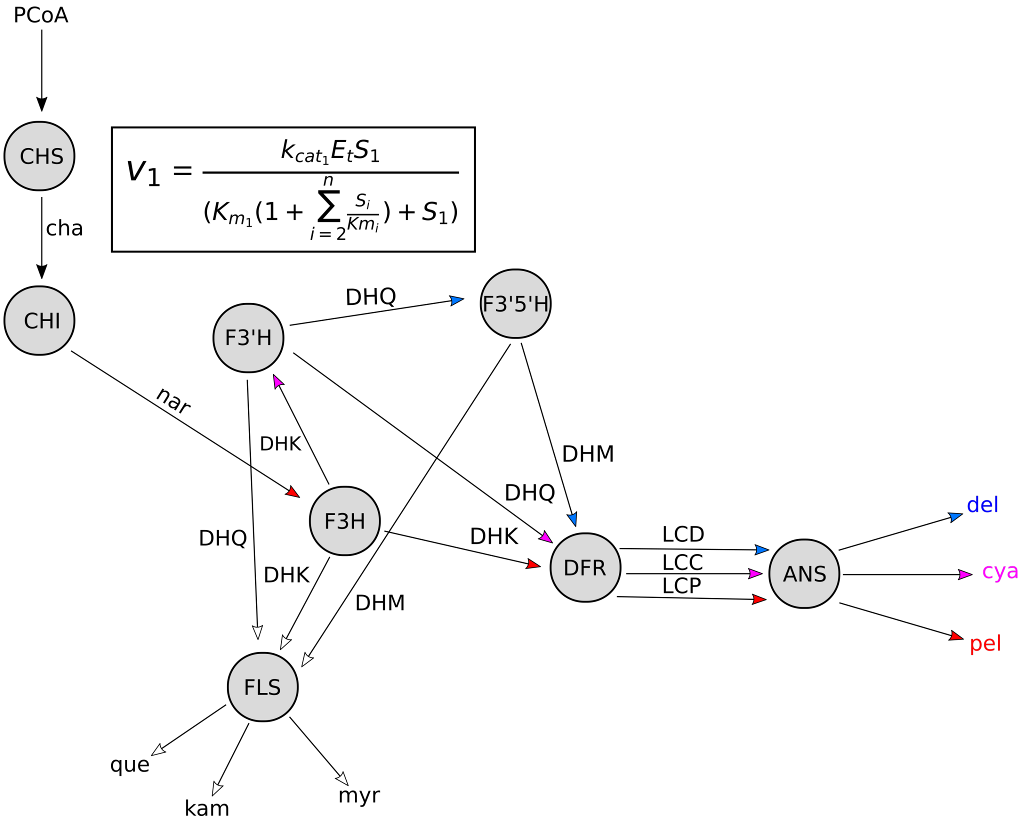

PCoA
CHS
cha
CHI
$v_1 = \frac{k_{cat_1} E_t S_1}{(K_{m_1}(1 + \sum_{i=2}^{n} \frac{S_i}{Km_i}) + S_1)}$
F3'H
DHQ
F3'5'H
nar
DHK
DHM
DHQ
F3H
DHQ
DHK
DFR
LCD
LCC
LCP
ANS
del
cya
pel
DHK
DHM
FLS
que
kam
myr

Fig 1. Diagram of the simplified anthocyanin pathway model. The simplified model is depicted as an enzyme-centric pathway diagram. It includes the key anthocyanin and flavonol branches. 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 [53]. 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. 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). Blue arrows lead to delphinidin, red arrows lead to pelargonidin, purple arrows lead to cyanidin.

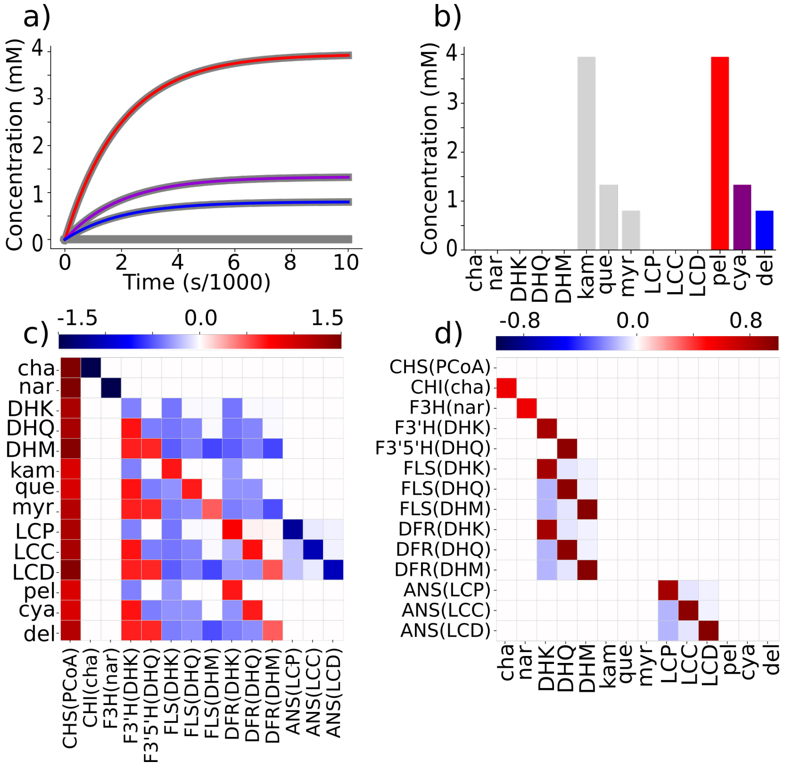


Fig 2. Kinetic behavior and metabolic control analysis of the naive starting state model. a) time-course simulation of the unbiased starting state model showing anthocyanidin concentrations over time: pelargondin (red), cyanidin (purple), and delphinidin (blue). The other compounds are shown in gray. The symmetrical flavonol equivalents of each anthocyanidin (kamperfol, quercetin, and myricetin) can be seen in gray behind the red, purple, and blue lines, respectively. b) Steady state concentrations of each floating species in the naive model. c) Heatmap of concentration control coefficients for all pathway reactions except for "sinks" at boundary conditions. d) Heatmap of elasticity matrix for all pathway reactions except for diffusion "sinks" at boundaries. 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.

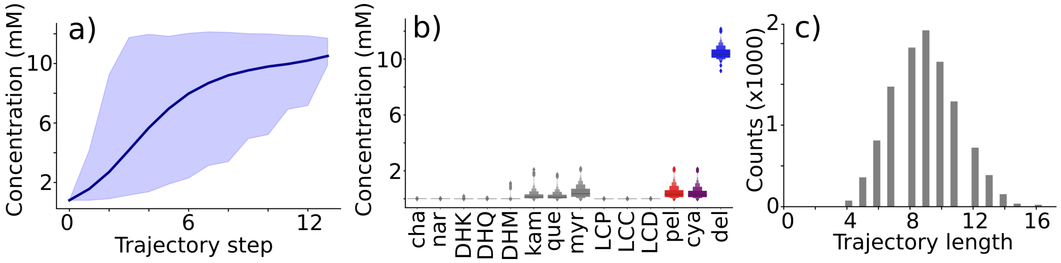


Fig 3. Evolutionary trajectories and optimum state steady state concentration distributions. a) Trajectories through delphinidin space. Dark blue line shows mean trajectory, averaged across all 9,985 simulations. Shaded blue area shows envelope containing 95% of trajectories. Top limit is maximum value 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).

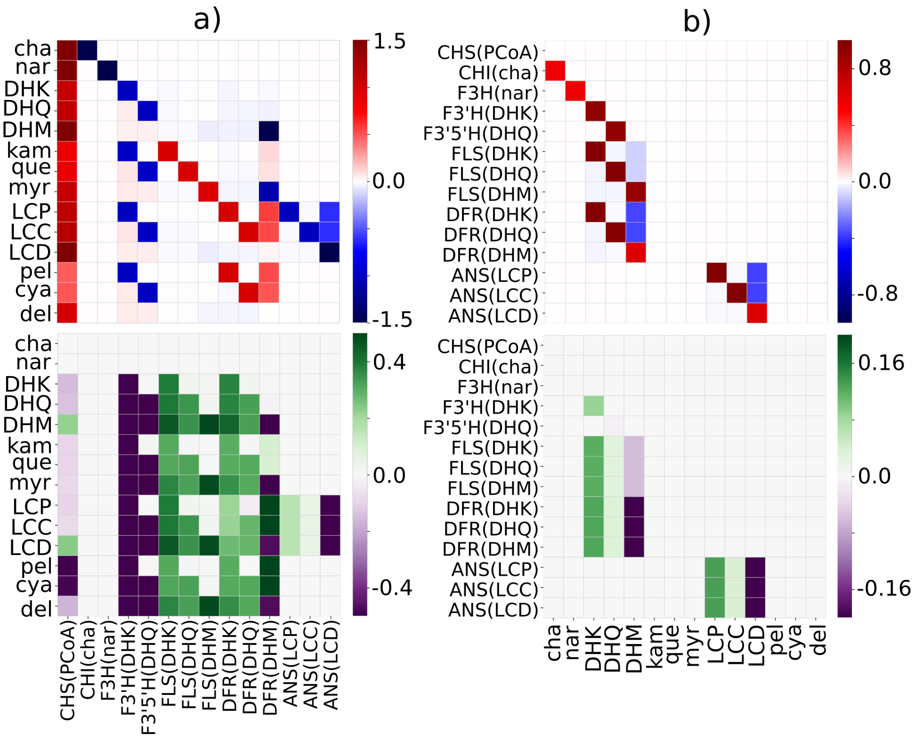
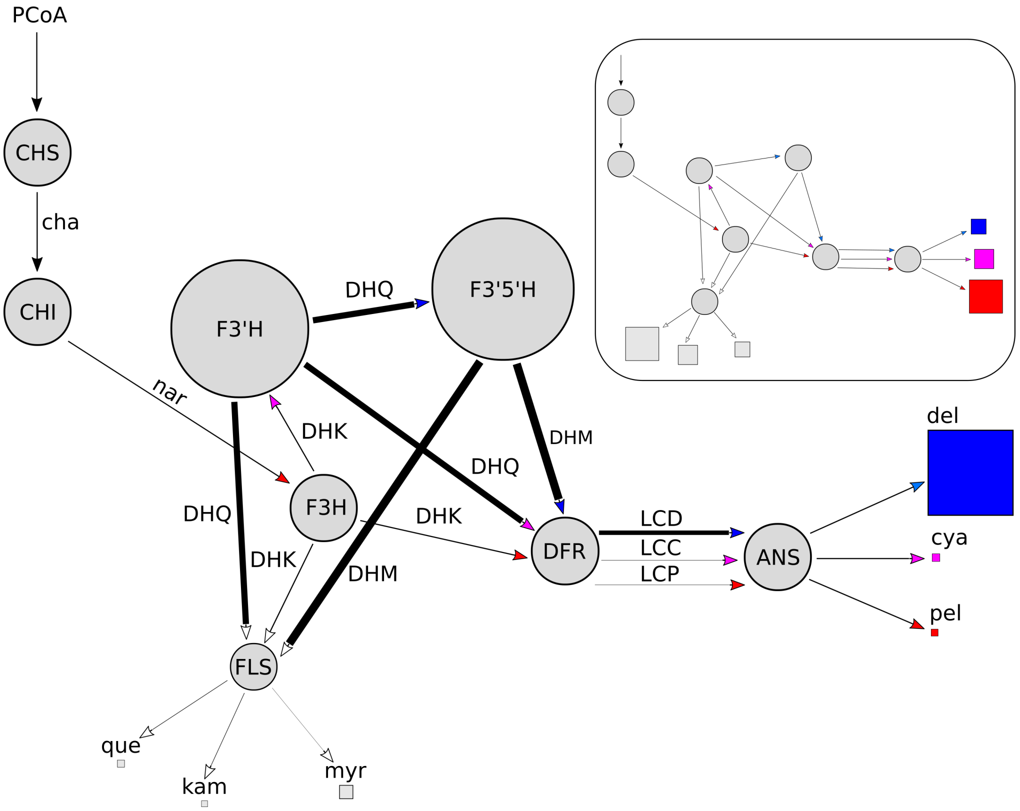


Fig 4. Pathway control shifts to acheive a bias toward the delphinidin branch. a) Top panel shows heatmap of median optimal concentration control coefficient matrix, taken over all 9,985 simulations. Bottom panel is the difference between starting state and median optimal end state concentration control coefficient matrices. b) Top panel shows a heatmap of median optimal elasticity coefficient matrix, taken over all 9,985 simulations. Bottom panel is the difference between starting state and median optimal end state elasticity coefficient matrices. Heatmaps axes are labeled as in fig. 2. 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.



PCoA
CHS
cha
CHI
nar
F3'H
DHQ
F3'5'H
DHK
DHM
DHQ
F3H
DHQ
DHK
DHK
DHM
DFR
LCD
LCC
LCP
ANS
del
cya
pel
FLS
que
kam
myr

Fig 5. Parameters of branching enzymes are the dominant components of pathway evolution. 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 concentration of the indicated floating species produced by a model initialized with mean parameter values. Inset shows the starting state model, in which all parameters of a given type are equal (see methods).

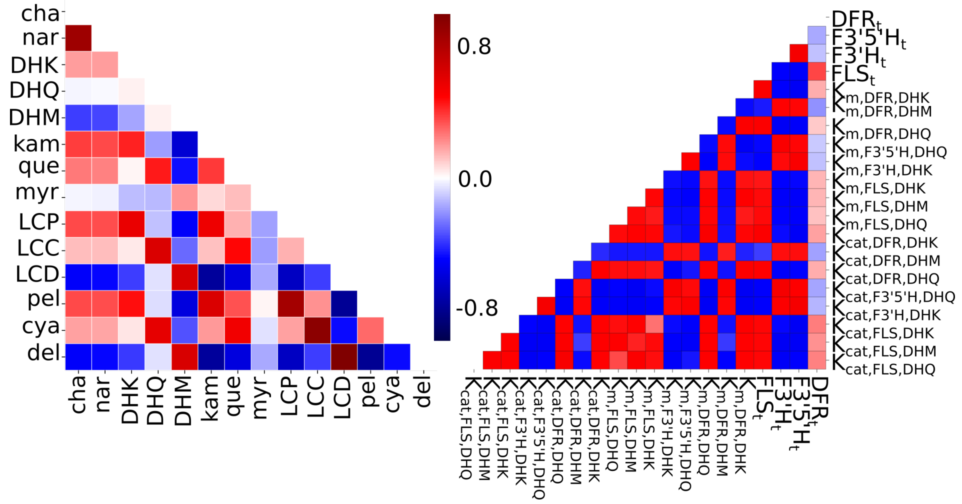


Fig 6. Pleiotropic trade-offs result from branching and competition. a) Heatmap of mean Spearman correlations between floating species concentrations across all 9,985 trajectories. b) Heatmap of mean Spearman correlations between parameter changes across all 9,985 trajectories for those parameters accounting for more than 5% of fixation events. Correlations are shown between enzyme concentrations (E_t), catalytic constants (K_{cat}), and the *inverse* of Michaelis constants ($1/K_M$ in units of M^{-1}) to make interpretation more intuitive, because higher K_M (in mM units) corresponds to weaker binding.