A general evolutionary model for the emergence of novel characters from serial homologs

Daohan Jiang[a], Matt Pennell[b] & Lauren Sallan[a,1]

[a]*Macroevolution Unit, Okinawa Institute of Science and Technology Graduate University, Japan*
[b]*Department of Computational Biology, Cornell University, USA*
[1]Corresponding author: lauren.sallan@oist.jp

Keywords: novelty, development, gene expression, serial homology, cell type

Abstract

The origin of morphological characters with novel forms and functions is among the most fascinating phenomena in evolution, and understanding its underlying principles has been a fundamental goal of evolutionary biology. An essential means by which novelties evolve is the divergence of repeated body parts, known as serial homologs, into various forms, as in the case of tetrapod limbs and insect wings. However, the mechanisms underlying such processes are poorly understood, with systematic investigations hindered by the lack of a generalized model that links selective, genetic, and developmental mechanisms and the expected patterns of evolutionary dynamics. To fill this gap, we propose a generalizable model for the evolution of serial homologs that is founded on recent advances in developmental biology, where the development of each serial homolog is controlled via a hierarchical gene regulatory network consisting of genes that specify character identity and those that encode the specific phenotypic state. Under this model, we investigated two types of evolutionary novelties: divergence in character states between body parts with conserved identities and switching of character identity by turning on and off master regulators. Using population genetic simulations, we demonstrate how selection and developmental constraints interact to shape dynamics of phenotypic evolution and conditions under which each type of novelty is likely to evolve. Together, our results provide general insights into how novelties could evolve from serial homologs, and offer a modeling framework where the developmental evolution of a broad range of phenotypic characters can be studied.

Introduction

To understand mechanisms underlying the origin of morphological characters with novel forms and functions has been a fundamental goal of evolutionary biology. Central to the problem of evolutionary novelties is how developmental processes can constrain and bias the trajectory of of evolution. It has been proposed that developmental constraints, in general, can have a dispositional effect on phenotypic evolution by channeling the effect of mutations on the phenotype and thereby biasing the availability of genetic variations [1–9]. Advancements in developmental biology over the past few decades have shed light on developmental mechanisms underlying phenotypic divergence between species and between body parts of the same organism. For example, phenotypic similarity between homologous body parts of different species and that between repeated body parts of the same organism (serial homologs) results from similar gene expression programs during their development—i.e., which genes are expressed, their expression levels, and timing of their expression; by the same principle, difference in the gene expression programs is responsible for phenotypic variation among the otherwise identical body parts [10–15].

There is abundant evidence that phenotypic changes result from the interplay between genes at different levels of an organizational hierarchy during development. This suggests that mutations in some genes have larger effects on character form than others depending on their position within this hierarchy. Observed morphological differences between species or between serial homologs have, in many cases, been hypothesized or confirmed as attributable to differential local expression of upstream "master regulator" genes that activate or suppress many other genes and/or pathways. Expression of such regulatory genes can specify the identity of a body part, but not necessarily specific character states (the observable realized phenotype, such as size, shape, or color). That is, the expression profile of these genes is shared by of homologous body parts of different species, regardless of how different their forms are [11–13]. A well documented case of such regulator is the *Ultrabithorax* (*Ubx*) gene: *Ubx* expression specifies the identity of hindwings in insects, and the loss of *Ubx* expression causes forewings to grow at the position of hindwings, regardless of specific size or morphology of the wings [16–24]. For instance, in fruit flies, the perturbation causes the halteres to be replaced by wing blades, whereas in beetles, it is the hindwings being replaced by elytra. Another example of well-studied character identity determination mechanisms is that of tetrapod limbs, in which case two *T-box* genes, *Tbx4* and *Tbx5*, specify forelimb and hindlimb identities respectively but not specific limb morphologies [25–31]. In addition to morphological characters, identities of cell types are also found to be regulated in similar ways (e.g., [32–39]). This suggests that modeling the interactions between master regulators and their downstream effectors can be used to infer the resultant phenotypes and phenotypic divergence for many traits.

Despite the progress in understanding the underlying developmental mechanisms linking genotypes and phenotypes, there has been little systematic study of their implications for morphological evolution and particularly the origin of novel characters. One major obstacle is that there is no existing quantitative, mechanistic model linking genetic changes and developmental processes. In addition, developmental models that make quantitative predictions about developmental evolution are rare to our knowledge. The few examples of such models include those for the growth mollusk shells [40, 41], which lacked a genetic basis, and for mammalian molars [42], which was based on a well characterized pathway. The latter allowed developmental constraints to be incorporated in phylogenetic comparative analyses to study the evolutionary dynamics [43], thus suggesting the utility of such models for evolutionary study of phenotypic change.

In the absence of quantitative models and sufficient information on the developmental basis for most traits, empirical observations have routinely been used to infer processes underlying the emergence of new forms. These inferences then form the basis of qualitative models and hypotheses, which are often used to make predictions about the plausibility of specific and potential character states. Qualitative models for

novelty are then assumed to be supported if predicted intermediate forms exist in the fossil record or matches normal ontogenetic stages (recapitulation) or mutant phenotypes arising from mutations. They are typically rejected if forms predicted to be implausible are observed in any such data. For instance, the qualitative model for feather development and evolution proposed by Prum [44] makes predictions about possible forms of early feathers, which are supported by findings in various early-diverging coelurosaur groups [45]. Another case of a qualitative model (un)supported by fossil evidence is the teleost caudal fin, where a growth series of the non-teleost ray-finned fish *Aetheretmon valentiacum* overturned a classic model of recapitulation of evolutionary stages during development while supporting an alternative model of derived states for non-homologous structures [46]. The notochord-bearing "tail" in early embryos and adult fossils was proposed to be serially homologous with tetrapod limbs in terms of shared dependence on homeobox (*Hox*) gene-regulated pathways, based on existing qualitative models of limb and other vertebrate outgrowth [46]. This new hypothesis is in turn supported by subsequent developmental observations; genetic manipulation of *Hox13* paralogs in zebra fish changes the growth of caudal structures in expected ways [47], and outgrowth of structures such as the zebrafish cloaca, fin rays and tetrapod digits originated from co-option of *Hox13* and are serially homologous [48]. There are many other qualitative models, hypotheses, and proposed mechanisms which remain supported only by empirical observations. For most, there has been no way objectively test the validity and logic of their underlying assumptions, simulate outcomes, identify the minimum requirements, or make predictions of quantitative data should be structured like if the model were accurate.

To address these issues, we need a generalizable quantitative model under which one can investigate and analyze how features of developmental processes shape phenotypic evolution. Here, we present a mathematical model that incorporates key general features of gene regulatory networks (GRNs) underlying the development of morphological characters and cell types and explore its implications for phenotypic evolution and the emergence of novelties. Under this model, the character state of a body part is controlled via a hierarchical gene regulatory network consisting of two classes of genes: the regulatory genes ("regulators" for short) and the effector genes ("effectors" for short). The regulators encode *trans*-regulatory factors (e.g., transcription factors) that interact with *cis*-elements (e.g., promoters and enhancers) of the effect genes and regulate the effectors' expression. A regulator expression program (combination of expression levels of regulators) that specifies a character identity is referred to as a character identity network (ChIN) [12, 13, 49, 50], which is also known as a core regulatory complex in the context of cell type identity determination [51].

Using our quantitative modeling framework, we investigated two types of evolutionary novelties. The first type of novelty is the divergence in character states between serially homologous body parts with conserved identities. These characters have distinct ChINs that are conserved during evolution, and divergence in character states is due to substitutions in *cis*-elements of the effectors. Prime examples of such characters include the aforementioned cases of tetrapod limbs and insect wings, among others. The second type of evolutionary change we examined is that a body part switches to an alternative identity by expressing a pre-existing alternative ChIN. As each regulator can regulate the expression of many effectors, alternation of their expression can lead to substantial transcriptomic change, potentially leading to drastic phenotypic changes. To explore principles underlying these two types of evolutionary novelties, we performed individual-based population genetic simulations using SLiM [52]. By examining evolutionary dynamics in a variety of evolutionary scenarios, we explored conditions under which each type of novelty is likely to evolve, demonstrating how patterns of evolution are connected to the underlying mechanisms.

Results

An generalizable model for the developmental control of character identity and state

In our model, we assume that there are n regulatory genes that specify the character identity and m effector genes that encode the character state. Expression levels of the effector genes are then given by $\mathbf{Y} = \exp(\mathbf{AX})$, where $\mathbf{Y}$ is a vector of length m and $\mathbf{A}$ is an $m \times n$ matrix summarizing per-unit-expression regulatory effects of the regulators on the expression of the effectors (an exponential function is used here to capture regulatory effects; see also Methods). Element $A_{i,j}$ of $\mathbf{A}$ represents the effect of the j-th regulator on the expression of the i-th effector, and its values captures both strength and direction of interactions between the regulator and *cis*-regulatory elements of the effector: a positive value represents an activation effect, whereas a negative value represents repression. The $\mathbf{A}$-matrix can further be decomposed as a product of two matrices $\mathbf{A} = \alpha\beta$, where α captures effects of *cis*-elements on the effector genes and β captures binding affinity between gene products of the regulators and *cis*-elements of the effectors. At last, the realized character state, which is d-dimensional and represented by a vector of length d, is given by $\mathbf{z} = \mathbf{B} \ln \mathbf{Y}$, where $\mathbf{B}$ is a $d \times m$ matrix characterizing the per-unit-expression-level effect of the effectors on the phenotype (see also Methods). Homologous body parts of different organisms are expected to have the same $\mathbf{X}$, but may have different states due to differences in their $\mathbf{A}$-matrices. By contrast, serially homologous body parts with distinct identities within the same organism will have difference(s) in $\mathbf{X}$, and potentially different character states depending on the structure of $\mathbf{A}$. The origin of a novel character requires the evolution of a unique $\mathbf{X}$ that distinguish it from pre-existing body parts.

Character state divergence between body parts with conserved identities

We considered two serially homologous body parts with distinct identities, each specified by a respective regulator. Character states of the two body parts (denoted by z_1 and z_2, respectively) are produced by the same set of effectors. Each effector is regulated either by a single *cis*-element that interacts with both regulators, or two *cis*-elements that respectively interact with the two regulators. We considered scenarios where different proportions of effectors fall into these two categories. Expression levels of the regulators in the two body parts are conserved over time, and evolutionary changes in character states are mediated by mutations in *cis*-elements that affect the strength of *cis-trans* interactions (β). Mutations in a *cis*-element that is a shared target of two regulators will tend to have concordant effects on its binding affinity with both regulators, thereby causing concordant effects on effector gene expression in two body parts. By contrast, mutations in *cis*-elements that are not shared targets of the two regulators can result in differential expression of the effectors between two body parts, thereby character state divergence between them (Fig. 1A).

We considered three scenarios of directional selection: in the first scenario, z_1 is under directional selection for a greater value and z_2 is under stabilizing selection; in the second scenario, which is referred to as a scenario of divergent selection, the optimum of z_1 is greater than the ancestral state and whereas that of z_2 is smaller than the ancestral state; in the last scenario, which is referred to as a scenario of concordant selection, greater values of both z_1 and z_2 are selected for. In the absence of any genetic constraint, phenotypic divergence between serial homologs is expected to increase in the first two scenarios but not the third one. For each regime of selection, we simulated 10 replicate populations for 10000 generations in SLiM [52], and examined the mean level character state divergence between two body parts across populations and the mean evolutionary change in z_1 relative to the ancestral state.

As predicted, in the first two scenarios, character state divergence between two body parts increased

over time in the first two scenarios (Fig. 1B-C) but not the third one (Fig. 1D). We also found that the rate of character state divergence to be generally lower when more effector genes are regulated via shared targets of two regulators (Fig. 1B-C). Evolutionary change of z_1, which had the same optimal value in all scenarios, agrees with these observations (Fig. S1). We also examined scenarios where correlational selection is present, including scenarios of positive correlational selection (optimal scaling line has a positive slope) and negative correlational selection (optimal scaling line has a negative slope). When two body parts had different optimal states and correlational selection is positive , the rate of character state divergence and that of z_1 evolution decreased with the number of targets shared by two regulators (Fig. S2A-B, Fig. S3A-B); in other scenarios with correlational selection, evolutionary dynamics was rather insensitive to target sharing (Fig. S2C-F, Fig. S3C-F). Together, our results demonstrate the impact of developmental constraint on adaptive divergence between character states of serial homologs.

Switching between pre-existing identities

We considered a single body part that can assume one of two potential identities (denoted I_1 and I_2, of which I_1 is ancestral) by expressing two different regulators that respectively activate two distinct sets of effector genes. Effectors regulated by two regulators affect the character state differently, so the two identities correspond to two distinct character states. Which regulator to express in a given body part is determined by the local concentration of a morphogen, denoted c. When c is equal to or above a cutoff, the regulator specifying I_1 is expressed; otherwise, the one specifying I_2 is expressed. Mutations that alter c can turn off one regulator while turning on the other, resulting in change in expression states of their targets and drastic change in character state (Fig. 2A). The ancestral value of c is just near the cutoff to produce I_1, and a mutation that lowers c will switch the identity; that is, there is minimal genetic constraint on identity switching on an ancestral background.

We performed simulations in SLiM under five regimes of selection characterized by different optimal states (z_{opt}), including the character state produced by the ancestral identity I_1 (given ancestral regulatory parameters; the same for the rest), the phenotype produced by the non-ancestral identity I_2, an intermediate state right in between those produced by I_1 and I_2, an intermediate phenotype that is closer to that produced by I_1, and an intermediate phenotype that is closer to that produced by I_2. For each regime of selection, we simulated 20 replicate populations and examined the proportion of populations where I_2 is fixed. As predicted, I_2 was fixed in most, albeit not all populations when z_{opt} is the character state it produces (Fig. 2B, leftmost column). Similar results were observed when the optimum is not the same but relatively close to the character state produced by I_2 (Fig. 2B, second column), indicating an identity shift can happen even if a shift does not lead to the exact optimal phenotype. Identity shift happened relatively infrequently under other regimes of selection as well as under neutrality, as expected (Fig. 2B). Mean expression of the regulator underlying I_2 showed consistent results (Fig. 2C). In the two scenarios where identity switching was frequent, the mean character states across populations ended up close to the respective optima (S4), indicating complementation by evolutionary changes in regulator-effector interactions. Together, our results show that identity shift can facilitate adaptation when the new optimum is in the vicinity of a character state that can potentially be produced by an alternative character identity.

Discussion

In this study, we present a generalizable model of developmental evolution and used it to explore principles of evolutionary innovations. We focused on a simplified, two-level model where two classes of genes, namely

regulators that specify character identities and effector genes that produce character states, and regulatory effect of the regulators on the effectors, were considered. This abstraction, while allowing the mathematical models and simulations to be simpler and more manageable, also captures essential properties of hierarchical GRNs underlying the development of morphological characters formation of cell types. Within-group regulation and feed-back loops, on the other hand, can potentially have their own evolutionary consequences, as structural properties of GRNs do in general [53–55], and are of interest to future studies. Importantly, interactions between regulators, such as mutual suppression that mutually exclusive expression and positive feedback loops that drive co-expression, are involved in specification identities of morphological characters and cell types [31, 36]. Our simulations, while not explicitly modeling interactions between regulators, can be interpreted as having some incorporated implicitly. For instance, interactions driving co-expression of regulators can be readily modeled if each entry of $\mathbf{X}$ is interpreted as a set of co-expressed genes instead of a single gene. One can also model expression levels of two or more regulators as functions of a latent variable (morphogen concentration) to recapitulate specific patterns of co-expression. In addition, it should be noted that, the that the same gene can be involved in the development of different body parts that are not serially homologous and have different roles in the corresponding GRNs (e.g., various *Hox* genes [56] and *Shh* [57]). Such pleiotropy is beyond the scope of this study, but will be essential to modeling the coevolution of distinct types of organs.

In our simulations, structure of GRNs underlying development underwent minimal change. Specifically, *de novo* birth of new *cis*-elements (new column of α and new row of β) was not modeled. Modeling such events requires information on their frequency the distribution of effects of new *cis*-elements. Recent years have also seen progress made in predicting regulatory potential of putative *cis*-elements and detecting selection on regulatory activities [58–65]; findings in this area may shed light on how gene regulatory networks underlying development could potentially be rewired during evolution and help better parameterize models for developmental evolution. Modeling the birth of new *cis*-elements and the resulting change in GRN structure could also help understand the role of second-order selection in evolutionary innovations. It has been hypothesized that second-order selection for mutational robustness and/or evolvability has been responsible for modularity of GRNs controlling development [66–70], and understanding the mutational input for this evolutionary process is key to understanding its dynamics.

We examined models of selection where the relationship between phenotype and fitness is specified while mechanisms underlying the relationship are not. However, such simple models can also represent more specific scenarios of potential interest. For example, negative correlational selection could be mediated by selection for division of labor in response to trade-off between functions performed by the same body part [71]. Our modeling framework can readily be developed to explore implications of specific mechanisms of selection; one can introduce variables that mediate the effect of the phenotype on fitness—e.g., locomotion and foraging performance and model them as custom-define functions of the character state. Models of adaptation examined in this study are simple models concerning a single episode of adaptive evolution with the optimal character state pre-specified. Another class of models that would be of interest is models with moving adaptive landscapes, which may better explain variation in the distribution of phenotypes across divergent clades. In addition, movement of the adaptive landscape can potentially interact with developmental constraints to shape patterns macroevolutionary divergence—e.g., when there is strong developmental (and thereby genetic) constraint, populations track the moving optimum less well and may end up stranded near an ancestral state and display less evolutionary divergence [72]. It remains an open question, however, whether adaptive landscapes' movement is generally predictable, and if so, how they move [73, 74]. For body parts with positional identities, there could also be correlation between their identities and functional constraints they are subject to; that is, how well a body part can perform a given task depends on where it is located (e.g., position along an axis). Such constraints could bias the movement of the adaptive peak and might explain some observed patterns of evolution: for example, bipedal tetrapods typically support their weight

using their hindlimbs rather than forelimbs [75], and arthropod appendages specialized for assisting feeding need to posited near the anterior end [76–78].

Under our modeling framework, we examined two types of evolutionary novelties, the first being character state divergence given conserved identities, mediated by evolutionary changes in interactions between regulators and *cis*-elements. Under certain regimes of selection, the rate of adaptive divergence in character states of two body parts is faster when regulators specifying different identities share a smaller portion of targets (Fig. 1B-D). This is consistent with the notion that decoupling of developmental regulation of different serial homologs enhances evolvability and allows more diverse combinations to evolve [79]. Our model for such characters can potentially help us understand evolutionary mechanisms underlying a broad range of novel character states. A potential case of this type of evolutionary change the evolution of long arms in bats: substitutions in an enhancer increased the expression level of *Prx1*, a gene that does not specify limb identities, in developing forelimb bones, resulting in forelimb elongation [80]. Similarly, limb size reduction, which has taken place repeatedly in various tetrapod lineages [81, 82], could have been mediated by similar mechanisms as well (we note that the complete loss of limbs is more appropriately a loss of character identity [83] rather than an extreme character state, albeit potentially driven by similar selection regimes).

The second type of novelty we examined is the switching of character identity via turning on and off regulators. We found such evolutionary changes are likely to occur when the optimum is closer to the phenotype produced by the derived identity than that produced by the ancestral identity and can be complemented. Scenarios examined in our simulations can potentially explain drastic evolutionary changes between apparently discrete states arising from the same developmental cell lineages in similar positions. A putative example is the transition between feathers and scales states in birds as well as non-avian dinosaurs: mechanistically, such a conversion can be achieved by simply activating or suppressing *Sonic Hedgehog* (*Shh*) signaling (in given body regions) [84–86]. On the other hand, there is indeed substantial diversity in the spatial distribution of feathers among theropods [45, 87], which could have been mediated by genetic changes that altered the expression of master regulators like *Shh*.

When interpreting results of our evolutionary simulations, we can disentangle evolutionary changes of character identity and state, as we know the truth of our simulations. However, in empirical studies, it is not always clear if the observed states resulted from character identity changes, character state changes, or acquired through loss and replacement rather than fully homologous, which can potentially confound further evolutionary analyses. Thus, more developmental studies will be necessary to determine how broad a range of novelties are explained by either of the two types of scenarios examined in this study. A general approach to distinguishing evolutionary changes of character identity and state is to compare the effect of the same genetic or gene expression perturbation across species. Specifically, one can test if the same perturbation produce similar character states in different species, or affect homologous characters of different species while producing dissimilar character states. A prime example of such investigations is the aforementioned case of *Ubx*. In both fruit flies and beetles, hindwings are replaced by forewings upon loss of *Ubx* expression while the resulting character state changes differ in the two groups, indicating the gene is responsible for specifying hindwing identity but not specific character states of halteres or 'normal' wings [12, 16–23]. Hence, evolutionary changes mediated by *Ubx* expression evolution shall be interpreted as identity instead of state changes. More such investigations will be necessary to clarify the nature of other instances of novelties and understand the underlying evolutionary mechanisms using our model.

Our modeling framework can be extended to study the dynamics of evolutionary innovations mediated by a broader range of developmental genetic mechanisms. One type of evolutionary change of interest is change in the number of serially homologous body parts within an organism; specifically, how the duplication of body parts could potentiate the evolution of novel character identities under selection for novel functions or division of labor. Mechanisms mediating developmental and selective constraints on the number of serial

homologs, however, are poorly understood and likely vary among different types of characters, making it difficult to model the evolution of the number of serial homologs. In particular, it is challenging to distinguish constraints resulting from selection on the number *per se* and indirect selective constraint mediated by pleiotropic effects of mutations that alter the number of them (e.g., mutations responsible for Pallister–Hall syndrome cause polydactyly along with other symptoms [88]). Pleiotropy is known to constrain the evolution of traits that are not themselves under selection, resulting in reduced level of within-population phenotypic variation and evolutionary rate [89–92]. For discrete traits showing little variation among species, however, it can be much more challenging to tease apart direct and indirect effects of selection. A type of character that may potentially offer implications for this problem are serial homologs whose number is conserved within some clades but variable in other clades and between clades. For example, the number of cervical vertebrae is highly conserved among mammals (typically seven, with few exceptions) but much more variable in sauropsids, likely due to different levels constraint resulting from pleiotropy [93–97]. It would be of interest for future studies to compare developmental processes as well as selective constraints in different clades to provide insights into mechanisms underlying constraints on the number of serial homologs.

Together, in this study, we present a generalizable model for the evolution of novel morphological characters and cell types, incorporating their underlying developmental mechanisms. Using this model, we investigated three types of evolutionary novelties and demonstrate the interplay of selective and developmental constraints in shaping the evolutionary dynamics. Readily extendable to a broader range of novel characters, our model offers a framework for understanding principles of evolutionary innovations.

Methods

Model of gene regulation

Let there be n regulatory genes (regulators) that are not directly involved in producing the phenotype, but can potentially specify character identities, and m effector genes (effectors) regulated by the regulators. Their expression levels are represented by two column vectors, $\mathbf{X} = \begin{bmatrix} X_1 \\ \dots \\ X_n \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \dots \\ Y_m \end{bmatrix}$, respectively. Per-unit-expression effect of the regulators on the expression of the effectors is captured by a matrix $\mathbf{A}$, where element $A_{i,j}$ is the j-th regulator's effect of on the i-th effector's expression.

We considered two types of functions to model the regulatory effect of the regulators on the effectors but focused on only one of them in this study, as described below.

Power function For the i-th effector gene, the rate of change of its gene product abundance is given by

$$\frac{dY_i}{dt} = \prod_{j=1}^{n} X_j^{A_{i,j}} - \gamma Y_i. \tag{1}$$

where γ is the decay rate of the gene product of the effector. The term $X_j^{A_{i,j}}$ represents the overall regulatory effect of the j-th regulator on the i-th effector, with $A_{i,j}$ representing regulatory effect per unit expression of the j-th regulator. When $A_{i,j} > 0$, $X_j^{A_{i,j}}$ increases with X_j, and the regulatory effect is activation; in contrast, when $A_{i,j} < 0$, $X_j^{A_{i,j}}$ declines with X_j, and the regulatory effect is repression. When $X_j^{A_{i,j}} = 1$, which is reached when $A_{i,j} = 0$, the j-th regulator has no regulatory effect on on Y_i. When $X_j = 1$, Y_i

is unaffected by the value of $A_{i,j}$, marking a critical level of regulator expression at which the regulatory parameter does not matter. The equilibrium of Y_i, which is reached when $\frac{dY_i}{dt} = 0$, is

$$Y_i = \frac{\prod_{j=1}^{n} X_j^{A_{i,j}}}{\gamma}, \tag{2}$$

which, when $\gamma = 1$, becomes

$$Y_i = \prod_{j=1}^{n} X_j^{A_{i,j}}. \tag{3}$$

Log-transforming both sides of the above equation gives

$$\ln Y_i = \sum_{j=1}^{n} A_{i,j} \ln X_j. \tag{4}$$

Expression of all effector genes can be written together as

$$\ln \mathbf{Y} = \begin{bmatrix} \ln Y_1 \\ \dots \\ \ln Y_m \end{bmatrix} = \begin{bmatrix} A_{1,1} & \dots & A_{1,n} \\ \dots & \dots & \dots \\ A_{m,1} & \dots & A_{m,n} \end{bmatrix} \begin{bmatrix} \ln X_1 \\ \dots \\ \ln X_n \end{bmatrix} = \mathbf{A} \ln \mathbf{X}. \tag{5}$$

Under this model, when an effector gene is regulated by two or more regulators, their effect on the effector's expression is multiplicative, which makes the effector's expression is sensitive to low expression levels of the regulators. This property makes this model less appropriate to be applied in certain biological scenarios. For instance, if the the i-th effector gene is activated by two regulators, and the two regulators are paralogs with no sequence divergence such that they have exactly the same regulatory effect ($A_{i,1} = A_{i,2} > 0$), Y_i should be the same when only one paralog is present and when only one paralog is expressed. However, under the above model, the latter condition means extremely negative $A_{i,1} \ln X_1$ or $A_{i,2} \ln X_2$, thereby a very negative $\ln Y_i$, which is obviously not sensible.

Exponential function For the i-th effector gene, the rate of change of its gene product abundance is given by

$$\frac{dY_i}{dt} = \prod_{j=1}^{n} \exp\left(A_{i,j} X_j\right) - \gamma Y_i = \exp\left(\sum_{j=1}^{n} A_{i,j} X_j\right) - \gamma Y_i, \tag{6}$$

where γ is the decay rate of the gene product of the effector. The term $\exp\left(A_{i,j} X_j\right)$ represents the overall regulatory effect of the j-th regulator on the i-th effector, with $A_{i,j}$ representing regulatory effect per unit expression of the j-th regulator. When $A_{i,j} > 0$, $\exp\left(A_{i,j} X_j\right)$ increases with X_j, and the regulatory effect is activation; in contrast, when $A_{i,j} < 0$, $\exp\left(A_{i,j} X_j\right)$ declines with X_j, and the regulatory effect is repression. When $\exp\left(A_{i,j} X_j\right) = 1$, which is reached when $X_j = 0$ or $A_{i,j} = 0$, the j-th regulator has no regulatory effect on on Y_i. The equilibrium of Y_i, which is reached when $\frac{dY_i}{dt} = 0$, is

$$Y_i = \frac{\exp\left(\sum_{j=1}^{n} A_{i,j} X_j\right)}{\gamma}. \tag{7}$$

When $\gamma = 1$, as assumed throughout this study, the above equation becomes

$$Y_i = \exp\left(\sum_{j=1}^{n} A_{i,j} X_j\right). \tag{8}$$

Together, there is

$$\mathbf{Y} = \exp\left(\mathbf{AX}\right) = \exp\left(\begin{bmatrix} A_{1,1} & \dots & A_{1,n} \\ \dots & \dots & \dots \\ A_{m,1} & \dots & A_{m,n} \end{bmatrix} \begin{bmatrix} X_1 \\ \dots \\ X_n \end{bmatrix}\right) = \exp\left(\begin{bmatrix} \sum_{i=1}^{n} A_{1,i} X_1 \\ \dots \\ \sum_{i=1}^{n} A_{n,i} X_i \end{bmatrix}\right). \tag{9}$$

Here, $\mathbf{A}$ is an $m \times n$ matrix that summarizes the strength and direction of each regulator's effect on each effector.

This model is suited for modeling scenarios where an effector gene is activated by two or more regulators and each regulator is sufficient but not necessary for the expression of the effector gene. If an effector is regulated by two regulators, its expression is $Y_i = \exp\left(A_{i,1}X_1 + A_{i,2}X_2\right)$. When it is only the first regulator that is expressed ($X_2 = 0$), there is $Y_i = \exp\left(A_{i,1}X_1\right)$; similarly, if it is the other regulator that is expressed ($X_1 = 0$), there is $Y_i = \exp\left(A_{i,2}X_2\right)$. If the two regulators have exactly the same effect ($A_{i,1} = A_{i,2}$),, which would be the case if they are paralogs resulting from a very recent duplication and have no sequence divergence, the same dosage of X_1 and X_2 will result in the same Y_i. Because of this property, we modeled $\mathbf{Y}$ as an exponential instead of power function of $\mathbf{X}$ when modeling the effect of genes that specify character identities.

Mapping between *cis*-elements and effector genes

The above models implicitly assume each effector gene is regulated through exactly one *cis*-element such that the interaction between a given regulator and a given effector is captured by a single parameter (an element of $\mathbf{A}$). It is plausible, however, that a gene is regulated by multiple *cis*-elements that interact with different *trans*-elements. To recapitulate this possibility, we introduce a more generalized model where the regulation matrix $\mathbf{A}$ is decomposed into two matrices that capture different aspects of regulation. Let there be l *cis*-elements that could potentially bind to the j-th *trans*-factor to regulate the expression of the i-th effector gene. The parameter $A_{i,j}$ can be re-written as

$$A_{i,j} = \sum_{k=1}^{l} \alpha_{i,k}\beta_{k,j}. \tag{10}$$

Here, $\alpha_{i,k}$ captures the effect of the k-th *cis*-element on the i-th effector (per unit expression of the j-th regulator). When $\alpha_{i,k} > 0$, the *cis*-element has an activation effect (e.g., enhancer); when $\alpha_{i,k} < 0$, it is a repressive effect (e.g., silencer). Strength of interaction (binding affinity) between the k-th *cis*-element and the j-th regulator is captured by $\beta_{k,j} \geq 0$. Together, there is

$$\mathbf{A} = \alpha\beta = \begin{bmatrix} \alpha_{1,1} & \dots & \alpha_{1,l} \\ \dots & \dots & \dots \\ \alpha_{m,1} & \dots & \alpha_{m,l} \end{bmatrix} \begin{bmatrix} \beta_{1,1} & \dots & \beta_{1,n} \\ \dots & \dots & \dots \\ \beta_{l,1} & \dots & \beta_{l,n} \end{bmatrix}. \tag{11}$$

Under this modeling framework, mutations in *cis*-elements of the effector genes are modeled as affecting elements of β. With each row of β corresponding to one *cis*-element, each mutation can affect element(s) of only one row.

Developmental control of character state

The observed character state is modeled as a linear combination of log-transformed expression levels of the effector genes. For a given phenotypic trait (z_i), its value is given by

$$z_i = \sum_{j=1}^{m} B_{i,j} \ln Y_j. \tag{12}$$

Here, $B_{i,j}$ is a parameter characterizing the per-unit-expression effect of the j-th effector on z_i. When $B_{i,j} = 0$, Y_j has no effect on z_i. Note that z_i is meant to represent normalized trait values, but not quantities that has to be non-negative such as body mass or absolute abundance of a chemical.

If the phenotype of interest d-dimensional, Eqn. (12) becomes

$$\mathbf{z} = \mathbf{B} \ln \mathbf{Y} = \begin{bmatrix} B_{1,1} & \dots & B_{1,m} \\ \dots & \dots & \dots \\ B_{d,1} & \dots & B_{d,m} \end{bmatrix} \begin{bmatrix} \ln Y_1 \\ \dots \\ \ln Y_m \end{bmatrix}. \tag{13}$$

Here, $\mathbf{z}$ is a vector of length d and $\mathbf{B}$ is a $d \times m$ matrix that summarizes effector genes' per-unit-expression effects on the trait values.

With the effectors' expression levels modeled as an exponential function of the regulators' (Eqn. (9)), Eqn. (13) can be written as

$$\mathbf{z} = \mathbf{BAX}. \tag{14}$$

Modeling characters with conserved identities

We considered a simple scenario where two distinct character identities exist, specified respectively by two distinct regulators (referred to as regulator 1 and regulator 2, respectively). Of the aforementioned models of regulation, the second model, where $\mathbf{Y}$ is modeled as a exponential function of $\mathbf{X}$, was used. Expression profiles of the regulatory genes in two serially homologous body parts were set to be $\mathbf{X}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{X}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, respectively. The two regulators activate the same set of 50 effectors.

An effector may be regulated by a single *cis*-element that ancestrally interacts with both regulators, or by two *cis*-elements that interact respectively with two regulators. If the i-th effector gene is regulated solely by the j-th *cis*-element, the i-th row of $\mathbf{A}$ can be written as

$$\begin{bmatrix} A_{i,1} & A_{i,2} \end{bmatrix} = \begin{bmatrix} \alpha_{i,j} \end{bmatrix} \begin{bmatrix} \beta_{j,1} & \beta_{j,2} \end{bmatrix} = \begin{bmatrix} \alpha_{i,j}\beta_{j,1} & \alpha_{i,j}\beta_{j,2} \end{bmatrix}$$

When a mutation takes place in the j-th *cis*-element, both $\beta_{j,1}$ and $\beta_{j,2}$ will be affected, and the i-th effector gene's expression in both body parts will be affected, resulting in pleiotropic effects on character states of both body parts. We had $\alpha_{i,j} = 1$ throughout the simulations and $\beta_{j,1} = \beta_{j,2} = 1$ ancestrally. Together, the ancestral expression levels of an effector of this category in the two body parts are $Y_{i,1} = \exp \beta_{j,1} = \exp(1)$ and $Y_{i,2} = \exp \beta_{j,2} = \exp(1)$, respectively.

If the i-th effector is regulated by the j-th and k-th *cis*-regulators, each of which interacts with the first and the second regulators, respectively, there is

$$\begin{bmatrix} A_{i,1} & A_{i,2} \end{bmatrix} = \begin{bmatrix} \alpha_{i,j} & \alpha_{i,k} \end{bmatrix} \begin{bmatrix} \beta_{j,1} & \beta_{j,2} \\ \beta_{k,1} & \beta_{k,2} \end{bmatrix} = \begin{bmatrix} \alpha_{i,j}\beta_{j,1} + \alpha_{i,k}\beta_{k,1} & \alpha_{i,j}\beta_{j,2} + \alpha_{i,k}\beta_{k,2} \end{bmatrix}.$$

If a mutation takes place in the j-th *cis*-element, only $\beta_{j,1}$ will be affected; similarly, if a mutation takes place in the k-th *cis*-element, only $\beta_{k,2}$ will be affected. We had $\alpha_{i,j} = \alpha_{i,k} = 1$ throughout the simulations, and $\beta_{j,1} = \beta_{k,2} = 1$ and $\beta_{j,2} = \beta_{k,1} = 0$ ancestrally. Together, the ancestral expression levels of an effector of this category in the two body parts are $Y_{i,1} = \exp(\beta_{j,1} + \beta_{k,1}) = \exp(\beta_{j,1}) = \exp(1)$ and $Y_{i,2} = \exp(\beta_{j,2} + \beta_{k,2}) = \exp(\beta_{k,2}) = \exp(1)$, respectively.

We considered scenarios where 0, 10, 20, 30, 40, and 50 out of the 50 effector genes are regulated by shared targets of the two regulators. The character state of interest of each body part is a single phenotypic dimension. In a diploid organism, each effector has two copies, so $\mathbf{B}$ is a 1×100 matrix. All elements of $\mathbf{B}$ were set to be 0.1. The ancestral state of each body part is thus $100 \times 0.1 \times \ln(\exp 1) = 10$, and the ancestral states of two body parts can be written as $\mathbf{z}_a = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$. We considered three scenarios of directional selection in which the optima are $\mathbf{z}_{opt} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$ (directional selection on z_1 only), $\mathbf{z}_{opt} = \begin{bmatrix} 20 \\ 20 \end{bmatrix}$ (concordant selection), and $\mathbf{z}_{opt} = \begin{bmatrix} 20 \\ 0 \end{bmatrix}$ (divergent selection), respectively.

Fitness is a multivariate Gaussian function of the character state $\mathbf{z}$. The strength of selection along different phenotypic dimensions was characterized by a covariance matrix $\mathbf{S}$. Fitness, denoted ω, is calculated as

$$\omega = \exp\left(-\left(\sqrt{\sum_{i=1}^{n} \frac{P_i^2}{2E_i}}\right)\right). \tag{15}$$

Here, E_i is the i-th eigenvalue of $\mathbf{S}$ and P_i is the i-th element of $\mathbf{P} = (\ln \mathbf{z} - \ln \mathbf{z}_{opt})\mathbf{E}$, where $\mathbf{E}$ is the eigenvector matrix of $\mathbf{S}$. Diagonal elements of $\mathbf{S}$ were all equal to one in the simulations. Off diagonal elements that were considered in our simulations include zero (no correlational selection), 0.9 (positive correlational selection), and -0.9 (negative correlational selection).

Evolutionary simulations were performed using SLiM ([52]). We had population size $N = 1000$ and all simulations lasted for $10N = 10000$ generations. Recombination rate was 0.5 (free recombination) for all simulations. Ten replicate populations were simulated for each parameter combination. Mutation rate was set as 2×10^{-7} per *cis*-element per generation in the simulations. Ten replicate populations were simulated for each parameter combination. For easier implementation of the model in SLiM, we parameterized mutations' effects as effects on character states without explicitly modeling effects on individual regulatory parameters. Each mutation's regulatory effect is summarized as a vector $\vec{\delta} = (\delta_1, \delta_2)$. If the mutation takes place in a *cis*-element that is ancestrally an exclusive target of regulator 1, δ_1 would be sampled from a normal distribution $\mathcal{N}(0, 0.01)$ whereas $\delta_2 = 0$. Similarly, if the *cis*-element is ancestrally an exclusive target of regulator 2, there is $\delta_1 = 0$ and δ_2 would be sampled from $\mathcal{N}(0, 0.01)$. If the mutation takes place in a *cis*-element that is ancestrally a shared target of the two regulators, $\vec{\delta}$ would be sampled from a multivariate normal distribution with zero means and (co)variances characterized by a 2×2 covariance matrix $\mathbf{m}$ with diagonal elements equal to 10^{-4} and off-diagonal elements equal to 9×10^{-5}. For a genotype that carries t mutations (as compared to the ancestral genotype), effects of all the mutations can be written collectively as a $2 \times t$ matrix Δ. Character state of the genotype is calculated as $\mathbf{z} = \mathbf{z}_a \circ \begin{bmatrix} \exp(\sum_{i=1}^{t} \Delta_{1,i}) \\ \exp(\sum_{i=1}^{t} \Delta_{2,i}) \end{bmatrix} = \begin{bmatrix} 10\exp(\sum_{i=1}^{t} \Delta_{1,i}) \\ 10\exp(\sum_{i=1}^{t} \Delta_{2,i}) \end{bmatrix}$. By modeling this way, only mutations in shared targets of two regulators will pleiotropically affect states of two body parts, and mutations do not change signs of $\mathbf{z}$ as it is only elements of β but not α that are affected.

For each population, we calculated means of character states of two body parts across populations ($\bar{z_1}$ and $\bar{z_2}$), respectively) every 100 generations. Mean evolutionary change in z_1 relative to the ancestral value,

calculated $\Delta\bar{z_1} = \bar{z_1} - 10$, was used to represent the overall response to directional selection on z_1. We also calculated character state divergence between two body parts for each population as $|z_1 - z_2|$. The among-population mean $\mathrm{E}(|z_1 - z_2|)$ is used to represent the overall degree of phenotypic divergence between two body parts.

Modeling switching between pre-existing identities

We considered a single body part and its one-dimensional character state z. The body part can potentially assume two identities, denoted I_1 and I_2 respectively, each specified by a regulator. Each regulator activates 50 target effector genes, and each effector is regulated via a single *cis*-element. Targets of the two regulators do not overlap. For convenience, we had effectors regulated by the regulator specifying I_1 correspond to the first 50 rows of $\mathbf{A}$ and the first 50 elements of $\mathbf{B}$. The two sets of effector genes affect the phenotype differently: the first 50 of $\mathbf{B}$ are equal to 0.1 and the rest are equal to -0.1. All elements of α and β (thereby $\mathbf{A}$) are equal to one. Together, given ancestral regulatory parameters, the character state produced by I_1 is $z = 10$ and that produced by I_2 is $z = -10$. The effect of mutations on $\mathbf{X}$ is mediated by a latent trait c, which can be interpreted as corresponding to a morphogen's local concentration [98]. The ancestral value of c is zero. Effects of mutations on c were sampled from a normal distribution $\mathcal{N}(0, 1)$. When $c \geq 0$, $\mathbf{X} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$; when $c < 0$, $\mathbf{X} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. We considered scenarios of selection where the optimal state z_opt is -10, -5, 0, 5, and 10, respectively. Settings of SLiM simulations were the same as those described in the previous section, unless specified. Fitness was a Gaussian function of character state z and the optimum z_{opt}, computed as $\omega = \exp\left(-\frac{z - z_{opt}}{2}\right)$. For each regime of selection, we simulated 20 replicate populations we calculated the fraction of populations where I_2 is fixed (all individuals have $\mathbf{X} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$), the mean expression level of the regulator specifying I_2 across populations ($\bar{X_2}$), and the mean character state across populations ($\bar{z}$).

All analysis of simulation results and generation of plots were done in R [99].

Acknowledgments We thank Joanna Wolfe and Mark Kim for helpful comments to an earlier version of the manuscript. DJ and LS are supported by Okinawa Institute of Science and Technology Graduate University.

Data and materials availability: Code and data files are available at `https://github.com/RexJiangEvoBio/novelty_evo_model`.

References

[1] Gould, S. J & Lewontin, R. C. (1979) *Proceedings of the Royal Society B: Biological Sciences* **205**, 581–598.

[2] Smith, J. M, Burian, R, Kauffman, S, Alberch, P, Campbell, J, Goodwin, B, Lande, R, Raup, D, & Wolpert, L. (1985) *The Quarterly Review of Biology* **60**, 265–287.

[3] Yampolsky, L. Y & Stoltzfus, A. (2001) *Evolution & Development* **3**, 73–83.

[4] Hendrikse, J. L, Parsons, T. E, & Hallgrímsson, B. (2007) *Evolution & Development* **9**, 393–401.

[5] Kavanagh, K. D, Shoval, O, Winslow, B. B, Alon, U, Leary, B. P, Kan, A, & Tabin, C. J. (2013) *Proceedings of the National Academy of Sciences* **110**, 18190–18195.

[6] Fritz, J. A, Brancale, J, Tokita, M, Burns, K. J, Hawkins, M. B, Abzhanov, A, & Brenner, M. P. (2014) *Nature Communications* **5**, 3700.

[7] Stoltzfus, A. (2021) *Mutation, randomness, and evolution*. (Oxford University Press).

[8] Rohner, P. T & Berger, D. (2023) *Proceedings of the National Academy of Sciences* **120**, e2211210120.

[9] Rohner, P. T & Berger, D. (2025) *Nature Ecology & Evolution* pp. 1–13.

[10] Carroll, S. B. (1995) *Nature* **376**, 479–485.

[11] Wagner, G. P & Stadler, P. F. (2003) *Journal of Theoretical Biology* **220**, 505–527.

[12] Wagner, G. P. (2007) *Nature Reviews Genetics* **8**, 473–479.

[13] Wagner, G. P. (2014) *Homology, genes, and evolutionary innovation*. (princeton university press).

[14] Musser, J. M & Wagner, G. P. (2015) *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **324**, 588–604.

[15] Parker, J & Pennell, M. (2025) *Current Biology* **35**, R626–R637.

[16] Warren, R. W, Nagy, L, Selegue, J, Gates, J, & Carroll, S. (1994) *Nature* **372**, 458–461.

[17] Weatherbee, S. D, Halder, G, Kim, J, Hudson, A, & Carroll, S. (1998) *Genes & Development* **12**, 1474–1482.

[18] Weatherbee, S. D, Nijhout, H. F, Grunert, L. W, Halder, G, Galant, R, Selegue, J, & Carroll, S. (1999) *Current Biology* **9**, 109–115.

[19] Deutsch, J. (2005) *Bioessays* **27**, 673–675.

[20] Tomoyasu, Y, Wheeler, S. R, & Denell, R. E. (2005) *Nature* **433**, 643–647.

[21] Hersh, B. M, Nelson, C. E, Stoll, S. J, Norton, J. E, Albert, T. J, & Carroll, S. B. (2007) *Developmental Biology* **302**, 717–727.

[22] Tomoyasu, Y, Arakane, Y, Kramer, K. J, & Denell, R. E. (2009) *Current Biology* **19**, 2057–2065.

[23] Tong, X, Hrycaj, S, Podlaha, O, Popadic, A, & Monteiro, A. (2014) *Developmental Biology* **394**, 357–366.

[24] Matsuoka, Y & Monteiro, A. (2022) *Development* **149**, dev200781.

[25] Ohuchi, H, Takeuchi, J, Yoshioka, H, Ishimaru, Y, Ogura, K, Takahashi, N, Ogura, T, & Noji, S. (1998) *Development* **125**, 51–60.

[26] Agarwal, P, Wylie, J. N, Galceran, J, Arkhitko, O, Li, C, Deng, C, Grosschedl, R, & Bruneau, B. G. (2003) *Development* **130**, 623–633.

[27] Naiche, L & Papaioannou, V. E. (2003) *Development* **130**, 2681–2693.

[28] Minguillon, C, Del Buono, J, & Logan, M. P. (2005) *Developmental Cell* **8**, 75–84.

[29] Ouimette, J.-F, Jolin, M. L, L'honoré, A, Gifuni, A, & Drouin, J. (2010) *Nature Communications* **1**, 35.

[30] Duboc, V & Logan, M. P. (2011) *Developmental Dynamics* **240**, 1017–1027.

[31] Petit, F, Sears, K. E, & Ahituv, N. (2017) *Nature Reviews Genetics* **18**, 245–258.

[32] Shirasaki, R & Pfaff, S. L. (2002) *Annual Review of Neuroscience* **25**, 251–281.

[33] Boyer, L. A, Lee, T. I, Cole, M. F, Johnstone, S. E, Levine, S. S, Zucker, J. P, Guenther, M. G, Kumar, R. M, Murray, H. L, Jenner, R. G, et al. (2005) *Cell* **122**, 947–956.

[34] Lee, S, Lee, B, Joshi, K, Pfaff, S. L, Lee, J. W, & Lee, S.-K. (2008) *Developmental Cell* **14**, 877–889.

[35] Arendt, D. (2008) *Nature Reviews Genetics* **9**, 868–882.

[36] Graf, T & Enver, T. (2009) *Nature* **462**, 587–594.

[37] Hashimshony, T, Feder, M, Levin, M, Hall, B. K, & Yanai, I. (2015) *Nature* **519**, 219–222.

[38] Kin, K, Nnamani, M. C, Lynch, V. J, Michaelides, E, & Wagner, G. P. (2015) *Cell Reports* **10**, 1398–1409.

[39] Dai, Y, Zhong, Y, Pan, R, Yuan, L, Fu, Y, Chen, Y, Du, J, Li, M, Wang, X, Liu, H, et al. (2024) *Nature Ecology & Evolution* pp. 1–18.

[40] Raup, D. M. (1966) *Journal of Paleontology* pp. 1178–1190.

[41] Raup, D. M. (1968) *Journal of Paleontology* **42**, 50–63.

[42] Salazar-Ciudad, I & Jernvall, J. (2010) *Nature* **464**, 583–586.

[43] Machado, F. A, Mongle, C. S, Slater, G, Penna, A, Wisniewski, A, Soffin, A, Dutra, V, & Uyeda, J. C. (2023) *Nature Ecology & Evolution* pp. 1–11.

[44] Prum, R. O. (1999) *Journal of Experimental Zoology* **285**, 291–306.

[45] Norell, M. A & Xu, X. (2005) *Annual Review of Earth and Planetary Sciences* **33**, 277–299.

[46] Sallan, L. (2016) *Current Biology* **26**, R1224–R1225.

[47] Cumplido, N, Arratia, G, Desvignes, T, Muñoz-Sánchez, S, Postlethwait, J. H, & Allende, M. L. (2024) *Science Advances* **10**, eadj5991.

[48] Hintermann, A, Bolt, C. C, Hawkins, M. B, Valentin, G, Lopez-Delisle, L, Gitto, S, Gómez, P. B, Mascrez, B, Mansour, T. A, Nakamura, T, et al. (2025) *Nature*.

[49] DiFrisco, J, Love, A. C, & Wagner, G. P. (2020) *Biology & Philosophy* **35**, 44.

[50] DiFrisco, J, Wagner, G. P, & Love, A. C. (2023) *Reframing research on evolutionary novelty and co-option: character identity mechanisms versus deep homology*. (Elsevier), Vol. 145, pp. 3–12.

[51] Arendt, D, Musser, J. M, Baker, C. V, Bergman, A, Cepko, C, Erwin, D. H, Pavlicev, M, Schlosser, G, Widder, S, Laubichler, M. D, et al. (2016) *Nature Reviews Genetics* **17**, 744–757.

[52] Haller, B. C & Messer, P. W. (2023) *The American Naturalist* **201**, E127–E139.

[53] Hether, T. D & Hohenlohe, P. A. (2014) *Evolution* **68**, 950–964.

[54] Jiang, D, Cope, A. L, Zhang, J, & Pennell, M. (2023) *Molecular Biology and Evolution* **40**, msad169.

[55] Pouzet, S & Le Rouzic, A. (2025) *Evolution* p. qpaf068.

[56] Mallo, M. (2018) *Trends in Genetics* **34**, 209–217.

[57] Ingham, P. W & Placzek, M. (2006) *Nature Reviews Genetics* **7**, 841–850.

[58] Tang, L, Hill, M. C, Wang, J, Wang, J, Martin, J. F, & Li, M. (2020) *Genome Research* **30**, 1835–1845.

[59] Liu, J & Robinson-Rechavi, M. (2020) *Science Advances* **6**, eabc9863.

[60] Vaishnav, E. D, de Boer, C. G, Molinet, J, Yassour, M, Fan, L, Adiconis, X, Thompson, D. A, Levin, J. Z, Cubillos, F. A, & Regev, A. (2022) *Nature* **603**, 455–463.

[61] Schweizer, G & Wagner, A. (2021) *Genome Biology and Evolution* **13**, evab273.

[62] Zhang, X, Fang, B, & Huang, Y.-F. (2023) *Nature Communications* **14**, 783.

[63] Xu, H, Li, C, Xu, C, & Zhang, J. (2023) *Nature Communications* **14**, 1826.

[64] Smith, G. D, Ching, W. H, Cornejo-Páramo, P, & Wong, E. S. (2023) *Genome Biology* **24**, 116.

[65] Jiang, D & Zhang, J. (2024) *Molecular Biology and Evolution* **41**, msad284.

[66] Wagner, G. P & Altenberg, L. (1996) *Evolution* **50**, 967–976.

[67] Wagner, G. P, Pavlicev, M, & Cheverud, J. M. (2007) *Nature Reviews Genetics* **8**, 921–931.

[68] Clune, J, Mouret, J.-B, & Lipson, H. (2013) *Proceedings of the Royal Society B: Biological Sciences* **280**, 20122863.

[69] Goswami, A, Smaers, J. B, Soligo, C, & Polly, P. D. (2014) *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130254.

[70] Ciliberti, S, Martin, O. C, & Wagner, A. (2007) *Proceedings of the National Academy of Sciences* **104**, 13591–13596.

[71] Rueffler, C, Hermisson, J, & Wagner, G. P. (2012) *Proceedings of the National Academy of Sciences* **109**, E326–E335.

[72] Holstad, A, Voje, K. L, Opedal, Ø. H, Bolstad, G. H, Bourg, S, Hansen, T. F, & Pélabon, C. (2024) *Science* **384**, 688–693.

[73] Arnold, S. J. (2023) *Evolutionary Quantitative Genetics*. (Oxford University Press, Oxford, UK).

[74] Pennell, M & Jiang, D. (2024) The macroevolutionary adaptive landscape: more than a metaphor?

[75] Maher, A. E, Burin, G, Cox, P. G, Maddox, T. W, Maidment, S. C, Cooper, N, Schachner, E. R, & Bates, K. T. (2022) *Nature Communications* **13**, 4340.

[76] Scholtz, G & Edgecombe, G. D. (2006) *Development Genes and Evolution* **216**, 395–415.

[77] Jockusch, E. L. (2017) *Integrative and Comparative Biology* **57**, 533–545.

[78] Ortega-Hernández, J, Janssen, R, & Budd, G. E. (2017) *Arthropod Structure & Development* **46**, 354–379.

[79] Young, N. M, Wagner, G. P, & Hallgrímsson, B. (2010) *Proceedings of the National Academy of Sciences* **107**, 3400–3405.

[80] Cretekos, C. J, Wang, Y, Green, E. D, Martin, J. F, Rasweiler, J. J, Behringer, R. R, Program, N. C. S, et al. (2008) *Genes & Development* **22**, 141–151.

[81] Greer, A. E. (1991) *Journal of Herpetology* pp. 166–173.

[82] Bejder, L & Hall, B. K. (2002) *Evolution & Development* **4**, 445–458.

[83] Kvon, E. Z, Kamneva, O. K, Melo, U. S, Barozzi, I, Osterwalder, M, Mannion, B. J, Tissières, V, Pickle, C. S, Plajzer-Frick, I, Lee, E. A, et al. (2016) *Cell* **167**, 633–642.

[84] Harris, M. P, Fallon, J. F, & Prum, R. O. (2002) *Journal of Experimental Zoology* **294**, 160–176.

[85] Prin, F & Dhouailly, D. (2004) *The International Journal of Developmental Biology* **48**, 137–148.

[86] Cooper, R. L & Milinkovitch, M. C. (2023) *Science Advances* **9**, eadg9619.

[87] Benton, M. J. (2014) *Vertebrate palaeontology*. (John Wiley & Sons).

[88] Kang, S, Graham Jr, J. M, Olney, A. H, & Biesecker, L. G. (1997) *Nature Genetics* **15**, 266–268.

[89] Turelli, M. (1985) *Genetics* **111**, 165–195.

[90] Zhang, X.-S & Hill, W. G. (2005) *Trends in Ecology & Evolution* **20**, 468–470.

[91] McGuigan, K, Rowe, L, & Blows, M. W. (2011) *Trends in Ecology & Evolution* **26**, 22–29.

[92] Jiang, D & Zhang, J. (2020) *Evolution* **74**, 2158–2167.

[93] Galis, F. (1999) *Journal of Experimental Zoology* **285**, 19–26.

[94] Galis, F, Dooren, T. J. V, Feuth, J. D, Metz, J. A, Witkam, A, Ruinard, S, Steigenga, M. J, & Wunaendts, L. C. (2006) *Evolution* **60**, 2643–2654.

[95] Varela-Lasheras, I, Bakker, A. J, van der Mije, S. D, Metz, J. A, van Alphen, J, & Galis, F. (2011) *EvoDevo* **2**, 1–27.

[96] Böhmer, C, Rauhut, O. W, & Wörheide, G. (2015) *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150077.

[97] Cerbus, R. T, Hiratani, I, & Kawaguchi, K. (2024) *Proceedings of the National Academy of Sciences* **121**, e2411421121.

[98] Wolpert, L. (1969) *Journal of Theoretical Biology* **25**, 1–47.

[99] R Core Team. (2025) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).

Figures

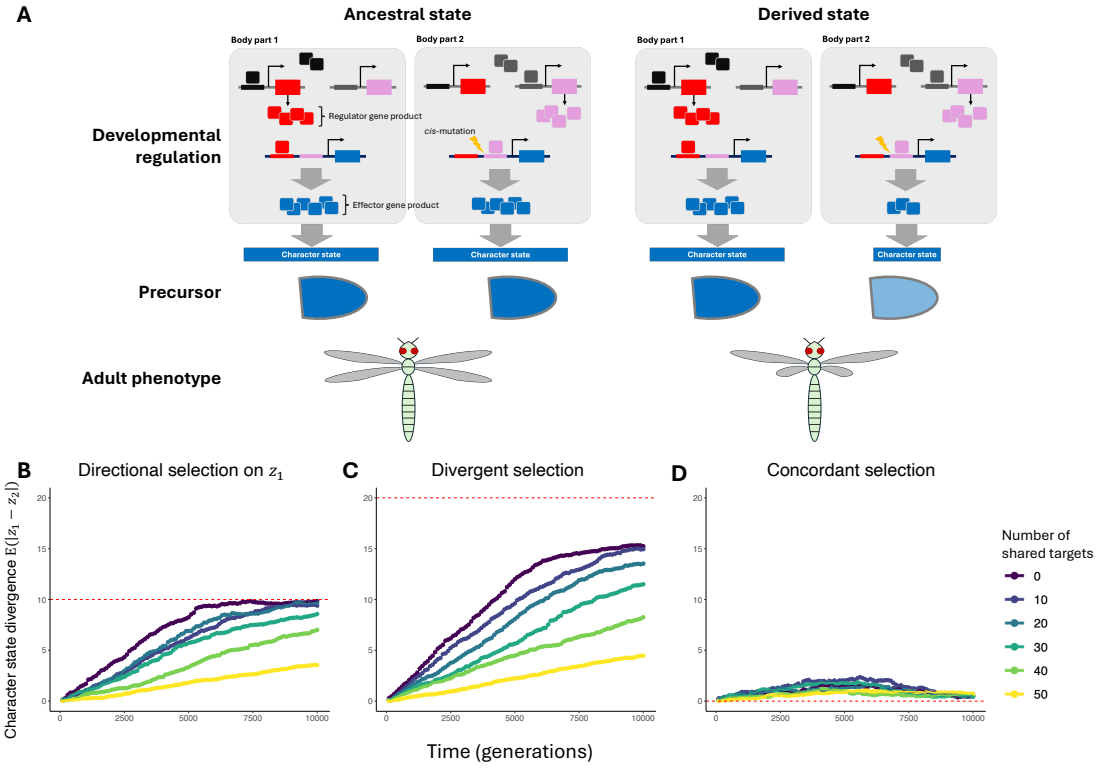

Figure 1: Character state divergence between body parts with conserved identities. (A) Schematic illustration of the scenario, where two regulators specifying distinct identities activate the same effector gene via different *cis*-elements, and the effector gene's expression level determines the character state (i.e., length of wings). The *cis*-mutation shown affects the interaction between a regulator and a *cis*-element, thereby effector expression level in body part 2. Upper row: developmental regulation in two body parts. Middle row: the developing precursor part, with color corresponding to concentration of the effector gene product. Lower row: character state in the adult organism. (B-D) Character state divergence between body parts with conserved identities over time, calculated every 100 generations, with color corresponds to the number of effector genes regulated via shared targets of two regulators. (B) Only z_1 is under directional selection. (B) Two body parts are under divergent selection. (C) Two body parts are under concordant selection.

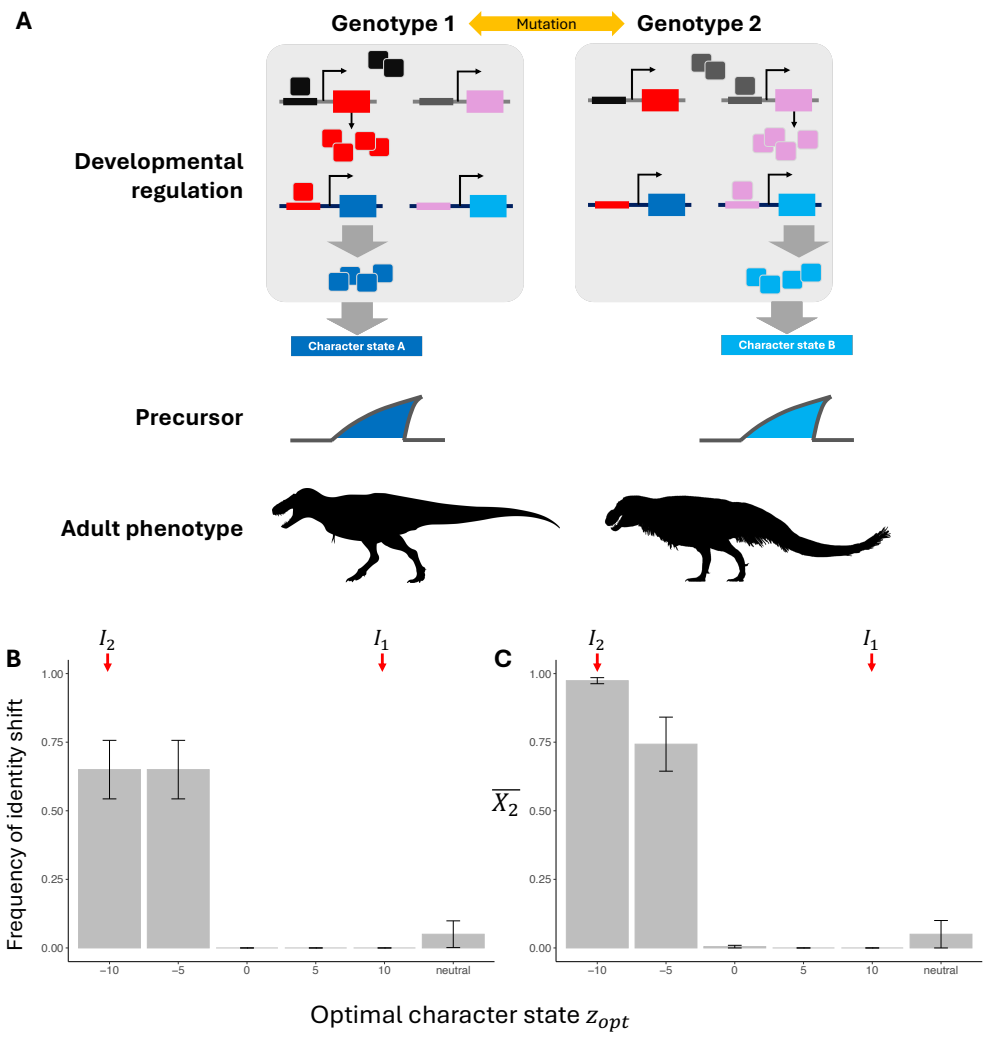

Figure 2: Switching between pre-existing identities. (A) Schematic illustration of the scenario. Left and right columns represent two genotypes where different regulators are expressed, which results in different character identities and states. Upper row: developmental regulation in two body parts. Middle row: the developing precursor part, with color indicating which effector genes are expressed. Lower row: character state in the adult organism. (B) Fraction of simulated populations where the non-ancestral identity I_2 is fixed. (C) Mean expression level of the regulator underlying I_2 across populations. Height of each error bar in (B) and (C) is twice the standard error (sample size $n = 20$). Character states produced by I_1 and I_2 give ancestral regulatory parameters are marked with arrows in (B) and (C). *T. rex* silhouettes in (B) are from PhyloPic (`https://www.phylopic.org/`), and were created by Matt Dempsey (left, non-feathered) and Matt Martyniuk (right, feathered), respectively.

Supplementary Materials

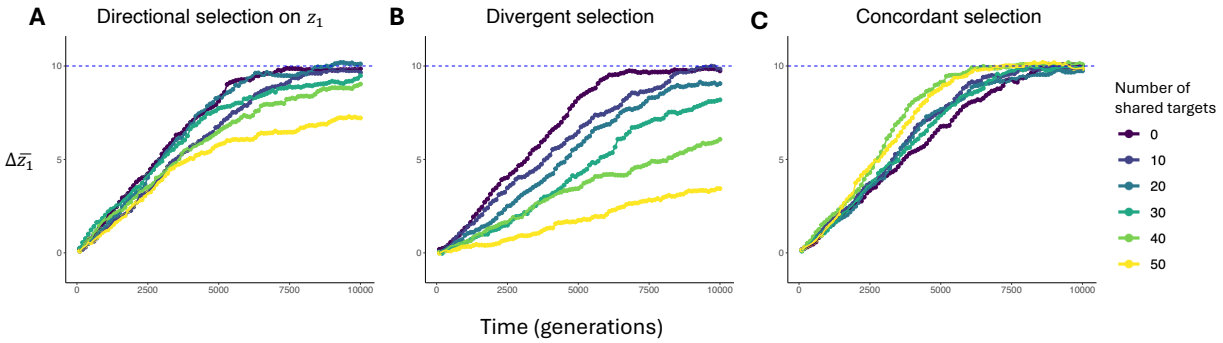

Figure S1: Mean evolutionary change in z_1 from the ancestral value over time, calculated every 100 generations. Scenarios of selection are the same as those in Fig. 1. (B) Color corresponds to the number of effector genes regulated via shared targets of two regulators.

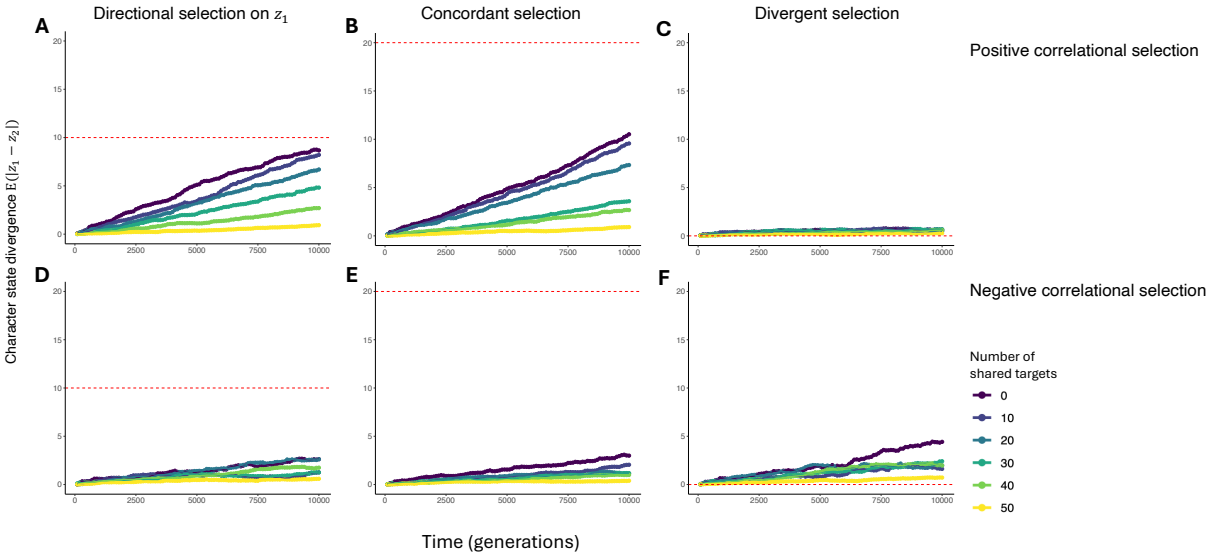

Figure S2: Character state divergence between body parts with conserved identities over time in the presence of both directional and correlational selection, calculated every 100 generations. (A-C) Positive correlational selection. (D-F) Negative correlational selection. Scenarios of directional selection are designated and arranged in the same way as in Fig. 1B-D. Color corresponds to the number of effector genes regulated via shared targets of two regulators.

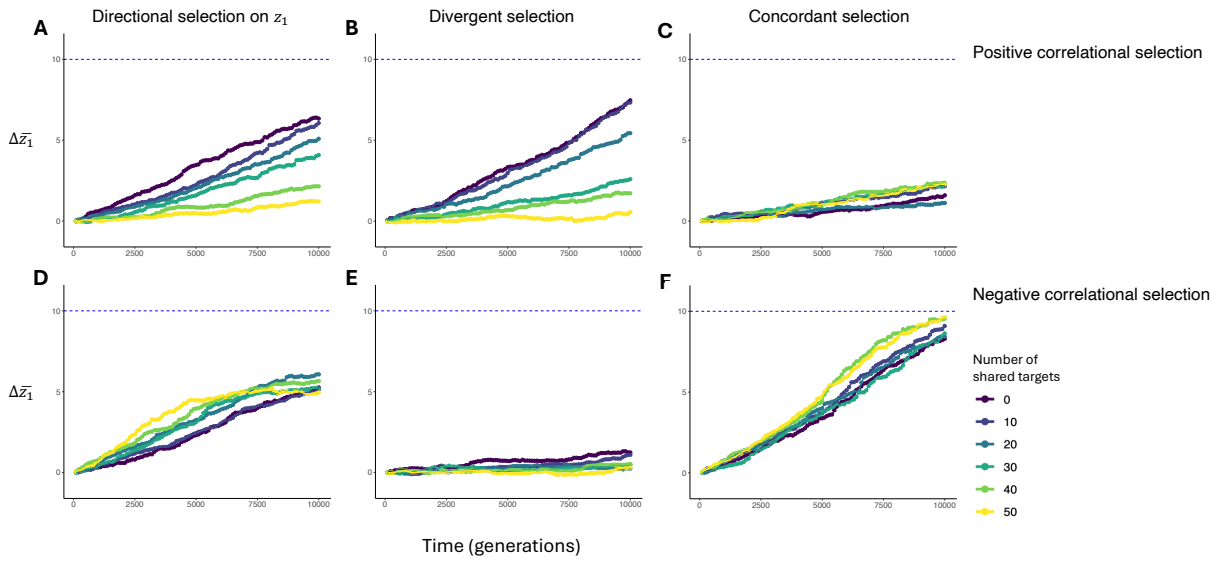

Figure S3: Mean evolutionary change in z_1 from the ancestral value over time in the presence of correlational selection, calculated every 100 generations. Scenarios of selection are the same as those in Fig. S2. Color corresponds to the number of effector genes regulated via shared targets of two regulators.

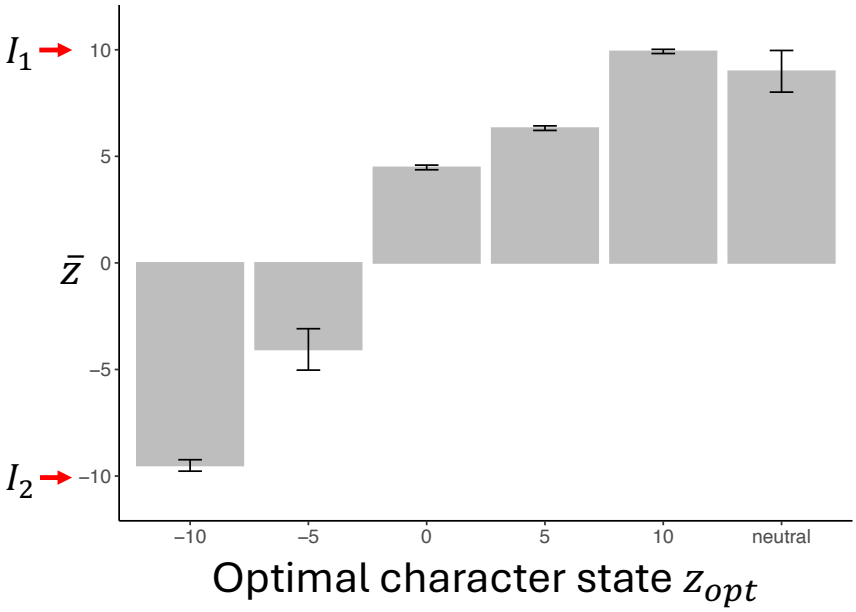

Figure S4: Mean character state of a body part that can switch between pre-existing identities. Character states produced by I_1 and I_2 give ancestral regulatory parameters are marked with arrows. Height of each error bar is twice the standard error (sample size $n = 20$).