

Article

A Family of Fitness Landscapes Modeled through Gene Regulatory Networks

Chia-Hung Yang ^{1,*}  and Samuel V. Scarpino ^{1,2,3,4,5,6,*}

¹ Network Science Institute, Northeastern University, Boston, MA 02115, USA

² Physics Department, Northeastern University, Boston, MA 02115, USA

³ Roux Institute, Northeastern University, Boston, MA 02115, USA

⁴ Institute for Experiential AI, Northeastern University, Boston, MA 02115, USA

⁵ Santa Fe Institute, Santa Fe, NM 87501, USA

⁶ Vermont Complex Systems Center, University of Vermont, Burlington, VT 05405, USA

* Correspondence: yang.chi@northeastern.edu (C.-H.Y.); s.scarpino@northeastern.edu (S.V.S.)

Abstract: Fitness landscapes are a powerful metaphor for understanding the evolution of biological systems. These landscapes describe how genotypes are connected to each other through mutation and related through fitness. Empirical studies of fitness landscapes have increasingly revealed conserved topographical features across diverse taxa, e.g., the accessibility of genotypes and “ruggedness”. As a result, theoretical studies are needed to investigate how evolution proceeds on fitness landscapes with such conserved features. Here, we develop and study a model of evolution on fitness landscapes using the lens of Gene Regulatory Networks (GRNs), where the regulatory products are computed from multiple genes and collectively treated as phenotypes. With the assumption that regulation is a binary process, we prove the existence of empirically observed, topographical features such as accessibility and connectivity. We further show that these results hold across arbitrary fitness functions and that a trade-off between accessibility and ruggedness need not exist. Then, using graph theory and a coarse-graining approach, we deduce a mesoscopic structure underlying GRN fitness landscapes where the information necessary to predict a population’s evolutionary trajectory is retained with minimal complexity. Using this coarse-graining, we develop a bottom-up algorithm to construct such mesoscopic backbones, which does not require computing the genotype network and is therefore far more efficient than brute-force approaches. Altogether, this work provides mathematical results of high-dimensional fitness landscapes and a path toward connecting theory to empirical studies.

Keywords: fitness landscapes; gene regulatory networks; coarse-graining; biological computation; graph theory



Citation: Yang, C.-H.; Scarpino, S.V. A Family of Fitness Landscapes Modeled through Gene Regulatory Networks. *Entropy* **2022**, *24*, 622. <https://doi.org/10.3390/e24050622>

Academic Editors: David Wolpert and Jessica Flack

Received: 2 December 2021

Accepted: 26 April 2022

Published: 29 April 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since its introduction by Wright [1], the concept of fitness landscapes has grown and matured into a cornerstone of biology [2–4]. A fitness landscape consists of a space of genotypes that are mutually accessible through mutations and a fitness value associated with the phenotype each genotype encodes. In this context, fitness describes the evolutionary potential of each genotype, and the set of navigable genotypes on these landscapes is termed the genotype network [5]. Continuing with this metaphor, the evolution of a population can be depicted as a trajectory wandering on the fitness landscape. As a consequence, the topography of a fitness landscape sheds light on various evolutionary processes, including constraints on adaptation [6–9], speciation via genetic incompatibilities [10,11], (dis)advantages of sexual reproduction and recombination [12–14], the repeatability/reversibility (or not) of evolutionary trajectories [15–18], and the role of neutral networks—components of the genotype network with the same fitness—in epochal evolution [19–24].

Despite being introduced by Wright [1], Fisher's 1930 geometric model of adaptation is the first mathematical model of evolution on what we now call fitness landscapes [15,25,26]. Later work by Kingman [27] and Kauffman and Levin [28] constructed what they termed a "house of cards" (HoC) model where fitness values for each genotype are drawn independently from a specified probability distribution. Building on the HoC model, Kauffman and Weinberger [29] introduced the NK model, which forces each locus to interact with a fixed number of other loci and where a genotype's fitness becomes the sum of the fitness contributions of every interaction group. More recently, the "rough Mount Fuji model" [30,31] combines the HoC landscape with an additional field penalizing a genotype's Hamming distance away from a referenced genotype with the optimal fitness. The dependence of a genotype's fitness on that of neighboring genotypes is thought to be a key feature of empirical fitness landscapes.

Over the past three decades, the fitness landscapes for various organisms, including bacteria [32–34], fungi [35,36], and fruit flies [37], have been empirically reconstructed. While the number of genotypes included in these early landscapes was limited, modern sequencing techniques and high-throughput analyses have enabled the construction of many large landscapes. Notable studies have been conducted in HIV [38,39], yeast [40], *E. coli* [41], jellyfish [42], human cancers [43], human stem cells [44], and DNA/RNA networks [45–49]. Comprehensive landscapes for multiple eukaryotic species have also been analyzed based on the binding affinity of transcription factors [50] and after accounting for the ecological context the species experiences [51]. What emerged from these studies is a set of prominent topographical features conserved across diverged taxa [52].

Together, empirical and modeled fitness landscapes exhibit three key topographical features. First, fitness landscapes are more often "rugged" than smooth [52]. The degree of ruggedness can be assessed via a variety of measures, such as the roughness of the slope ratio [53,54] and the number of local fitness maxima [55], which are often strongly correlated with each other [4]. Empirical studies typically show moderate ruggedness in the observed fitness landscapes [32–34,36,50,56]. The degree of ruggedness in these empirical landscapes is less than the HoC model assumes and comparable to a fine-tuned NK model or rough Mount Fuji model [4]. Second, fitness landscapes reveal mutational trajectories from one genotype to another where the fitness is non-decreasing, which implies accessibility (typically to a fitness optimum) across the landscape [32,57–60]. Lastly, whereas the inaccessible region in the HoC model expands when distant from the fitness optimum [61], other models find accessible trajectories despite high genotypic dimensionality [62,63].

Due to the often pervasive interaction between loci, determining phenotype from genotype can have a high degree of computational complexity [64,65]. Many existing fitness landscape models have dealt with this complexity by strongly constraining the state-space of possible genotypic interactions and/or reducing the complexity of how information is processed when mapping genotype to phenotype. For example, studies have focused on the folded structure of short RNA sequences, where the resulting stability or affinity is a fitness proxy [66–70], networks of molecular/genetic pathways whose expression pattern or homeostasis determines fitness [71–75], and modular mutational effects at different loci in Fisher's geometric model [76].

Here, we model the genotype–phenotype map using the pathway framework of gene regulatory networks (GRNs), where mechanistic knowledge of how phenotypes are computed from genotypes is encoded in the GRN (see [77,78] for a more formal introduction). To study the fitness landscapes induced by GRN evolution, we integrate the pathway framework into a family of fitness landscape models where the fitness value is uniquely determined by the phenotype corresponding to the regulatory outcome of a genotype. For a fitness landscape of GRNs, we first prove the existence of two key topographical features: (a) GRNs with the same phenotype are themselves connected in the underlying genotypic network, and (b) there exists accessible trajectories between all pairs of GRNs with similar phenotypes. Second, utilizing the idea of symmetries and automorphisms in the genotype network, we coarse-grain GRNs into groups with equivalent roles in the fitness landscapes

and deduce an underlying mesoscopic structure with which we can predict the trajectory of evolution with minimal complexity. Lastly, using this coarse-graining, we develop a bottom-up algorithm for constructing the underlying fitness landscape of GRNs, which does not require computing the genotype network and is thus more efficient than the conventional brute-force approach.

2. Methods

Here, we introduce a family of fitness landscape models where the genotype-phenotype mapping is constructed from regulatory interactions. We first summarize a modeling framework of GRNs proposed in our previous work [77,78], termed the pathway framework, and then fitness landscape models of GRNs built upon the pathway framework.

2.1. Pathway Framework of GRNs

Genotypes in the pathway framework of GRNs contain all necessary information to construct a regulatory network [77,78]. More specifically, alleles at each locus include both a transcription activator and a protein product, which means the regulatory interactions among the loci can be deduced by connecting genes whose expression product corresponds to the activator of another. Compared to existing work on regulatory circuits—where mutations are modeled as rewiring a single interaction between genes [79,80]—the pathway framework considers a mutation as changing the activator/product of a gene. Lastly, the phenotype is determined by the set of loci reached in a regulatory cascade induced by external stimuli. These stimuli could be completely external to the individual or simply come from another regulatory network in the organism. For additional details on the pathway framework, see [77,78] and Figure 1 for illustration.

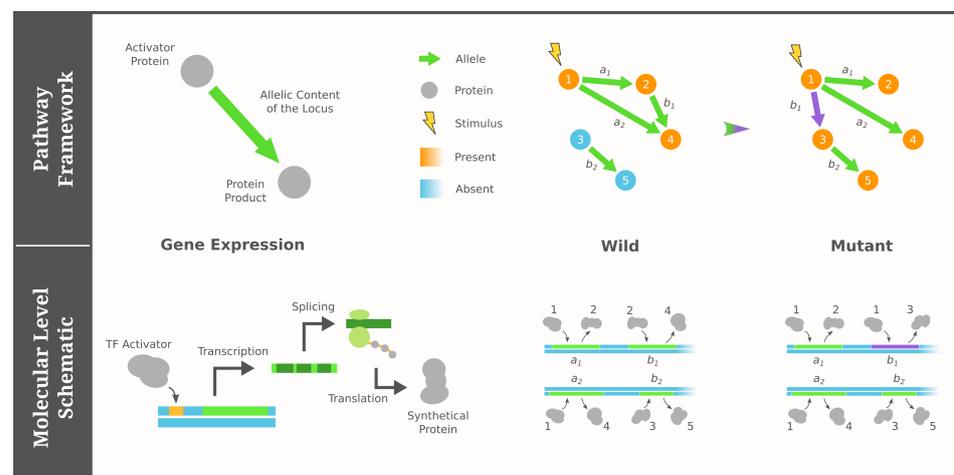


Figure 1. Cartoon illustration of the pathway framework of GRNs adapted from [77,78]. Under our four simplified assumptions, a GRN (genotype) consists of a fixed number of proteins as nodes and a constant number of directed edges depicting the activator/product pairs of genes. The phenotype is modeled as the Boolean states of proteins (colored), which are determined by their reachability from the external stimulus (lightning icon).

In this work, when building the family of fitness landscape models, we restrict the pathway framework with four assumptions. First, we consider a fixed set of genes underlying the genotypes, i.e., gene duplication and deletion events are excluded. Second, we assume a fixed underlying collection of proteins that can possibly exist in the organism. Third, we consider the case where a gene's expression is activated by a specific protein, and it generates only one protein product. Fourth, we assume that the associated chemical state of each protein is modeled as a Boolean/binary variable (present or absent), and external environmental signals stimulate the existence of specific proteins in the organism. As a

consequence, the Boolean state of a phenotype-related protein is determined by whether it is reached by a regulatory cascade starting from an initial stimulus.

While the above assumptions seem naive, as we will show in Section 3, this simplified model still predicts the topographical features observed in empirical landscapes (see Section 1). As a result of these assumptions, we are able to derive rigorous theoretical insights into GRN evolution and obtain fitness landscapes consistent with far more complicated models. We believe these assumptions are conservative with respect to the biology and a justified starting point for modeling fitness landscapes, and we discuss the implications of these assumptions and possible extensions to the model in Section 4.

2.2. Fitness Landscape of GRNs under the Pathway Framework

Let Γ and Ω be the fixed, underlying collection of loci and proteins, respectively. A genotype is represented by its GRN g such that every locus $\gamma \in \Gamma$ is associated with a protein activator/product pair $e_g(\gamma) = (u, v)$, $u, v \in \Omega$. Equivalently, any GRN is a directed graph with $|\Omega|$ nodes labeled by the proteins Ω and $|\Gamma|$ edges labeled by the loci Γ . In the rest of this paper, we will use the terminology “source/target” node of edge γ interchangeably to refer to the protein activator/product of locus γ . We also write \mathcal{G} to be the set of all GRNs with the underlying loci Γ and proteins Ω .

The backbone of a fitness landscape of GRNs, i.e., the genotype network, is an undirected network of networks encoding the mutational relationship between the GRNs. Let G be the genotype network, and we denote its mega-nodes by $V(G) = \mathcal{G}$ and its edges by $E(G)$. There is an edge $(g_1, g_2) \in E(G)$ between any GRNs $g_1, g_2 \in \mathcal{G}$ when they only differ by the allele of a single locus γ , $e_{g_1}(\gamma) \neq e_{g_2}(\gamma)$. In other words, g_1 and g_2 are connected in G when they can be transformed into each other through one edge rewiring.

Furthermore, we write x_ω to be the binary state of protein $\omega \in \Omega$, where $x_\omega = 1$ indicates the presence of ω , and $x_\omega = 0$ designates its absence. We also partition Ω into three disjoint groups: (a) proteins Ω_0 whose presence is externally stimulated by the given environment, (b) proteins $\hat{\Omega}$ whose states influence the fitness value, and (c) the remaining ones, which we call the dummy proteins Ω' since their specific identities are irrelevant to the external environment and the resultant phenotypes/fitness. (In this paper, we assume that the stimuli Ω_0 must be proteins that cannot be produced by expression, and we leave no constraint to the fitness-relevant and dummy proteins $\hat{\Omega}$ and Ω').

A phenotype is then treated as a vector of zeros and ones, where each entry corresponds to the binary state of a protein in $\hat{\Omega}$. The resultant phenotype $\mathbf{x}_{\hat{\Omega}}(g)$ of a GRN g is determined by the reachability in g : For any $\omega \in \hat{\Omega}$, $x_\omega = 1$ if and only if there is a stimulus $\omega_0 \in \Omega_0$ and a path from ω_0 to ω in g , which represents a chain of sequentially expressed genes that generates protein ω . Finally, the fitness f is simply a function of the phenotype $\mathbf{x}_{\hat{\Omega}}(g)$.

Combined, a fitness landscape of GRNs is characterized by three key elements: the genotype network G , the external stimuli Ω_0 , and the fitness function of phenotype f (which implicitly identifies the fitness-relevant proteins $\hat{\Omega}$). The genotype network G serves as the skeleton of the fitness landscape, whereas the environment-dependent stimuli Ω_0 and fitness function f determine the phenotypes of GRNs and their selective advantages.

3. Results

In this work, we derive three theoretical insights into fitness landscape models using GRNs as the embedded genotype–phenotype mapping. First, we show that the resulting family of fitness landscapes must always contain two topographical properties: connectivity, i.e., GRNs with the same phenotype can be mutually reached via mutations, and accessibility, i.e., that any GRN can be reached from an arbitrary less-fit GRN (once certain similarity criterion is met). Second, we propose a mesoscopic coarse-graining for fitness landscapes, which is a more compact alternative to analyzing evolutionary processes than the original landscape. This mesoscopic backbone recognizes “symmetries” in the genotype network, and it aggregates GRNs with the same role in the fitness landscape into a

single representative genotype. Third, we provide a bottom-up approach to algorithmically construct this mesoscopic backbone and demonstrate its efficiency over coarse-graining the genotype network using brute force.

3.1. Connectivity and Accessibility in a Fitness Landscape of GRNs

A fitness landscape model of GRNs features a handful of properties that have either been discovered in empirical fitness landscapes or investigated mathematically. First, its underlying space, i.e., the genotype network G , presents immense dimensionality. Second, the fitness function f is flexible and can effectively tune the ruggedness of the fitness landscape. For example, a highly rugged “holey” landscape can be modeled by a binary f such that any GRN $g \in \mathcal{G}$ has high fitness once some single protein $\omega \in \Omega$ is present, $x_\omega = 1$, and otherwise, g has low/zero fitness. Because one can always find several mutational neighbors of g whose phenotype shows an opposite state x_ω , the resultant fitness landscape is inevitably rugged. In what follows, we further show that fitness landscape models of GRNs must hold the characteristics of connectivity and accessibility.

Let \mathbf{y} be a phenotype and denote by $\mathcal{G}_\mathbf{y}$ the set of all GRNs with phenotype \mathbf{y} , i.e., $\mathbf{x}_{\Omega}(g) = \mathbf{y}$ for $g \in \mathcal{G}_\mathbf{y}$, under the given external stimuli Ω_0 . We also write $\Omega_\mathbf{y}^+$ to be the required-present proteins in the phenotype \mathbf{y} , so $x_\omega = 1$ for $\omega \in \Omega_\mathbf{y}^+$ and $x_{\omega'} = 0$ for any other $\omega' \in \Omega \setminus \Omega_\mathbf{y}^+$. Note that the number of required-present proteins $|\Omega_\mathbf{y}^+|$ is bounded from above by the number of loci $|\Gamma|$ since any present protein that is not a stimulus must be triggered by the expression of some locus.

We observe that some GRNs $\tilde{\mathcal{G}}_\mathbf{y} \subset \mathcal{G}_\mathbf{y}$ play a “central” role among GRNs with the same phenotype \mathbf{y} . Specifically, for any $\tilde{g} \in \tilde{\mathcal{G}}_\mathbf{y}$, all the edges in \tilde{g} point from the stimuli Ω_0 to the required-present proteins $\Omega_\mathbf{y}^+$, and each $\omega \in \Omega_\mathbf{y}^+$ is targeted by at least one edge in \tilde{g} . We demonstrate an example of such \tilde{g} in Figure 2a. These $\tilde{\mathcal{G}}_\mathbf{y}$ are deemed central because they can be reached by any GRN $g \in \mathcal{G}_\mathbf{y}$ through mutations among $\mathcal{G}_\mathbf{y}$ themselves: First, for every edge in g that points to an $\omega \in \Omega_\mathbf{y}^+$, we rewire the edge such that it still points to ω but now from an $\omega_0 \in \Omega_0$. Arbitrarily rewiring the remaining edges between Ω_0 and $\Omega_\mathbf{y}^+$ then leads to some central GRN in $\tilde{\mathcal{G}}_\mathbf{y}$ (see Figure 2a).

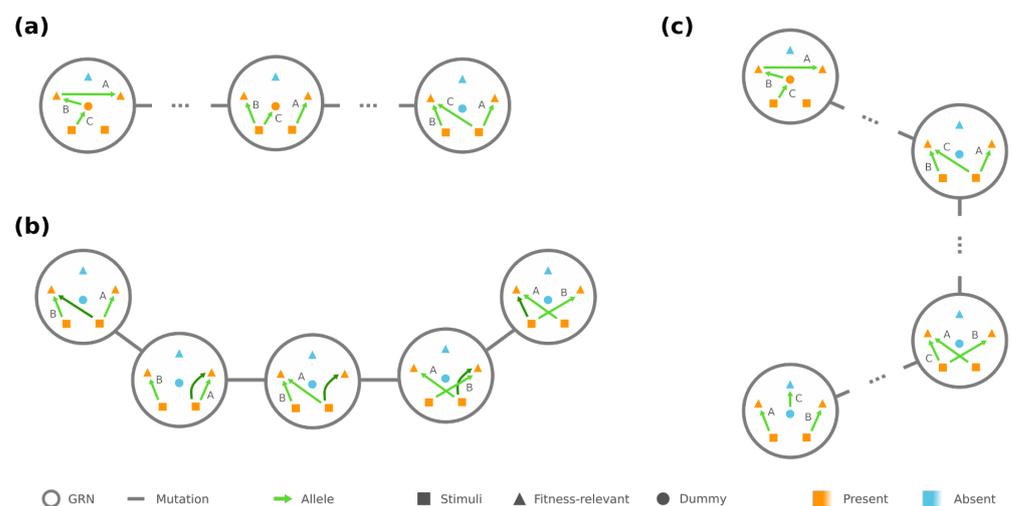


Figure 2. Connectivity exists between all GRNs of the same phenotype. (a) Any GRN can be rewired/mutated into a “central” GRN (shown on the right). (b) A redundant edge (dark green) makes it feasible to turn any central GRN into another via edge rewiring. (c) There is a mutational trajectory between any GRNs of the same phenotype through the central GRNs.

In addition, if the phenotype \mathbf{y} has strictly less required proteins than the number of loci, the central GRNs $\tilde{\mathcal{G}}_\mathbf{y}$ are mutually reachable by edge rewiring among $\tilde{\mathcal{G}}_\mathbf{y}$. There is always a redundant edge whose rewiring makes no change to the phenotype, and it helps

us rewire each edge to any desired source/target pair between Ω_0 and Ω_y^+ (see Figure 2b), which subsequently creates a chain of mutations between any $\tilde{g}, \tilde{g}' \in \tilde{\mathcal{G}}_y$. These results implicate that, for any phenotype y with $|\Omega_y^+| < |\Gamma|$ and any $g_1, g_2 \in \mathcal{G}_y$, there is always a mutational trajectory between g_1 and g_2 that only traverses over GRNs in \mathcal{G}_y , especially through the central ones $\tilde{\mathcal{G}}_y$ (see Figure 2c). In the extreme case where $|\Omega_y^+| = |\Gamma|$, however, \mathcal{G}_y fragments into multiple connected components (detailed in Appendix A).

Next, we turn to accessibility between GRNs of different phenotypes y and y' , where without loss of generality $f(y') \geq f(y)$. We observe that, if $|\Omega_y^+ \cup \Omega_{y'}^+| \leq |\Gamma| + 1$, there are always two “peripheral” GRNs $\hat{g} \in \mathcal{G}_y$ and $\hat{g}' \in \mathcal{G}_{y'}$, which only differ by one edge rewiring. To be more specific, there are two independent chains in \hat{g} , one of which begins with a stimulus $\omega_0 \in \Omega_0$ and sequentially connects the proteins required to be present in y but not in y' , i.e., $\Omega_y^+ \setminus \Omega_{y'}^+$, while the other consecutively joins $\Omega_{y'}^+ \setminus \Omega_y^+$. The rest of the edges in \hat{g} merely point from Ω_0 to $\Omega_y^+ \cap \Omega_{y'}^+$, and each $\omega \in \Omega_y^+ \cap \Omega_{y'}^+$ is targeted by at least one edge (see example in Figure 3, left). The other GRN \hat{g}' only differ from \hat{g} by the first edge in the chain of $\Omega_y^+ \setminus \Omega_{y'}^+$, which is rewired such that it points from the stimulus ω_0 to the first node in the chain of $\Omega_{y'}^+ \setminus \Omega_y^+$ (Figure 3, right).

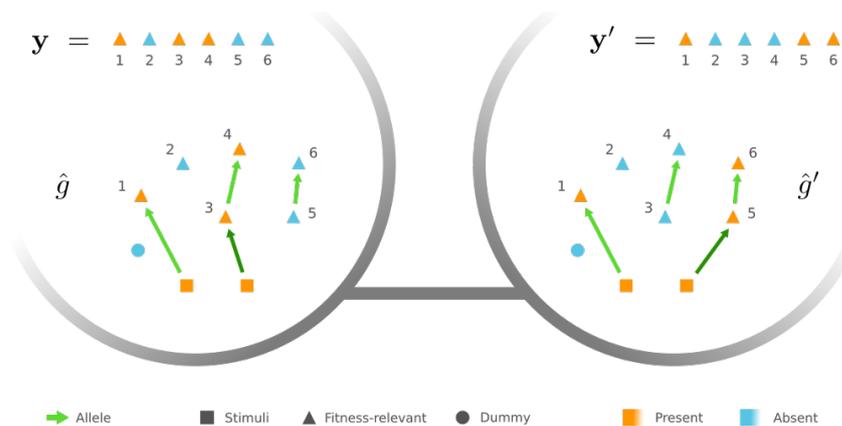


Figure 3. Example of peripheral GRNs connecting two different phenotypes—a peripheral GRN \hat{g} of phenotype y in this example. There is a chain that triggers the presence state of proteins $\Omega_y^+ \setminus \Omega_{y'}^+ = \{3, 4\}$. However, the other peripheral GRN \hat{g}' of phenotype y' contains a chain of proteins $\Omega_{y'}^+ \setminus \Omega_y^+ = \{5, 6\}$. \hat{g} and \hat{g}' are mutational neighbors since they only differ by rewiring the dark green edge, i.e., the first edge in either chain.

Our observation suggests that there is a sequence of mutations with non-decreasing fitness from any GRN $g \in \mathcal{G}_y$ to any GRN $g' \in \mathcal{G}_{y'}$, as long as $|\Omega_y^+ \cup \Omega_{y'}^+| \leq |\Gamma| + 1$. In particular, when $|\Omega_y^+| < |\Gamma|$, the mutational trajectory starting at g first traverses within \mathcal{G}_y to a peripheral GRN and then transitions into $\mathcal{G}_{y'}$ to reach g' . An analogous trajectory exists even under the extreme scenario $|\Omega_y^+| = |\Gamma|$ (see Appendix A). We also note that if the number of fitness-relevant proteins is $|\hat{\Omega}| \leq |\Gamma| + 1$, then the condition $|\Omega_y^+ \cup \Omega_{y'}^+| \leq |\Gamma| + 1$ is assuredly satisfied for any two phenotypes y and y' . As a corollary, if $|\hat{\Omega}| \leq |\Gamma| + 1$, the fitness optimum will always be accessible.

3.2. Mesoscopic Skeleton Derived from “Symmetries” in the Genotype Network of GRNs

Because the number of possible GRNs grows super-exponentially as the underlying loci and proteins expand, constructing the genotype network becomes extremely challenging beyond a small Γ and Ω . Here, we present a more compact skeleton of the fitness landscape of GRNs based on “symmetries” in the genotype network.

As the underlying space of a fitness landscape of GRNs, the genotype network G appears to contain redundant information. On the one hand, GRNs leading to the identical phenotype are deemed to have equal fitness. On the other hand, given any GRN, for

example, the mega-node rounded by orange in Figure 4a, one can always find some other GRN such that their neighborhoods in G are locally similar, e.g., the mega-node rounded by blue. This simple demonstration suggests that the structure of the genotype network G is not arbitrary; instead, some structural symmetries exist.

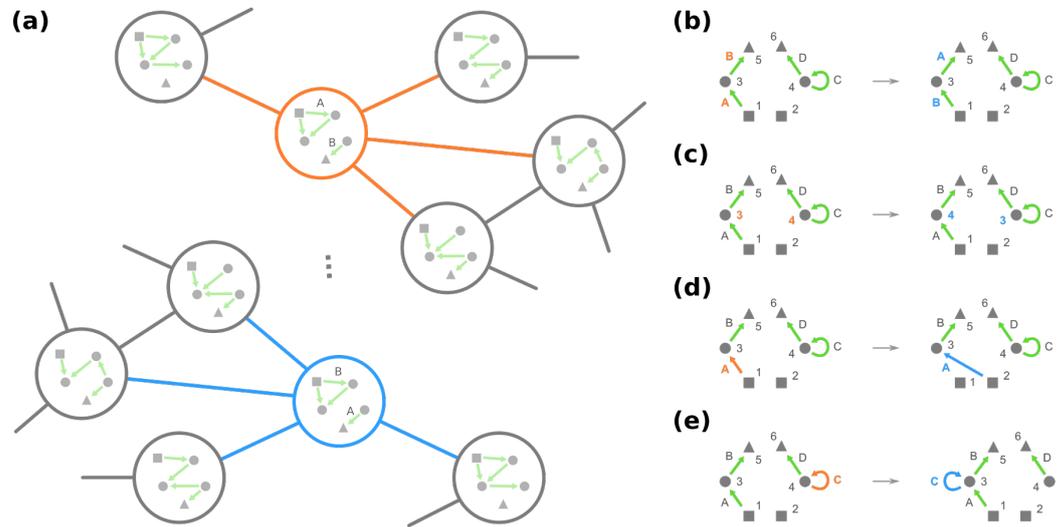


Figure 4. The genotype network has symmetry such that multiple GRNs have similar local neighborhoods, as we demonstrate in (a) since the corresponding GRNs only differ by exchanging the role of loci A and B. More formally, these GRNs constitute an equivalence class under phenotype-preserving automorphisms, which can be found by graphical operations of (b) permuting loci, (c) permuting dummy proteins (circles), (d) exchanging edges pointing from two different stimuli (squares), and (e) exchanging self-loops at two different nodes.

In graph theory, symmetries in a network are formally described through the network’s automorphisms. An automorphism of a graph is a way to shuffle the labels of its nodes such that the graph remains identical before and after shuffling. For instance, in Figure 5b, exchanging nodes 2 and 3 generates the same network and is thus an automorphism, whereas exchanging nodes 2 and 4 is not because there is an edge from 2 to 3 after shuffling. Formally, an automorphism of the genotype network G is a permutation σ of all plausible GRNs $\mathcal{G} = V(G)$ such that, for any $g_1, g_2 \in \mathcal{G}$, $(\sigma(g_1), \sigma(g_2)) \in E(G)$ if and only if we also have $(g_1, g_2) \in E(G)$. (A permutation of \mathcal{G} is a mapping $\sigma : \mathcal{G} \rightarrow \mathcal{G}$ where no two GRNs are mapped to the same GRN, i.e., $\sigma(g_1) \neq \sigma(g_2)$ if $g_1 \neq g_2$ for any $g_1, g_2 \in \mathcal{G}$.) Once two GRNs g and g' are related through an automorphism σ of G , e.g., $g' = \sigma(g)$, they share the same mega-node properties that are fully determined by the connections in the genotype network (see Proposition A1).

Furthermore, automorphisms partition the GRNs by their roles in the genotype network through the mathematical concept of equivalence classes. For a high-level and general description, imagine a set of elements and a group of operations acting on them. Each operation turns one element into another, and these two elements are related by the operation, which describes the similarity between them. An equivalence class consists of elements that are mutually related by any operation, and the set of elements is said to be partitioned into equivalence classes under the action of the operations (see Figure 5a for an illustrative example). For automorphisms $\Sigma(G')$ of a graph G' , the equivalence classes of nodes $V(G')$ under the action of $\Sigma(G')$ then gather nodes with a similar “structural position” in G' (Figure 5c).

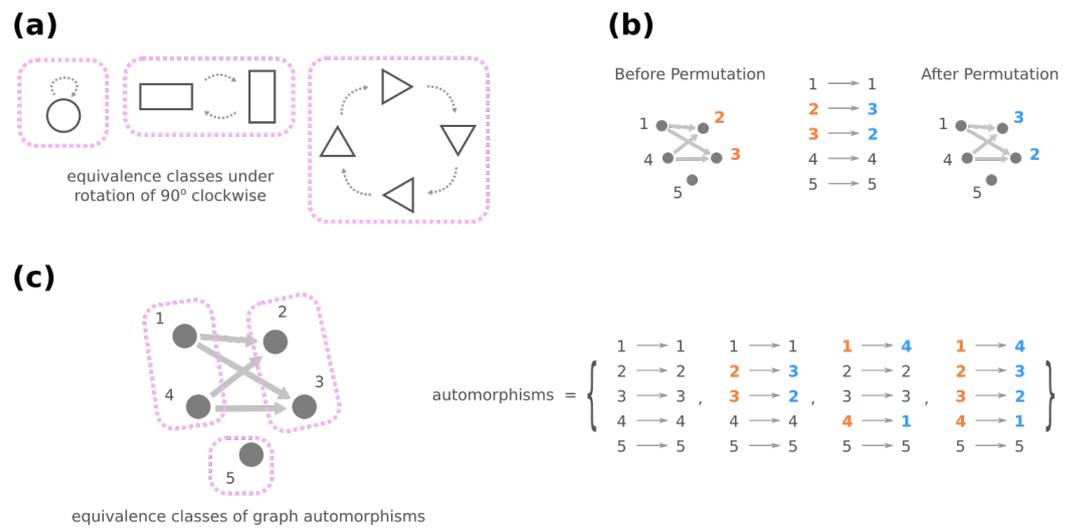


Figure 5. (a) As a minimal example, imagine an operation that rotates a geometric object 90 degrees clockwise. The rotation maps one object onto another (dashed arrows), and it leads to equivalence classes where objects are grouped by their symmetry under rotation (pink rectangles). (b) An automorphism of a graph is a permutation of nodes that retains the same graph. (c) Equivalence classes under graph automorphisms bring together nodes that have similar roles connection-wise in the graph.

However, to reveal GRNs with identical roles in a fitness landscape, these automorphisms also need to preserve the phenotype. Denote by $\Sigma_x(G)$ the set of such automorphisms of G , i.e., for any $\sigma \in \Sigma_x(G)$, and GRN $g \in \mathcal{G}$, $\sigma(g)$ and g have the same phenotype. The equivalence classes of mega-nodes $V(G)$ under the action of phenotype-preserving automorphisms $\Sigma_x(G)$ then unite GRNs that (a) show similar mutational relationships with others and (b) lead to the same fitness due to their identical phenotype. We will mildly abuse the terminology to call them the *equivalence classes of GRNs*, which we denote by Θ , and each $\theta \in \Theta$ is a set of GRNs related through $\Sigma_x(G)$. Crucially, since the mutational relationship and the resultant phenotype are the two components that characterize a GRN in the fitness landscape, GRNs in a $\theta \in \Theta$ are deemed equivalent semantically, and they can be reduced to an arbitrary representative among them. Therefore, the equivalence classes of GRNs provide an efficient way to depict the underlying space of the fitness landscape.

However, what exactly composes the phenotype-preserving automorphisms $\Sigma_x(G)$ of the genotype network? From a sufficiency direction, we show that there exist a few graphical operations on the GRNs that produce phenotype-preserving automorphisms. These graphical operations involve permuting/shuffling different sorts of elements in a GRN:

- (i) The identities of loci Γ , e.g., exchanging edge labels of loci A and B in Figure 4b;
- (ii) The identities of dummy proteins Ω' , e.g., exchanging node labels of proteins 3 and 4 in Figure 4c.

Then, potentially rewiring a given edge (see details in Definitions A1 and A2):

- (iii) Change the source node of an edge from one stimulus to another stimulus and vice versa, e.g., in Figure 4d, moving an edge pointing from node 1 to node 3 to pointing from node 2. (Note that this operation is not necessarily equivalent to permuting the identities of stimuli since at most only the single focal edge will be affected.)
- (iv) Move a self-loop at one node to another node and vice versa, for example, re-allocating a self-loop at node 3 to node 4 in Figure 4e.

For the formal proofs, we point the reader to Theorem A1. Additionally, from a necessity direction, one can computationally obtain a partition $\hat{\Theta}$ of the GRNs \mathcal{G} that is coarser than the equivalence classes Θ . (A partition \mathcal{P} is coarser than another partition \mathcal{P}' if any group in \mathcal{P}' is included in some group in \mathcal{P} .) Specifically, start with a partition $\hat{\Theta}_0$ where GRNs with the same resultant phenotype are grouped together. We create a

sequence of partitions of \mathcal{G} through the following iterative procedure: Given the partition $\hat{\Theta}_i$, the next partition $\hat{\Theta}_{i+1}$ is obtained by further dividing groups into $\hat{\Theta}_i$ (if needed) such that for each group $\theta \in \hat{\Theta}_i$ and $\theta' \in \hat{\Theta}_{i+1}$, any two GRNs in θ' have the same number of neighbors among θ . This iterative procedure is terminated when no further division is required, i.e., $\hat{\Theta}_{k+1} = \hat{\Theta}_k$ for some integer k (see Figure 6a for an illustrative cartoon of the iterative procedure). We then have $\hat{\Theta} = \hat{\Theta}_k$ to be our desired partition of GRNs.

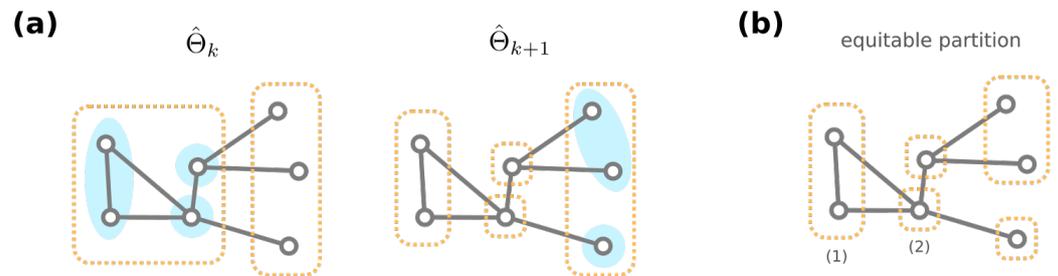


Figure 6. (a) Consider a toy example genotype network of GRNs. (Here, we omit the exact content of GRNs.) Given the partition $\hat{\Theta}_k$, note that mega-nodes in a group (dashed orange rectangle) may share a different number of connections among other groups (blue shaded circles), and they are further divided to generate the next partition $\hat{\Theta}_{k+1}$. (b) Both the equivalence classes of GRNs and the stationary partition from our iterative procedure are equitable, e.g., each mega-node in group (1) has one connection among (1), another connection with (2), and none with other groups.

To see why the proposed iterative procedure generates a coarser partition $\hat{\Theta}$ than the equivalence classes Θ of GRNs, we stress that the equivalence classes under automorphisms always form an equitable partition. A partition $\mathcal{P} = \{P_i\}_{i=1}^m$ of nodes of a graph is equitable [81] if for every $P_i, P_j \in \mathcal{P}$, any two nodes u, v in group P_i have the same number of neighbors in P_j (Figure 6b). Since GRNs in an equivalence class $\theta \in \Theta$ must have the same amount of neighbors for each different phenotype, we inductively show that any two GRNs $g_1, g_2 \in \theta$ are never separated during the iterative procedure that generates $\hat{\Theta}$ (see Theorem A2). Therefore, any equivalence class $\theta \in \Theta$ must be included in a computationally acquired group $\hat{\theta} \in \hat{\Theta}$.

Figure 7 demonstrates the coarser partition $\hat{\Theta}$ generated by the iterative procedure for an arbitrary toy example. The obtained $\hat{\Theta}$ contains 154 groups of GRNs, and the size of groups ranges from 2 to 96. We also count the number of different kinds of GRNs that can not be transformed through graphical operations (i) and (ii), and this number varies from 1 to 4 in our example $\hat{\Theta}$. Moreover, for every group in $\hat{\Theta}$, we observe that those different kinds of GRNs can be related by changing the stimulus that an edge is pointing from and re-allocating self-loops (e.g., see Figure 7b). $\hat{\Theta}$ is thus not simply a coarser partition than the equivalence classes; according to (i)–(iv), we know that groups in $\hat{\Theta}$ are exactly the equivalence classes Θ . This arguably general toy example implicates that there is no need for other graphical operations to determine the equivalence classes of GRNs.

As a result, we conjecture that all the phenotype-preserving automorphisms $\Sigma_x(G)$ of the genotype network can be generated by combining graphical operations (i) to (iv) on the GRNs. In other words, two GRNs g_1 and g_2 belong to the same equivalence class if and only if, after removing all the self-loops and merging stimuli Ω_0 into a single node, there exist permutations of loci Γ and dummy proteins Ω' that jointly transform g_1 into g_2 . This condition reconciles with the concept of isomorphisms between graphs. Whereas an automorphism is a mapping of nodes such that a graph preserves itself, an isomorphism is a mapping of nodes that transform one graph into another. We will borrow the terminology and call the two permutations of Γ and Ω' together a *phenotype-preserving isomorphism* from g_1 to g_2 .

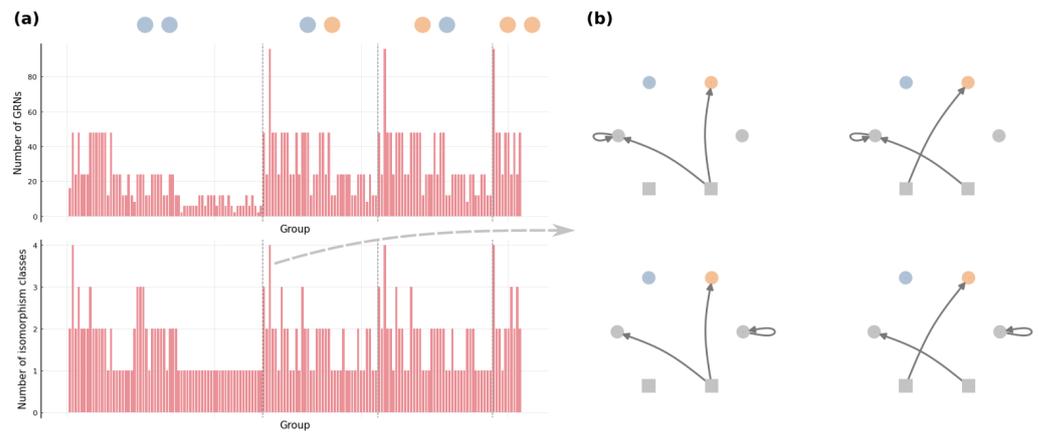


Figure 7. Example partition coarser than the equivalence classes of GRNs. We run the proposed iterative procedure with $|\Gamma| = 3$ and $|\Omega_0| = |\Omega'| = |\hat{\Omega}| = 2$, where stimuli Ω_0 are drawn as squares and the present/absent state of fitness-relevant proteins $\hat{\Omega}$ are colored by orange/blue. (a) The number of GRNs and the number of isomorphism classes of GRNs in each group of the obtained partition $\hat{\Theta}$, where the dashed lines separate groups of different phenotypes and (b) isomorphism classes of GRNs in a group.

3.3. Algorithmic Construction of the Mesoscopic Backbone of GRN Fitness Landscape

Next, we investigate algorithmic approaches to construct the mesoscopic backbone of a fitness landscape based on equivalence classes, where a representative GRN replaces all other GRNs in an equivalence class due to their identical role. In particular, the desired algorithm must (a) acquire the equivalence classes Θ from scratch and (b), for a representative GRN in any equivalence class, count the number of its mutational neighbors in other equivalence classes and also within the class it belongs to.

To avoid any confusion, we emphasize that, although drawing mutational connections between equivalence classes Θ can be achieved by grouping mega-nodes in the genotype network G , this naive exercise is unsuitable. First and foremost, grouping mega-nodes demands prior knowledge of the genotype network itself, but its construction is computationally heavy. Second, in contrast to coarse-graining nodes in a graph where the groups of nodes are pre-specified, listing all GRNs in an equivalence class requires examining pairs of GRNs and assuring a phenotype-preserving isomorphism between them after removing self-loops and merging stimuli. Determining the equivalence classes Θ from all the GRNs $\mathcal{G} = V(G)$ can thus be costly as well. These reasons again show the value of the equivalence classes Θ , which consolidate GRNs into their equivalent representatives.

Here, we present a bottom-up approach that enumerates each equivalence class of GRNs and simultaneously computes the number of mutational connections among them. To begin, recall from Section 2.2 that a mutation from a GRN $g_1 \in \mathcal{G}$ to another $g_2 \in \mathcal{G}$ corresponds to rewiring a single edge in g_1 , where g_1 may rewire a self-loop/non-self-loop edge to a self-loop/non-self-loop edge in g_2 . We observe that the number of non-self-loop edges in mutational neighbors g_1 and g_2 differ at most by one. We denote by $\Gamma'(g)$ the loci representing the non-self-loop edges in the GRN g , and $|\Gamma'(g)|$ the number of those non-self-loop edges. In other words, given equivalence classes $\theta, \theta' \in \Theta$ and representative GRNs $g \in \theta$ and $g' \in \theta'$, g has no mutational neighbors in θ' if $||\Gamma'(g)| - |\Gamma'(g')|| > 1$.

We can therefore build the mesoscopic backbone by incrementally examining each equivalence class with an increasing number of non-self-loop edges in the representative GRN. This strategy is envisioned in Figure 8, where the backbone can be viewed as “layers” of equivalence classes of GRNs. Let Θ_k be the set of equivalence classes where for every $\theta \in \Theta_k$, the representative GRN $g \in \theta$ has exactly k non-self-loop edges, $|\Gamma'(g)| = k$. We start with layer Θ_0 , which consists of the only equivalence class with no non-self-loop edges. Then, with layers $\Theta_0, \Theta_1, \dots, \Theta_k$ and all the mutational connections among them, we will

find the equivalence classes in the next layer Θ_{k+1} and their mutational connections with layer Θ_k and within themselves up until $k = |\Gamma|$, where all the edges are non-self-loops.

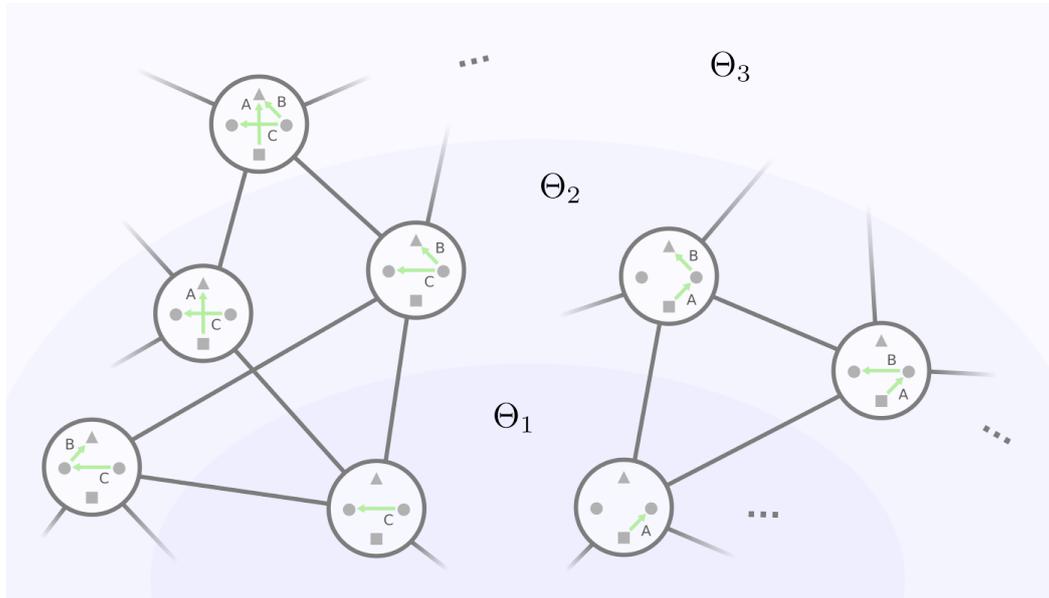


Figure 8. Layering the GRNs by their number of non-self-loop edges. A GRN’s mutational neighbors must fall into the same or the adjacent layers. For ease of illustration, we only show the non-self-loop edges and neglect the protein states in GRNs.

To be more precise, we introduce the concept of \mathcal{M}^+ neighborhood: For any GRN $g \in \mathcal{G}$, denote by $\mathcal{M}^+(g)$ the mutational neighbors of g that have one more non-self-loop edge than g . \mathcal{M}^+ neighborhoods are sufficient to capture the relationship between two mutational neighbors g and g' :

- If g' has one more non-self-loop edge than g , then $g' \in \mathcal{M}^+(g)$;
- If g' has one less non-self-loop edge than g , then we have $g \in \mathcal{M}^+(g')$;
- If g' has the same number of non-self-loop edges as g , and then they share a common mutational neighbor g'' , where the only different edge between g and g' is rewired to a self-loop and thus $g, g' \in \mathcal{M}^+(g'')$.

The mutational connections between equivalence classes can hence be uncovered by examining the \mathcal{M}^+ neighborhood of the representative GRNs. Moreover, the \mathcal{M}^+ neighborhood of representative GRNs in layer Θ_k reveals the equivalence classes in layer Θ_{k+1} because any GRN must have a mutational neighbor with one less non-self-loop edge. All that remains is to join different \mathcal{M}^+ neighbors into equivalence classes. In particular:

- (A) For an equivalence class $\theta \in \Theta_k$ and its representative GRN $g \in \theta$, under what condition will $g'_1, g'_2 \in \mathcal{M}^+(g)$ belong to the same equivalence class in layer Θ_{k+1} ?
- (B) For two distinct equivalence classes $\theta_1, \theta_2 \in \Theta_k$ and their representative GRNs $g_1 \in \theta_1$ and $g_2 \in \theta_2$, under what condition will $g'_1 \in \mathcal{M}^+(g_1)$ and $g'_2 \in \mathcal{M}^+(g_2)$ belong to the same equivalence class in layer Θ_{k+1} ?

For our ease of illustration, we hereafter choose the GRNs g, g_1, g_2, g'_1 and g'_2 such that only one stimulus node is incident to out-going edges.

To address (A), let $g'_1, g'_2 \in \mathcal{M}^+(g)$ belong to the same equivalence class, so there is a phenotype-preserving isomorphism π from g'_1 to g'_2 after self-loop removal. Recalling from Section 2.2, $e_g(\gamma) = (u, v)$ denotes that “the source–target pair of edge γ is (u, v) in GRN g .” Furthermore, we write $e_{g'_1}(\gamma_1) = (u_1, v_1)$ and $e_{g'_2}(\gamma_2) = (u_2, v_2)$, where γ_1 and γ_2 are the non-self-loop edges “added” to g that forms g'_1 and g'_2 , respectively. A few observations follow:

1. There is an integer p such that $\pi^p(\gamma_1) = \gamma_2$ and $(\pi^p(u_1), \pi^p(v_1)) = (u_2, v_2)$;

2. There is another integer $q < p$ such that $\pi^q(\gamma_1) = \gamma_2$ and $(\pi^q(u_1), \pi^q(v_1)) = (u_2, v_2)$;
3. $e_{g'_2}(\pi^k(\gamma_1)) = (\pi^k(u_1), \pi^k(v_1))$ for $k = 1, 2, \dots, q$;
4. $e_{g'_2}(\pi^k(\gamma_1)) \neq (\pi^k(u_1), \pi^k(v_1))$ for $k = q + 1, q + 2, \dots, p$;
5. For any locus γ and non-self-loop source–target pair (u, v) such that $(\gamma, u, v) \neq (\pi^k(\gamma), \pi^k(u_1), \pi^k(v_1))$ for $0 \leq k \leq q - 1$, we have $e_g(\pi(\gamma)) = (\pi(u), \pi(v))$ if and only if $e_{\bar{g}}(\gamma) = (u, v)$.

We detail the reasoning behind these observations in Lemma A1–A3. Critically, our fifth observation implies that, after self-loop removal, the isomorphism π between g'_1 and g'_2 is in fact a phenotype-preserving automorphism of a subgraph \bar{g} of the GRN g . In addition, observations 3. and 4. show that those edges in g —but not in \bar{g} —are sequentially mapped from one to another via this automorphism π , i.e., $\Gamma'(g) \setminus \Gamma'(\bar{g}) = \left\{ \pi^k(\gamma_1) \right\}_{k=1}^{q-1}$, and they bridge the newly added edges γ_1 and $\gamma_2 = \pi^q(\gamma_1)$. We show that the converse is also true (see Theorem A3): After self-loop removal, if we find a phenotype-preserving automorphism π of a subgraph \bar{g} of g where γ_1 is consecutively mapped to γ_2 through the edge differences $\Gamma'(g) \setminus \Gamma'(\bar{g})$, π is guaranteed a phenotype-preserving isomorphism from g'_1 to g'_2 .

The sufficient and necessary condition for two \mathcal{M}^+ neighbors of g to be in the same equivalence class, intriguingly, lies in the phenotype-preserving automorphisms of subgraphs of the representative GRN g . Here, we demonstrate a few simple examples in Figure 9a. In the top row, an automorphism of g directly maps between the two additional edges $(u_1, v_1) = (3, 5)$ and $(u_2, v_2) = (1, 4)$. In the middle row, the two edges $(u_1, v_1) = (1, 2)$ are consecutively mapped to $(u_2, v_2) = (3, 4)$ through edge $(2, 3)$, and (u_2, v_2) is consecutively mapped back to (u_1, v_1) through the non-edge $(4, 1)$, so we have $q = 2$ and $p = 4$. As a mixture of both, in the bottom row, $(u_1, v_1) = (2, 5)$ is consecutively mapped to $(u_2, v_2) = (3, 6)$ through edge $(1, 4)$, and this isomorphism is exactly an automorphism of a subgraph \bar{g} of g where edge $(1, 4)$ is removed.

Switching gears to the remaining question (B), suppose that g_1 and g_2 are the representative GRN in two different equivalence classes where $|\Gamma'(g_1)| = |\Gamma'(g_2)|$ and that $g'_1 \in \mathcal{M}^+(g_1)$ and $g'_2 \in \mathcal{M}^+(g_2)$ belong to the same equivalence class. Let γ_1 and γ_2 be the newly added edges to g_1 and g_2 that generate g'_1 and g'_2 , respectively, where $e_{g'_1}(\gamma_1) = (u_1, v_1)$ and $e_{g'_2}(\gamma_2) = (u_2, v_2)$, and let π be a phenotype-preserving isomorphism from g'_1 to g'_2 after self-loop removal. We observe that applying the permutation π on g_1 transforms it into another GRN \tilde{g}_1 in the same equivalence class. Since g_1 simply has one less edge γ_1 than g'_1 , and \tilde{g}_1 and g'_2 only differ by a missing edge $\pi(\gamma_1)$. Namely, we have $g'_2 \in \mathcal{M}^+(\tilde{g}_1)$ with the additional edge $e_{g'_2}(\pi(\gamma_1)) = (\pi(u_1), \pi(v_1))$. Moreover, since g'_2 also belongs to the \mathcal{M}^+ neighborhood of g_2 with the additional edge $e_{g'_2}(\gamma_2) = (u_2, v_2)$, by removing both the extra edges from g'_2 , we find a GRN g'' such that $\tilde{g}_1, g_2 \in \mathcal{M}^+(g'')$.

We again present an illustrative example in Figure 9b. Here, a GRN \tilde{g}_1 in the equivalence class of g_1 can be found via the isomorphism π between g'_1 and g'_2 . We note that the newly added edge $(u_1, v_1) = (4, 1)$ is transformed into $(\pi(u_1), \pi(v_1)) = (3, 4)$ in g'_2 , which is missing in \tilde{g}_1 . Removing both $(\pi(u_1), \pi(v_1)) = (3, 4)$ and $(u_2, v_2) = (3, 1)$ from g'_2 produces a GRN g'' , which is a common neighbor of g_2 and \tilde{g}_1 with one less non-self-loop edge.

Our observation resolves the necessary condition of (B): For the representative GRNs of two different equivalence classes $g_1 \in \theta_1$ and $g_2 \in \theta_2$, if their \mathcal{M}^+ neighbor $g'_1 \in \mathcal{M}^+(g_1)$ and $g'_2 \in \mathcal{M}^+(g_2)$ belong to the same equivalence class, then we can always find two GRNs \tilde{g}_1 and g'' such that (a) \tilde{g}_1 falls into the equivalence class of g_1 , and (b) \tilde{g}_1 and g_2 are \mathcal{M}^+ neighbors of g'' . Moreover, the converse is true as well (Theorem A4). Therefore, whether the \mathcal{M}^+ neighborhood of g_1 and g_2 reveal a common equivalence class depends on the existence of a GRN g'' that both the equivalence classes θ_1 and θ_2 are rooted from.

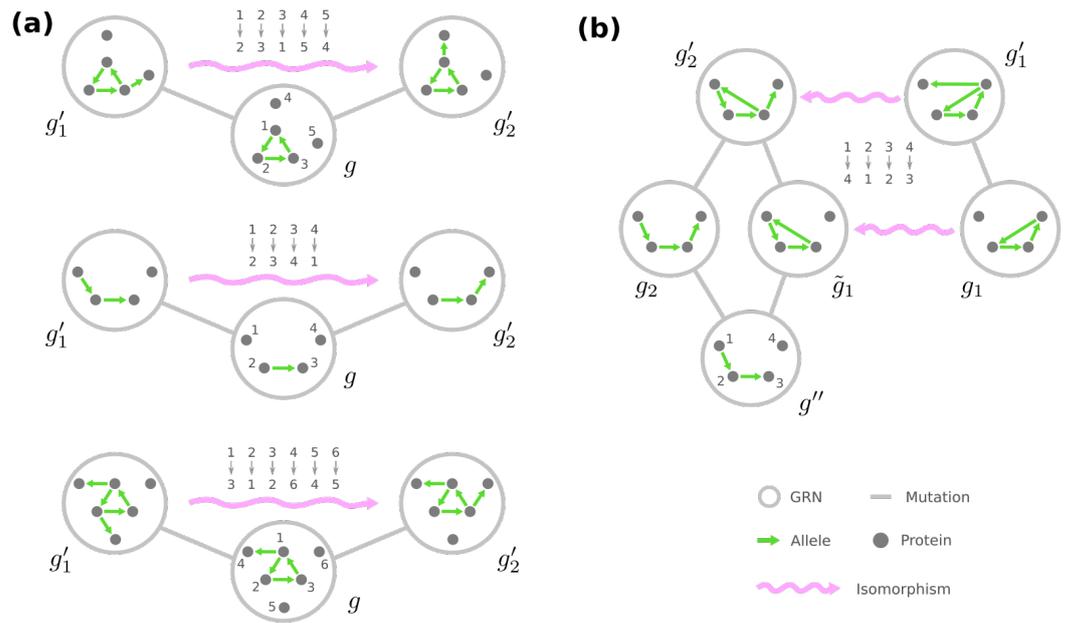


Figure 9. Sufficient conditions that two \mathcal{M}^+ neighbors belong to an equivalence class. For illustration purposes, we only show the dummy proteins and omit the protein states, edge labels and self-loops in (a) three examples such that two \mathcal{M}^+ neighbors of a GRN g are isomorphic, and (b) an example where the \mathcal{M}^+ neighbors g'_1 and g'_2 of GRNs g_1 and g_2 in different equivalence classes are isomorphic.

Our strategy to build the mesoscopic backbone is now complete, and here, we detail our algorithm that incrementally generates the equivalence classes Θ of GRNs and establishes the mutational connections among them. Suppose that we have already built layers of equivalence classes $\Theta_0, \Theta_1, \dots, \Theta_k$ and determined the mutational connections among them. For each representative GRN g in layer Θ_k and every $g' \in \mathcal{M}^+(g)$, we will view g' as the combination of g and an additional, non-self-loop edge $e_{g'}(\gamma) = (u, v)$, for which we write $g' = g \oplus (\gamma, u, v)$. All such combinations form a collection of \mathcal{M}^+ neighbors of the representative GRNs in layer Θ_k , for which we abuse the notation $\mathcal{M}^+(\Theta_k)$.

We initially put each $g' \in \mathcal{M}^+(\Theta_k)$ into an individual group, and we define a collection of operations Φ that join groups of \mathcal{M}^+ neighbors:

- (I) For every representative GRN g in Θ_k and every phenotype-preserving automorphism σ of g , there is an operation $\psi_{g,\sigma}$ that joins together the groups of $g'_1 = g \oplus (\gamma, u_1, v_1)$ and $g'_2 = g \oplus (\gamma, u_2, v_2)$, where $u_1, u_2 \in \Omega_0$ and $v_2 = \sigma(v_1)$;
- (II) For every representative GRN g in Θ_k and every phenotype-preserving automorphism $\bar{\sigma}$ of each subgraph \bar{g} of g such that the edge differences $\Gamma'(g) \setminus \Gamma'(\bar{g})$ are sequentially connected via $\bar{\sigma}$, there is an operation $\phi_{g,\bar{g},\bar{\sigma}}$ that joins together the groups of $g'_1 = g \oplus (\gamma_1, u_1, v_1)$ and $g'_2 = g \oplus (\gamma_2, u_2, v_2)$, where automorphism $\bar{\sigma}$ consecutively transforms edge γ_1 into γ_2 through $\Gamma'(g) \setminus \Gamma'(\bar{g})$;
- (III) For every representative GRN g'' in Θ_{k-1} and each $\tilde{g}_1 = g'' \oplus (\gamma''_1, u''_1, v''_1)$ and $\tilde{g}_2 = g'' \oplus (\gamma''_2, u''_2, v''_2)$ in two different equivalence classes θ_1 and θ_2 , such that we have phenotype-preserving isomorphisms π_1/π_2 from \tilde{g}_1/\tilde{g}_2 to the representative GRN g_1/g_2 after self-loop removal, there is an operation $\varphi_{g,\tilde{g}_1,\tilde{g}_2}$ that joins together the groups of $g'_1 = g_1 \oplus (\pi_2(\gamma''_2), \pi_2(u''_2), \pi_2(v''_2))$, and $g'_2 = g_2 \oplus (\pi_1(\gamma''_1), \pi_1(u''_1), \pi_1(v''_1))$.

The resulting groups of \mathcal{M}^+ neighbors, after applying the joining operations Φ , constitute the equivalence classes in the next layer Θ_{k+1} . We hereafter denote by $\mathcal{M}^+_{\Phi}(\theta')$ the corresponding consequent group of an equivalence class $\theta' \in \Theta_{k+1}$. We then choose an arbitrary \mathcal{M}^+ neighbor in $\mathcal{M}^+_{\Phi}(\theta')$ as the representative GRN of the equivalence class θ' , such that only one stimulus node is incident to out-going edges in the chosen representative GRN.

The joining operations Φ further provide useful information to count the number of mutation neighbors that a representative GRN $g \in \theta$ in layer Θ_k has among any equivalence class θ' , which we will denote by $A_g(\theta')$. Let us first consider $\theta' \in \Theta_{k+1}$. For any $\tilde{g}' \in \mathcal{M}_\Phi^+(\theta')$, \tilde{g}' is a mutational neighbor of g if it can be viewed as a combination of g and an arbitrary extra non-self-loop edge, and hence

$$A_g(\theta') = |\mathcal{M}_\Phi^+(\theta') \cap \mathcal{M}^+(g)|, \quad \text{for } \theta' \in \Theta_{k+1}. \tag{1}$$

Note that, in this case, $A_g(\theta')$ is easily acquired when building up the layer Θ_{k+1} through Φ .

Second, for $\theta' \in \Theta_{k-1}$, $A_g(\theta')$ can be computed given $A_{g'}(\theta)$, where g' is the representative GRN of θ' . Since the equivalence classes Θ generate an equitable partition of the genotype network G (see Section 3.2), we have $A_g(\theta') \times |\theta| = A_{g'}(\theta) \times |\theta'|$ equal to the total number of mutational connections between θ and θ' . Moreover, the size of the equivalence class θ is (see Appendix D)

$$|\theta| = \frac{|\Pi'|}{|\Sigma'(g)|} \times n_l(|\Gamma| - k) \times m_s(g) \times r(g), \tag{2}$$

where (a) we denote by Π' the set of all permutations of dummy proteins Ω' and denote by $\Sigma'(g)$ the set of automorphisms of the representative GRN g after self-loop removal that only permutes Ω' ; (b) $n_l(|\Gamma| - k)$ is the number of ways to allocate $|\Gamma| - k$ labeled self-loops among the proteins Ω ; (c) $m_s(g)$ is the number of ways to re-distribute the edges pointing from stimuli Ω_0 in g ; and (d) $r(g)$ is the number of ways to divide loci Γ into self-loops, non-self-loop edges pointing from stimuli, and others. As a result,

$$A_g(\theta') = A_{g'}(\theta) \times \frac{|\Sigma'(g)|}{|\Sigma'(g')|} \times \frac{n_l(|\Gamma| - k + 1)m_s(g')r(g')}{n_l(|\Gamma| - k)m_s(g)r(g)}, \quad \text{for } \theta' \in \Theta_{k-1}. \tag{3}$$

Third, we turn to the case where $\theta' \in \Theta_k$ but $\theta' \neq \theta$. Recall that, if any $\tilde{g}' \in \theta'$ is a mutational neighbor of g , then there is a GRN \tilde{g}'' in layer Θ_{k-1} , where $g, \tilde{g}' \in \mathcal{M}^+(\tilde{g}'')$, and such \tilde{g}'' is unique up to arbitrary self-loop re-allocation. Additionally, the extra edge in g and \tilde{g}' must correspond to the same locus, so

$$A_g(\theta') = \sum_{\theta'' \in \Theta_{k-1}} \frac{A_g(\theta'')}{n_l(1)} \times A_{g''}(\theta'), \quad \text{for } \theta' \in \Theta_{k-1}, \theta' \neq \theta, \tag{4}$$

in which we use g'' to be the representative GRN of equivalence class θ'' . Lastly, if $\theta' = \theta$, we also need to include the scenario that the mutational neighbor \tilde{g}' of g is generated by rewiring a self-loop to another self-loop. Therefore,

$$A_g(\theta) = (|\Gamma| - k) \times (n_l(1) - 1) + \sum_{\theta'' \in \Theta_{k-1}} \frac{A_g(\theta'')}{n_l(1)} \times [A_{g''}(\theta) - 1]. \tag{5}$$

In Algorithm 1, we summarize our proposed approach that constructs the mesoscopic backbone. It is apparent that the core of our algorithm is determining the joining operations Φ for a given layer Θ_k . This task can be achieved by pre-computing the phenotype-preserving automorphisms of every representative GRN once it is chosen. In addition, since these joining operations reflect the mutational neighbors and the phenotype-preserving isomorphisms in previous layers, the type-(III) Φ for layer Θ_k is generated as a composition of the already uncovered operations. Furthermore, the remaining Φ of type (II) consists of combinations of the uncovered joining operations and the newly computed automorphisms of representative GRNs in layer Θ_k . As a result, the only prerequisite in our proposed algorithm is producing the phenotype-preserving automorphisms of a GRN.

Algorithm 1 Constructing the underlying space of a fitness landscape of GRNs**Require:** The fixed underlying collections of loci Γ and proteins Ω of GRNs**Ensure:** The representative GRN g_θ of each equivalence class $\theta \in \Theta$, and its number of mutational neighbors $A_{g_\theta}(\theta')$ in any equivalence class $\theta' \in \Theta$

```

1:  $k \leftarrow 0$  ▷ initialization
2:  $g_{\theta_0} \leftarrow$  a GRN with no self-loop, where  $\theta_0$  is the only equivalence class in layer  $\Theta_0$ 
3: Store the phenotype-preserving automorphisms  $\Sigma_x(g_{\theta_0})$ .
4: Compute  $A_{g_{\theta_0}}(\theta_0)$  via Equation (5).
5: while  $k < |\Gamma|$  do ▷ incrementally find  $\Theta$ 
6:   Construct and store the joining operations  $\Phi$  for layer  $\Theta_k$ .
7:    $\mathcal{M}_\Phi^+ \leftarrow$  grouping of  $\mathcal{M}^+(\Theta_k)$  acted by  $\Phi$ 
8:    $\Theta_{k+1}$  corresponds to the groups in  $\mathcal{M}_\Phi^+$ .
9:   for all  $\theta' \in \Theta_{k+1}$  do
10:     $g_{\theta'} \leftarrow$  a GRN in  $\mathcal{M}_\Phi^+(\theta')$  ▷ choose the representative GRN
11:    Store the phenotype-preserving automorphisms  $\Sigma_x(g_{\theta'})$ .
12:   end for
13:   for all  $\theta \in \Theta_k, \theta' \in \Theta_{k+1}$  do ▷ count the number of mutational neighbors
14:     Compute  $A_{g_\theta}(\theta')$  and  $A_{g_{\theta'}}(\theta)$  via Equations (1) and (3).
15:   end for
16:   for all  $\theta'_1, \theta'_2 \in \Theta_{k+1}$  do
17:     Compute  $A_{g_{\theta'_1}}(\theta'_2)$  via Equations (4) and (5).
18:   end for
19:    $k \leftarrow k + 1$ 
20: end while
21: Set any remaining, not computed  $A_{g_\theta}(\theta')$  to zero.

```

4. Conclusions

In this work, we integrate mechanistic knowledge of how phenotypes are computed from genotypes via regulatory interactions into fitness landscape models. The resulting family of fitness landscape models features flexibility for tunable ruggedness and accessibility among phenotypes. Furthermore, we introduce the concept of equivalence classes of GRNs, where GRNs of the same phenotype and with similar structural positions in the genotype network are coarse-grained into a group. These equivalence classes of GRNs lead to a compact and informative description of the fundamental space of a fitness landscape. Using this coarse-graining, we develop a bottom-up, efficient algorithm for constructing the underlying space of a fitness landscape based on the equivalence classes. Critically, this algorithm does not require pre-computing the genotype network and therefore permits the exploration of substantially larger GRNs.

Naively, ruggedness and accessibility would seem to be contradictory characteristics of a fitness landscape. Indeed, reciprocal sign epistasis has been shown to yield a strong influence on a landscape's ruggedness and was regarded as an impediment to evolutionary accessibility when first introduced [2,32,55]. Nevertheless, recent studies suggest that fitness landscape models most closely aligned with empirical observations show that sign epistasis (and thus ruggedness) can co-exist with accessibility [63,82]. In addition to demonstrating that ruggedness and accessibility are not mutually exclusive, our model is compatible with three additional empirical observations. First, GRNs result in high dimensional genotype–phenotype maps [63]. Second, selection acts on the superposition of mutations and the background GRN rather than a few pairs of mutations [60]. Third, and perhaps most importantly, a GRN may experience a series of neutral mutations and then evolve into a nearby phenotype [3,8,83,84]. The accessibility induced in fitness landscapes of GRNs via neutral evolution agrees with the phenomenon of punctuated equilibrium/epochal evolution [23,85,86].

Our derived equivalence classes for GRNs provide a novel, mesoscopic, and optimally descriptive skeleton of a fitness landscape. Neither the genotypic space nor the phenotypic

space alone fully characterize a fitness landscape; however, models with even a relatively simple genotype–phenotype map are computationally intensive because they must retain all plausible genotypes [70,73–75]. Intuitively, the complexity of a genotype–phenotype map can be reduced by combining similar phenotypes into high-level descriptors [87]. The equivalence classes of GRNs, on the other hand, serve as an intermediate level between the genotypic and phenotypic space, which provides an optimal coarse-graining that encodes all necessary information to predict the evolutionary trajectory on the fitness landscape.

We argue that our proposed algorithm for coarse-graining GRN fitness landscapes is more efficient than brute-force approaches. First, because we consolidate an equivalence class into a single representative GRN, our method is less costly in memory and requires fewer computations when finding mutational neighbors. Second, suppose all plausible GRNs were organized into layers by the number of non-self-loop edges (see Section 3.3), every layer would still super-exponentially contain many GRNs. Our algorithm instead finds the equivalence classes in each layer iteratively. To construct the $(k + 1)$ -th layer, we only have to exhaust the representative GRNs in the k -th layer—along with any plausible additional non-self-loop edge(s)—this amount will be significantly fewer than the number of GRNs in the $(k + 1)$ -th layer. Lastly, existing heuristics for graph automorphisms [88,89] can be used to produce the phenotype-preserving automorphisms of the representative GRNs, which is the only prerequisite when joining together different GRN–edge pairs. Because the set of automorphisms becomes more limited as the complexity of GRNs increases, we expect only a minor overhead in the joining procedure as compared to the exhaustive, brute-force approach.

Despite our model being constrained to the pathway framework of GRNs [77,78] and a few naive assumptions described in Section 2, we believe our methodology to be flexible and, in what follows, we outline some potential directions to extend the framework. First, when GRNs are modeled through more complex computation, e.g., with different logic gates connecting multiple expression activators/suppressors/products, those GRNs that only consist of naive interactions are never excluded. Thus, the current model represents a subset of the complete landscape built by more complex gene regulation. The derived connectivity and accessibility among the naive GRNs still hold, and we expect these topographical features to manifest for complex GRNs if mutations between the simple and complex expressions are permitted. Second, hypergraphs [90] could be used to describe the expression behavior of genes where multiple activators/products appear. Third, stable motif identification [91] and target control [92] for Boolean network models could be used to explore the phenotypes of mutational neighbors of a focal complex GRN. Lastly, our methodologies are likely applicable to other classes of genotype–phenotype maps [93,94]. In particular, once the mapping and the genotype network are determined, one can simply follow the proposed iterative procedure (Figure 6) to obtain a genotype partition coarser than the equivalence classes.

More broadly, this work showcases the potential of combining biological computation across different scales along the hierarchy of living systems. Computing biological functionality on the organism level with genotype–phenotype mapping provides a blueprint of the overall fitness landscape, where evolutionary processes occur/compute on the population level. Furthermore, several intriguing perspectives arise from the proposed mesoscopic backbone if we consider evolution to be a random walk on the fitness landscape. The process of evolution not only manifests genotypes with higher fitness values but also reveals genotypes whose mutational neighbors are more fit [19,23,78]; in other words, the prevalence of different genotypes would reflect the connection counts between equivalence classes of GRNs. In addition, these “connection counts” could become associated with an analogous theory of computation in evolution that addresses questions such as how likely a genotype in an equivalence class is to evolve into a specified phenotype, as well as how likely it is to “reset” to another genotype in the same equivalence class and recover its position in the fitness landscape.

Author Contributions: Conceptualization, C.-H.Y. and S.V.S.; Formal analysis, C.-H.Y.; Funding acquisition, S.V.S.; Investigation, C.-H.Y.; Methodology, C.-H.Y.; Project administration, S.V.S.; Supervision, S.V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Central and Peripheral GRNs Where No Regulation Presents

Here, we show the connectivity of “central” and “peripheral” GRNs mentioned in Section 3.1 under the extreme scenario where no gene regulation appears, in particular, when $|\Omega_y^+| = |\Gamma|$. We observe that, since $|\Omega_y^+| = |\Gamma|$, each node $\omega \in \Omega_y^+$ is incident to one and only one incoming edge, and thus, each central GRN $\tilde{g} \in \tilde{\mathcal{G}}_y$ corresponds to a bijective mapping from Ω_y^+ to edge labels Γ .

First, we show that if two central GRNs $\tilde{g}_1, \tilde{g}_2 \in \tilde{\mathcal{G}}_y$ correspond to different mappings between Ω_y^+ and Γ , there is no mutational trajectory connecting them among \mathcal{G}_y . Start with assuming that such a mutational trajectory does exist. Due to the different associated mappings of \tilde{g}_1 and \tilde{g}_2 , there is an $\omega \in \Omega$ with distinct labels of the incident edge γ_1 and γ_2 in \tilde{g}_1 and \tilde{g}_2 , respectively. Because mutating \tilde{g}_1 into \tilde{g}_2 requires rewiring both γ_1 and γ_2 , there must exist a GRN g where either none or both the edges γ_1 and γ_2 point to ω . Nevertheless, g contradicts our observation following the constraint $|\Omega_y^+| = |\Gamma|$ so $g \notin \mathcal{G}_y$. As a result, under this extreme scenario, \mathcal{G}_y fragment into multiple connected components when only mutations among themselves are considered.

Next, for any phenotype y' for which $|\Omega_y^+ \cup \Omega_{y'}^+| \leq |\Gamma| + 1$, we show that there is a mutational trajectory among \mathcal{G}_y connecting an arbitrary central GRN $\tilde{g} \in \tilde{\mathcal{G}}_y$ and a peripheral GRN $\hat{g} \in \mathcal{G}_{y'}$ at the boundary of \mathcal{G}_y and $\mathcal{G}_{y'}$. Specifically, take $\omega \in \Omega_y^+ \setminus \Omega_{y'}^+$ and its incident edge pointing from $\omega_0 \in \Omega_0$. For each $\omega' \in \Omega_y^+ \setminus \Omega_{y'}^+$, where $\omega' \neq \omega$, one can sequentially rewire the incident edge of ω' to form a chain of $\Omega_y^+ \setminus \Omega_{y'}^+$ initiated by ω_0 , which leads to a resultant GRN \hat{g} . Moreover, since $|\Omega_y^+| = |\Gamma|$, we have $|\Omega_{y'}^+ \setminus \Omega_y^+| = 1$. Recall from Section 3.1, this \hat{g} is indeed a peripheral GRN between \mathcal{G}_y and $\mathcal{G}_{y'}$.

Appendix B. Phenotype-Preserving Automorphisms of the Genotype Network of GRNs

In this section, we demonstrate (a) why GRNs mapped by automorphisms of the genotype network G are equivalent, (b) four graphical operations that generate phenotype-preserving automorphisms of G , and (c) the correctness of our iterative procedure to obtain a coarser partition than the equivalence classes of GRNs.

Proposition A1. *Given an automorphism σ of the genotype network G and a mega-node function f_G that depends on the adjacency matrix \mathbf{A} of G , for any GRNs $g_1, g_2 \in \mathcal{G}$ where $g_2 = \sigma(g_1)$, $f_G(g_2) = f_G(g_1)$.*

Proof. Take $f'_G = f_G \circ \sigma$. Since for any $g_1, g_2 \in V(G)$, $(\sigma(g_1), \sigma(g_2)) \in E(G)$ if and only if $(g_1, g_2) \in E(G)$, the adjacency matrix A remains unchanged after permuting the mega-nodes through σ . As a result, we have $f'_G = f_G$, and $f_G(g_2) = f'_G(g_1) = f_G(g_1)$. \square

Let π and π' be a permutation of the loci Γ and the dummy proteins Ω' , respectively. It is not hard to see that π and π' also generate a permutation of the GRNs \mathcal{G} . For $g \in \mathcal{G}$, we abuse the notation $\pi(g)$ to be the GRN mapped through the locus permutation π , where an edge with $e_g(\gamma) = (u, v)$ is transformed into $e_{\pi(g)}(\pi(\gamma)) = (u, v)$. Similarly, in the GRN

$\pi'(g)$ mapped through the dummy protein permutation π' , an edge with $e_g(\gamma) = (u, v)$ is transformed into $e_{g'}(\gamma) = (\pi(u), \pi(v))$.

Furthermore, we have two more types of graphical operations on GRNs:

Definition A1. For a locus γ and two stimuli $\omega, \omega' \in \Omega_0$, $\rho_{\gamma, \omega, \omega'} : \mathcal{G} \rightarrow \mathcal{G}$ transforms a GRN g into g' such that edge γ becomes

$$\begin{cases} e_{g'}(\gamma) = (u, \omega') \text{ if } e_g(\gamma) = (u, \omega), \\ e_{g'}(\gamma) = (u, \omega) \text{ if } e_g(\gamma) = (u, \omega'), \\ \text{unchanged otherwise.} \end{cases}$$

Definition A2. For a locus γ and two nodes ω and ω' , $\varrho_{\gamma, \omega, \omega'} : \mathcal{G} \rightarrow \mathcal{G}$ transforms a GRN g into g' such that the edge γ becomes

$$\begin{cases} e_{g'}(\gamma) = (\omega', \omega') \text{ if } e_g(\gamma) = (\omega, \omega), \\ e_{g'}(\gamma) = (\omega, \omega) \text{ if } e_g(\gamma) = (\omega', \omega'), \\ \text{unchanged otherwise.} \end{cases}$$

Note both $\rho_{\gamma, \omega, \omega'}$ and $\varrho_{\gamma, \omega, \omega'}$ are permutations of \mathcal{G} as well, where pairs of GRNs are mutually mapped from one to the other.

These four graphical operations introduced above are, more importantly, automorphisms of the genotype network G that also preserve the phenotype of GRNs:

Theorem A1. The transformations π , π' , $\rho_{\gamma, \omega, \omega'}$, and $\varrho_{\gamma, \omega, \omega'}$ are phenotype-preserving automorphisms of G .

Proof. For $g_1, g_2 \in \mathcal{G}$, let $\Delta = \{\gamma \in \Gamma \mid e_{g_1}(\gamma) \neq e_{g_2}(\gamma)\}$. Since π' is a permutation of dummy proteins, it preserves Δ . Additionally, because $\rho_{\gamma, \omega, \omega'}$ and $\varrho_{\gamma, \omega, \omega'}$ can be viewed as permutations of the source–target pair of a single edge γ , they also preserve Δ . The permutation π of Γ , on the other hand, does not preserve Δ but maintains its size $|\Delta|$. Since $(g_1, g_2) \in E(G)$ if and only if $|\Delta| = 1$, the four transformations are automorphisms of G .

Furthermore, since π , π' and $\varrho_{\gamma, \omega, \omega'}$ simply change the labels of edges, labels of the intermediate nodes, and the location of a self-loop, they maintain any path from a stimulus $\omega_0 \in \Omega_0$ to a fitness-relevant protein $\hat{\omega} \in \hat{\Omega}$. $\rho_{\gamma, \omega, \omega'}$ may alter the path between ω_0 and $\hat{\omega}$, but the reachability from Ω_0 to $\hat{\omega}$ remains. Therefore, the four transformations also preserve the phenotype of GRNs. \square

Finally, we turn to the computationally acquired partition that can be shown to be coarser than the equivalence classes of GRNs. Recall from Section 3.2 that our iterative procedure starts from a partition $\hat{\Theta}_0$ where GRNs with the same phenotype are grouped together. Given the partition $\hat{\Theta}_i$, the next partition $\hat{\Theta}_{i+1}$ is obtained by further dividing groups into $\hat{\Theta}_i$ (if needed) such that for each group $\theta \in \hat{\Theta}_i$ and $\theta' \in \hat{\Theta}_{i+1}$, any two GRNs in θ' have the same number of neighbors among θ in the genotype network G . The procedure terminates when a stationary partition $\hat{\Theta}$ is reached. We then have:

Theorem A2. Every equivalence class $\theta \in \Theta$ is included in a group $\hat{\theta} \in \hat{\Theta}$, $\theta \subset \hat{\theta}$.

Proof. Recall that GRNs in an equivalence class have the same phenotype, so for each $\theta \in \Theta$, there is some $\theta_0 \in \Theta_0$ where $\theta \subset \theta_0$. Suppose that $\theta \subset \theta_i$ for each $\theta \in \Theta$ and some $\theta_i \in \Theta_i$. Since Θ forms an equitable partition, every $g \in \theta$ has the same number of neighbors in each $\theta' \in \Theta$ and thus also in each $\theta_i \in \Theta_i$. Consequently, no two GRNs in θ will be separated into two different groups in Θ_{i+1} , and the theorem is proved by induction. \square

Appendix C. Combining Mutational Neighbors into Equivalence Classes

In this section, we tackle the two questions raised in Section 3.3:

- (A) For an equivalence class $\theta \in \Theta_k$ and its representative GRN $g \in \theta$, under what condition will $g'_1, g'_2 \in \mathcal{M}^+(g)$ belong to the same equivalence class in layer Θ_{k+1} ?
- (B) For two distinct equivalence classes $\theta_1, \theta_2 \in \Theta_k$ and their representative GRNs $g_1 \in \theta_1$ and $g_2 \in \theta_2$, under what condition will $g'_1 \in \mathcal{M}^+(g_1)$ and $g'_2 \in \mathcal{M}^+(g_2)$ belong to the same equivalence class in layer Θ_{k+1} ?

Furthermore, recall that for ease of demonstration, we constrain the GRNs g, g_1, g_2, g'_1 and g'_2 where only one stimulus node is incident to out-going edges.

Definition A3. A phenotype-preserving isomorphism π from g'_1 to g'_2 after self-loop removal is a permutation of loci Γ and dummy protein Ω' such that for any locus γ and non-self-loop source–target pair (u, v) , $e_{g'_2}(\pi(\gamma)) = (\pi(u), \pi(v))$ if and only if $e_{g'_1}(\gamma) = (u, v)$.

Starting with the question (A), we write $e_{g'_1}(\gamma_1) = (u_1, v_1)$ and $e_{g'_2}(\gamma_2) = (u_2, v_2)$, where γ_1 and γ_2 are the non-self-loop edge newly rewired to generate g'_1 and g'_2 from g , respectively. A few observations appear when we assume g'_1 and g'_2 belong to the same equivalence class:

Lemma A1. Suppose a phenotype-preserving isomorphism π from g'_1 to g'_2 after self-loop removal. There are two integers $q < p$ such that $\pi^p(\gamma_1) = \gamma_1$, $(\pi^p(u_1), \pi^p(v_1)) = (u_1, v_1)$, and $\pi^q(\gamma_1) = \gamma_2$, $(\pi^q(u_1), \pi^q(v_1)) = (u_2, v_2)$.

Proof. Since π is a permutation of finite sets, it must have a finite period, i.e., an integer p such that $\pi^p(\gamma_1) = \gamma_1$ and $(\pi^p(u_1), \pi^p(v_1)) = (u_1, v_1)$.

If $\pi(\gamma_1) = \gamma_2$ and $(\pi(u_1), \pi(v_1)) = (u_2, v_2)$, then we have $q = 1$. Otherwise, it must map to a non-self-loop edge in g because γ_2 is the only additional non-self-loop edge in g_2 , i.e., $e_g(\pi(\gamma_1)) = (\pi(u_1), \pi(v_1))$. Assume there is no integer $q < p$ such that $\pi^q(\gamma_1) = \gamma_2$ and $(\pi^q(u_1), \pi^q(v_1)) = (u_2, v_2)$. Then, $\pi^{p-1}(\gamma_1)$ is a non-self-loop edge in g with $e_g(\pi^{p-1}(\gamma_1)) = (\pi^{p-1}(u_1), \pi^{p-1}(v_1))$. However, since γ_1 is not a non-self-loop edge in g_2 , the fact that $\pi^p(\gamma_1) = \gamma_1$ and $(\pi^p(u_1), \pi^p(v_1)) = (u_1, v_1)$ contradicts π being an isomorphism from g_1 to g_2 after self-loop removal. Therefore, there is an integer $q < p$ such that $\pi^q(\gamma_1) = \gamma_2$ and $(\pi^q(u_1), \pi^q(v_1)) = (u_2, v_2)$. \square

Lemma A2. Suppose a phenotype-preserving isomorphism π from g'_1 to g'_2 after self-loop removal. For integers p and q in Lemma A1, $e_{g'_2}(\pi^k(\gamma_1)) = (\pi^k(u_1), \pi^k(v_1))$ for $k = 1, 2, \dots, q$, and $e_{g'_2}(\pi^k(\gamma_1)) \neq (\pi^k(u_1), \pi^k(v_1))$ for $k = q + 1, q + 2, \dots, p$.

Proof. Since π is an isomorphism and $\gamma_2 = \pi^q(\gamma_1)$ is the only additional non-self-loop edge in g_2 , we have $\pi(\gamma_1), \pi^2(\gamma_1), \dots, \pi^{q-1}(\gamma_1)$ to be non-self-loop edges in g . Thus, $e_{g'_2}(\pi^k(\gamma_1)) = (\pi^k(u_1), \pi^k(v_1))$ for $k = 1, 2, \dots, q$. On the other hand, since $\gamma_2 = \pi^q(\gamma_1)$ is not a non-self-loop edge in g_1 , the isomorphism π guaranteed that the source–target pairs $(\pi^{q+1}(u_1), \pi^{q+1}(v_1)), (\pi^{q+2}(u_1), \pi^{q+2}(v_1)), \dots, (\pi^p(u_1), \pi^p(v_1))$ do not match to edges in g_2 , in particular, $e_{g'_2}(\pi^k(\gamma_1)) \neq (\pi^k(u_1), \pi^k(v_1))$ for $k = q + 1, q + 2, \dots, p$. \square

Lemma A3. Suppose a phenotype-preserving isomorphism π from g'_1 to g'_2 after self-loop removal. Given integer q in Lemma A1, for a locus γ and non-self-loop source–target pair (u, v) where there is no $0 \leq k \leq q - 1$ such that $(\gamma, u, v) = (\pi^k(\gamma), \pi^k(u_1), \pi^k(v_1))$, $e_g(\pi(\gamma)) = (\pi(u), \pi(v))$ if and only if $e_g(\gamma) = (u, v)$.

Proof. For $(\gamma, u, v) \notin \left\{ (\pi^k(\gamma), \pi^k(u_1), \pi^k(v_1)) \right\}_{k=0}^{p-1}$, since (γ_1, u_1, v_1) and (γ_2, u_2, v_2) are already excluded, and by Definition A3, we have $e_g(\pi(\gamma)) = (\pi(u), \pi(v))$ if and only

if $e_g(\gamma) = (u, v)$. Furthermore, for $k = q, q + 1, \dots, p - 1$, because the only additional non-self-loop edge in g'_2 follows $\pi^q(\gamma_1) = \gamma_2$ and $(\pi^q(u_1), \pi^q(v_1)) = (u_2, v_2)$, according to Lemma A2, we know that $e_g(\pi^k(\gamma_1)) \neq (\pi^k(u_1), \pi^k(v_1))$ and $e_g(\pi^{k+1}(\gamma_1)) \neq (\pi^{k+1}(u_1), \pi^{k+1}(v_1))$. As a result, the statement in Lemma A3 is true for any locus γ and non-self-loop source–target pair (u, v) where $(\gamma, u, v) \notin \left\{ (\pi^k(\gamma), \pi^k(u_1), \pi^k(v_1)) \right\}_{k=0}^{q-1}$. \square

The following theorem resolves the necessary and sufficient condition in question (A):

Theorem A3. Let $g'_1, g'_2 \in \mathcal{M}^+(g)$ with the additional non-self-loop edge $e_{g'_1}(\gamma_1) = (u_1, v_1)$ and $e_{g'_2}(\gamma_2) = (u_2, v_2)$, respectively. g'_1 and g'_2 belong to the same equivalence class if and only if there exist two integers $q < p$ and a phenotype-preserving automorphism σ of a subgraph \bar{g} of g such that

- i. $(\sigma^q(\gamma_1), \sigma^q(u_1), \sigma^q(v_1)) = (\gamma_2, u_2, v_2)$;
- ii. $(\sigma^p(\gamma_1), \sigma^p(u_1), \sigma^p(v_1)) = (\gamma_1, u_1, v_1)$;
- iii. $\left\{ \sigma^k(\gamma_1) \right\}_{k=1}^{q-1} = \Gamma'(g) \setminus \Gamma'(\bar{g})$ and $e_g(\sigma^k(\gamma_1)) = (\sigma^k(u_1), \sigma^k(v_1))$ for $k = 1, 2, \dots, q - 1$;
- iv. $e_g(\sigma^k(\gamma_1)) \neq (\sigma^k(u_1), \sigma^k(v_1))$ for $k = q, q + 1, \dots, p$.

Proof. For one direction, suppose a phenotype-preserving isomorphism π from g'_1 to g'_2 after self-loop removal. According to Lemmas A1–A3, we see that π is a phenotype-preserving automorphism of a subgraph \bar{g} of g where $\Gamma'(g) \setminus \Gamma'(\bar{g}) = \left\{ \pi^k(\gamma_1) \right\}_{k=1}^{q-1}$. Taking $\sigma = \pi$ satisfies all four conditions.

For the other direction, we show that such a phenotype-preserving automorphism σ is also a phenotype-preserving isomorphism from g'_1 to g'_2 after self-loop removal. First, regarding any locus γ and non-self-loop target pair (u, v) such that $(\gamma, u, v) \notin \left\{ (\sigma^k(\gamma_1), \sigma^k(u_1), \sigma^k(v_1)) \right\}_{k=1}^p$, we have $e_{g'_2}(\sigma(\gamma)) = (\sigma(u), \sigma(v))$ if and only if $e_{g'_1}(\gamma) = (u, v)$ because σ is an automorphism of \bar{g} and γ_1, γ_2 , and $\Gamma'(g) \setminus \Gamma'(\bar{g})$ are already excluded. Second, for $k = 0, 1, \dots, q - 1$, the conditions i.–iii. ensure that $e_{g'_1}(\sigma^k(\gamma)) = (\sigma^k(u), \sigma^k(v))$ and $e_{g'_2}(\sigma^{k+1}(\gamma)) = (\sigma^{k+1}(u), \sigma^{k+1}(v))$. Lastly, for $k = q, q + 1, \dots, p - 1$, the conditions i., ii., and iv. indicate $e_{g'_1}(\sigma^k(\gamma)) \neq (\sigma^k(u), \sigma^k(v))$ and $e_{g'_2}(\sigma^{k+1}(\gamma)) \neq (\sigma^{k+1}(u), \sigma^{k+1}(v))$. \square

Next, the necessary and sufficient condition of question (B) is described in the theorem below:

Theorem A4. Let g_1 and g_2 be of two different equivalence classes, and let $g'_1 \in \mathcal{M}^+(g_1)$ and $g'_2 \in \mathcal{M}^+(g_2)$ with the additional non-self-loop edge $e_{g'_1}(\gamma_1) = (u_1, v_1)$ and $e_{g'_2}(\gamma_2) = (u_2, v_2)$, respectively. g'_1 and g'_2 belong to the same equivalence class if and only if there exist GRNs \tilde{g}_1 and g'' such that

- i. \tilde{g}_1 and g_1 belong to the same equivalence class;
- ii. $\tilde{g}_1, g_2 \in \mathcal{M}^+(g'')$;
- iii. $g'_2 \in \mathcal{M}^+(\tilde{g}_1)$.

Proof. For one direction, suppose a phenotype-preserving isomorphism π from g'_1 to g'_2 after self-loop removal. Take $\tilde{g}_1 = \pi(g_1)$, i.e., $e_{\tilde{g}_1}(\pi(\gamma)) = (\pi(u), \pi(v))$ for each $\gamma \in \Gamma'(g_1)$ with $e_{g_1}(\gamma) = (u, v)$, so π also becomes a phenotype-preserving isomorphism from g_1 to \tilde{g}_1 . Moreover, due to the isomorphism π , we note $g'_2 \in \mathcal{M}^+(\tilde{g}_1)$ with the additional non-self-loop edge $e_{g'_2}(\pi(\gamma_1)) = (\pi(u_1), \pi(v_1))$. Let g'' be a GRN obtained by rewiring edges $\pi(\gamma_1)$ and γ_2 from g'_2 to two arbitrary self-loops. We have $\tilde{g}_1, g_2 \in \mathcal{M}^+(g'')$ with the additional non-self-loop edge $e_{\tilde{g}_1}(\gamma_2) = (u_2, v_2)$ and $e_{g_2}(\pi(\gamma_1)) = (\pi(u_1), \pi(v_1))$, respectively. Observe that the GRNs \tilde{g}_1 and g'' satisfy the conditions i.–iii.

For the other direction, suppose there exist GRNs \tilde{g}_1 and g'' where the conditions i.–iii. hold. Let π be the phenotype-preserving isomorphism from g_1 to \tilde{g}_1 after self-loop removal such that the additional non-self-loop edge of $g_2 \in \mathcal{M}^+(g'')$ complies $e_{g_2}(\pi(\gamma_1)) = (\pi(u_1), \pi(v_1))$. Since $\tilde{g}_1, g_2 \in \mathcal{M}^+(g'')$, $g'_2 \in \mathcal{M}^+(g_2)$, and $g'_2 \in \mathcal{M}^+$, we also have the addition non-self-loop edge in $g'_2 \in \mathcal{M}^+$ to follow $e_{g'_2}(\pi(\gamma_1)) = (\pi(u_1), \pi(v_1))$. Therefore, π is also a phenotype-preserving isomorphism from g'_1 to g'_2 after self-loop removal. \square

Appendix D. Size of an Equivalence Class of GRNs

Here, given the representative GRN g of an equivalence class θ , we calculate the number of GRNs in θ in Equation (2). Observe that $|\theta|$ is proportional to (a) the number of GRNs to which there is a phenotype-preserving isomorphism from g , (b) the number of ways to arbitrarily allocate self-loops on g , and (c) the number of ways to arbitrarily rewire source nodes of edges among the stimuli Ω_0 .

First, for part (a), if we temporarily ignore labels on the edges, the set of permutations over dummy proteins Ω' is partitioned into groups of isomorphisms from g to different GRNs. In addition, the size of each group is exactly the number of automorphisms of g since the composition of an automorphism of g and an isomorphism from g to g' also generates an isomorphism from g to g' . Thus, there are in total $\frac{|\Pi'|}{|\Sigma'(g)|}$ different GRNs isomorphic to g , where Π' is the set of permutations over Ω' , and $\Sigma'(g)$ is the set of automorphisms of g that only permutes Ω' .

Second, for part (b), every possible allocation distributes $|\Gamma| - |\Gamma'(g)|$ self-loops over $|\Omega \setminus \Omega_0|$ nodes. Suppose that the labels of self-loops $\Gamma \setminus \Gamma'(g)$ are given, we have

$$n_l(|\Gamma \setminus \Gamma'(g)|) = |\Omega \setminus \Omega_0|^{|\Gamma \setminus \Gamma'(g)|}.$$

Third, we write $k_s(g)$ as the number of incident edges to stimuli Ω_0 in g . Any possibility in part (c) chooses a source node among Ω_0 for each of the $k_s(g)$ incident edges. Providing that the labels of incident edges of the stimuli are already known, we have

$$m_s(g) = |\Omega_0|^{k_s(g)}.$$

Combined, the size of the equivalence class θ becomes

$$\begin{aligned} |\theta| &= \frac{|\Gamma|!}{|\Gamma \setminus \Gamma'(g)|! k_s(g)!} \frac{|\Pi'|}{|\Sigma'(g)|} n_l(|\Gamma \setminus \Gamma'(g)|) m_s(g) \\ &= \frac{|\Gamma|!}{|\Gamma \setminus \Gamma'(g)|! k_s(g)!} \frac{|\Pi'|}{|\Sigma'(g)|} |\Omega \setminus \Omega_0|^{|\Gamma \setminus \Gamma'(g)|} |\Omega_0|^{k_s(g)} \end{aligned}$$

where the first fraction represents first selecting combinations of $|\Gamma \setminus \Gamma'(g)|$ and $k_s(g)$ labels for self-loops and edges incident to stimuli and then permuting the remaining labels, which also contributes to different phenotype-preserving isomorphisms from g but was previously omitted.

References

1. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In Proceedings of the Sixth International Congress on Genetics, Ithaca, NY, USA, 24–31 August 1932; pp. 356–366.
2. De Visser, J.A.G.M.; Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **2014**, *15*, 480–490. [[CrossRef](#)] [[PubMed](#)]
3. Fragata, I.; Blanckaert, A.; Louro, M.A.D.; Liberles, D.A.; Bank, C. Evolution in the light of fitness landscape theory. *Trends Ecol. Evol.* **2019**, *34*, 69–82. [[CrossRef](#)] [[PubMed](#)]
4. Szendro, I.G.; Schenk, M.F.; Franke, J.; De Visser, J.A.G. Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech. Theory Exp.* **2013**, *2013*, P01005. [[CrossRef](#)]
5. Wagner, A. *The Origins of Evolutionary Innovations: A Theory of Transformative Change in Living Systems*; Oxford University Press: Oxford, UK, 2011.
6. Jain, K.; Krug, J. Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes. *Genetics* **2007**, *175*, 1275–1288. [[CrossRef](#)] [[PubMed](#)]

7. Kryazhimskiy, S.; Tkačik, G.; Plotkin, J.B. The dynamics of adaptation on correlated fitness landscapes. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18638–18643. [[CrossRef](#)]
8. Draghi, J.A.; Parsons, T.L.; Wagner, G.P.; Plotkin, J.B. Mutational robustness can facilitate adaptation. *Nature* **2010**, *463*, 353–355. [[CrossRef](#)]
9. Wu, N.C.; Dai, L.; Olson, C.A.; Lloyd-Smith, J.O.; Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **2016**, *5*, e16965. [[CrossRef](#)]
10. Gavrillets, S. *Fitness Landscapes and the Origin of Species*; Princeton University Press: Princeton, NY, USA, 2004.
11. Fraïsse, C.; Gunnarsson, P.A.; Roze, D.; Bierne, N.; Welch, J.J. The genetics of speciation: Insights from fisher’s geometric model. *Evolution* **2016**, *70*, 1450–1464. [[CrossRef](#)]
12. de Visser, J.A.G.; Park, S.C.; Krug, J. Exploring the effect of sex on empirical fitness landscapes. *Am. Nat.* **2009**, *174*, S15–S30. [[CrossRef](#)]
13. Otto, S.P. The evolutionary enigma of sex. *Am. Nat.* **2009**, *174*, S1–S14. [[CrossRef](#)]
14. Watson, R.A.; Weinreich, D.M.; Wakeley, J. Genome structure and the benefit of sex. *Evol. Int. J. Org. Evol.* **2011**, *65*, 523–536. [[CrossRef](#)] [[PubMed](#)]
15. Orr, H.A. The genetic theory of adaptation: A brief history. *Nat. Rev. Genet.* **2005**, *6*, 119–127. [[CrossRef](#)] [[PubMed](#)]
16. Lobkovsky, A.E.; Wolf, Y.I.; Koonin, E.V. Predictability of evolutionary trajectories in fitness landscapes. *PLoS Comput. Biol.* **2011**, *7*, e1002302. [[CrossRef](#)] [[PubMed](#)]
17. Salverda, M.L.; Dellus, E.; Gorter, F.A.; Debets, A.J.; Van Der Oost, J.; Hoekstra, R.F.; Taw, D.S.; de Visser, J.A.G. Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* **2011**, *7*, e1001321. [[CrossRef](#)] [[PubMed](#)]
18. Bank, C.; Matuszewski, S.; Hietpas, R.T.; Jensen, J.D. On the (un) predictability of a large intragenic fitness landscape. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 14085–14090. [[CrossRef](#)]
19. Van Nimwegen, E.; Crutchfield, J.P.; Huynen, M. Neutral evolution of mutational robustness. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9716–9720. [[CrossRef](#)]
20. Van Nimwegen, E.; Crutchfield, J.P. Metastable evolutionary dynamics: Crossing fitness barriers or escaping via neutral paths? *Bull. Math. Biol.* **2000**, *62*, 799–848. [[CrossRef](#)]
21. Wilke, C.O. Adaptive evolution on neutral networks. *Bull. Math. Biol.* **2001**, *63*, 715–730. [[CrossRef](#)]
22. Smith, T.; Husbands, P.; O’Shea, M. Neutral networks and evolvability with complex genotype-phenotype mapping. In *European Conference on Artificial Life*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 272–281.
23. Aguirre, J.; Catalán, P.; Cuesta, J.A.; Manrubia, S. On the networked architecture of genotype spaces and its critical effects on molecular evolution. *Open Biol.* **2018**, *8*, 180069. [[CrossRef](#)]
24. Manrubia, S.; Cuesta, J.A.; Aguirre, J.; Ahnert, S.E.; Altenberg, L.; Cano, A.V.; Catalán, P.; Diaz-Uriarte, R.; Elena, S.F.; García-Martín, J.A.; et al. From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Phys. Life Rev.* **2021**, *38*, 55–106. [[CrossRef](#)]
25. Tenaillon, O. The utility of fisher’s geometric model in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* **2014**, *45*, 179–201. [[CrossRef](#)] [[PubMed](#)]
26. Fisher, R.A. *The Genetical Theory of Natural Selection*; Oxford University Press: Oxford, UK, 1930.
27. Kingman, J.F.C. A simple model for the balance between selection and mutation. *J. Appl. Probab.* **1978**, *15*, 1–12. [[CrossRef](#)]
28. Kauffman, S.; Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **1987**, *128*, 11–45. [[CrossRef](#)]
29. Kauffman, S.A.; Weinberger, E.D. The nk model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.* **1989**, *141*, 211–245. [[CrossRef](#)]
30. Aita, T.; Uchiyama, H.; Inaoka, T.; Nakajima, M.; Kokubo, T.; Husimi, Y. Analysis of a local fitness landscape with a model of the rough mt. fuji-type landscape: Application to prolyl endopeptidase and thermolysin. *Biopolym. Orig. Res. Biomol.* **2000**, *54*, 64–79. [[CrossRef](#)]
31. Neidhart, J.; Szendro, I.G.; Krug, J. Adaptation in tunably rugged fitness landscapes: The rough mount fuji model. *Genetics* **2014**, *198*, 699–721. [[CrossRef](#)]
32. Weinreich, D.M.; Delaney, N.F.; DePristo, M.A.; Hartl, D.L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **2006**, *312*, 111–114. [[CrossRef](#)]
33. Chou, H.-H.; Chiu, H.-C.; NDelaney, N.F.; Segrè, D.; Marx, C.J. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **2011**, *332*, 1190–1192. [[CrossRef](#)]
34. Khan, A.I.; Dinh, D.M.; Schneider, D.; Lenski, R.E.; Cooper, T.F. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **2011**, *332*, 1193–1196. [[CrossRef](#)]
35. De Visser, J.A.G.M.; Hoekstra, R.F.; van den Ende, H. Test of interaction between genetic markers that affect fitness in *aspergillus niger*. *Evolution* **1997**, *51*, 1499–1505. [[CrossRef](#)]
36. Hall, D.W.; Agan, M.; Pope, S.C. Fitness epistasis among 6 biosynthetic loci in the budding yeast *saccharomyces cerevisiae*. *J. Hered.* **2010**, *101* (Suppl. S1), S75–S84. [[CrossRef](#)] [[PubMed](#)]
37. Whitlock, M.C.; Bourguet, D. Factors affecting the genetic load in *drosophila*: Synergistic epistasis and correlations among fitness components. *Evolution* **2000**, *54*, 1654–1660. [[CrossRef](#)] [[PubMed](#)]

38. Hinkley, T.; Martins, J.; Chappey, C.; Haddad, M.; Stawiski, E.; Whitcomb, J.M.; Petropoulos, C.J.; Bonhoeffer, S. A systems analysis of mutational effects in hiv-1 protease and reverse transcriptase. *Nat. Genet.* **2011**, *43*, 487–489. [[CrossRef](#)] [[PubMed](#)]
39. Kouyos, R.D.; Leventhal, G.E.; Hinkley, T.; Haddad, M.; Whitcomb, J.M.; Petropoulos, C.J.; Bonhoeffer, S. Exploring the complexity of the hiv-1 fitness landscape. *PLoS Genet.* **2012**, *8*, e1002551. [[CrossRef](#)] [[PubMed](#)]
40. Hietpas, R.T.; Jensen, J.D.; Bolon, D.N.A. Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7896–7901. [[CrossRef](#)] [[PubMed](#)]
41. Otwinowski, J.; Nemenman, I. Genotype to phenotype mapping and the fitness landscape of the E. coli lac promoter. *PLoS ONE* **2013**, *8*, e61570. [[CrossRef](#)]
42. Sarkisyan, K.S.; Bolotin, D.A.; Meer, M.V.; Usmanova, D.R.; Mishin, A.S.; Sharonov, G.V.; Ivankov, D.N.; Bozhanova, N.G.; Baranov, M.S.; Soylemez, O.; et al. Local fitness landscape of the green fluorescent protein. *Nature* **2016**, *533*, 397–401. [[CrossRef](#)]
43. Rogers, Z.N.; McFarland, C.D.; Winters, I.P.; Seoane, J.A.; Brady, J.J.; Yoon, S.; Curtis, C.; Petrov, D.A.; Winslow, M.M. Mapping the in vivo fitness landscape of lung adenocarcinoma tumor suppression in mice. *Nat. Genet.* **2018**, *50*, 483–486. [[CrossRef](#)]
44. Watson, C.J.; Papula, A.L.; Poon, G.Y.P.; Wong, W.H.; Young, A.L.; Druley, T.E.; Fisher, D.S.; Blundell, J.R. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science* **2020**, *367*, 1449–1454. [[CrossRef](#)]
45. Rowe, W.; Platt, M.; Wedge, D.C.; Day, P.J.; Kell, D.B.; Knowles, J. Analysis of a complete DNA–protein affinity landscape. *J. R. Soc. Interface* **2010**, *7*, 397–408. [[CrossRef](#)]
46. Pitt, J.N.; Ferré-D’Amaré, A.R. Rapid construction of empirical rna fitness landscapes. *Science* **2010**, *330*, 376–379. [[CrossRef](#)] [[PubMed](#)]
47. Jiménez, J.I.; Xulvi-Brunet, R.; Campbell, G.W.; Turk-MacLeod, R.; Chen, I.A. Comprehensive experimental fitness landscape and evolutionary network for small rna. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14984–14989. [[CrossRef](#)] [[PubMed](#)]
48. Li, C.; Qian, W.; Maclean, C.J.; Zhang, J. The fitness landscape of a trna gene. *Science* **2016**, *352*, 837–840. [[CrossRef](#)] [[PubMed](#)]
49. Bendixsen, D.P.; Østman, B.; Hayden, E.J. Negative epistasis in experimental rna fitness landscapes. *J. Mol. Evol.* **2017**, *85*, 159–168. [[CrossRef](#)]
50. Aguilar-Rodríguez, J.; Payne, J.L.; Wagner, A. A thousand empirical adaptive landscapes and their navigability. *Nat. Ecol. Evol.* **2017**, *1*, 0045. [[CrossRef](#)]
51. Martin, C.H. Context dependence in complex adaptive landscapes: Frequency and trait-dependent selection surfaces within an adaptive radiation of caribbean pupfishes. *Evolution* **2016**, *70*, 1265–1282. [[CrossRef](#)]
52. Boucher, J.L.; Cote, P.; Flynn, J.; Jiang, L.; Laban, A.; Mishra, P.; Roscoe, B.P.; Bolon, D.N.A. Viewing protein fitness landscapes through a next-gen lens. *Genetics* **2014**, *198*, 461–471. [[CrossRef](#)]
53. Aita, T.; Iwakura, M.; Husimi, Y. A cross-section of the fitness landscape of dihydrofolate reductase. *Protein Eng.* **2001**, *14*, 633–638. [[CrossRef](#)]
54. Carneiro, M.; Hartl, D.L. Adaptive landscapes and protein evolution. *Proc. Natl. Acad. Sci. USA* **2010**, *107* (Suppl. S1), 1747–1751. [[CrossRef](#)]
55. Poelwijk, F.J.; Tănase-Nicola, S.; Kiviet, D.J.; Tans, S.J. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* **2011**, *272*, 141–144. [[CrossRef](#)]
56. Lozovsky, E.R.; Chookajorn, T.; Brown, K.M.; Imwong, M.; Shaw, P.J.; Kamchonwongpaisan, S.; Neafsey, D.E.; Weinreich, D.M.; Hartl, D.L. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12025–12030. [[CrossRef](#)] [[PubMed](#)]
57. Lunzer, M.; Miller, S.P.; Felsheim, R.; Dean, A.M. The biochemical architecture of an ancient adaptive landscape. *Science* **2005**, *310*, 499–501. [[CrossRef](#)] [[PubMed](#)]
58. Bridgham, J.T.; Carroll, S.M.; Thornton, J.W. Evolution of hormone-receptor complexity by molecular exploitation. *Science* **2006**, *312*, 97–101. [[CrossRef](#)] [[PubMed](#)]
59. Poelwijk, F.J.; Kiviet, D.J.; Tans, S.J. Evolutionary potential of a duplicated repressor-operator pair: Simulating pathways using mutation data. *PLoS Comput. Biol.* **2006**, *2*, e58. [[CrossRef](#)] [[PubMed](#)]
60. Poelwijk, F.J.; Kiviet, D.J.; Weinreich, D.M.; Tans, S.J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **2007**, *445*, 383–386. [[CrossRef](#)] [[PubMed](#)]
61. Franke, J.; Klözer, A.; de Visser, J.A.G.; Krug, J. Evolutionary accessibility of mutational pathways. *PLoS Comput. Biol.* **2011**, *7*, e1002134. [[CrossRef](#)]
62. Hegarty, P.; Martinsson, A. On the existence of accessible paths in various models of fitness landscapes. *Ann. Appl. Probab.* **2014**, *24*, 1375–1395. [[CrossRef](#)]
63. Zagorski, M.; Burda, Z.; Waclaw, B. Beyond the hypercube: Evolutionary accessibility of fitness landscapes with realistic mutational networks. *PLoS Comput. Biol.* **2016**, *12*, e1005218. [[CrossRef](#)]
64. Kell, D.B. Genotype–phenotype mapping: Genes as computer programs. *Trends Genet.* **2002**, *18*, 555–559. [[CrossRef](#)]
65. Kondrashov, D.A.; Kondrashov, F.A. Topological features of rugged fitness landscapes in sequence space. *Trends Genet.* **2015**, *31*, 24–33. [[CrossRef](#)]
66. Chan, H.S.; Bornberg-Bauer, E. Perspectives on protein evolution from simple exact models. *Appl. Bioinform.* **2002**, *50*, 121–144.
67. Cowperthwaite, M.C.; Economou, E.P.; Harcombe, W.R.; Miller, E.L.; Meyers, L.A. The ascent of the abundant: How mutational networks constrain evolution. *PLoS Comput. Biol.* **2008**, *4*, e1000110. [[CrossRef](#)] [[PubMed](#)]

68. Stich, M.; Lázaro, E.; Manrubia, S.C. Phenotypic effect of mutations in evolving populations of rna molecules. *BMC Evol. Biol.* **2010**, *10*, 46. [[CrossRef](#)] [[PubMed](#)]
69. Palmer, M.E.; Moudgil, A.; Feldman, M.W. Long-term evolution is surprisingly predictable in lattice proteins. *J. R. Soc. Interface* **2013**, *10*, 20130026. [[CrossRef](#)]
70. Bershtein, S.; Serohijos, A.W.R.; Shakhnovich, E.I. Bridging the physical scales in evolutionary biology: From protein sequence space to fitness of organisms and populations. *Curr. Opin. Struct. Biol.* **2017**, *42*, 31–40. [[CrossRef](#)] [[PubMed](#)]
71. Perfeito, L.; Ghozzi, S.; Berg, J.; Schnetz, K.; Lässig, M. Nonlinear fitness landscape of a molecular pathway. *PLoS Genet.* **2011**, *7*, e1002160. [[CrossRef](#)]
72. Chou, H.-H.; Delaney, N.F.; Draghi, J.A.; Marx, C.J. Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLoS Genet.* **2014**, *10*, e1004149. [[CrossRef](#)]
73. Friedlander, T.; Prizak, R.; Barton, N.H.; Tkačik, G. Evolution of new regulatory functions on biophysically realistic fitness landscapes. *Nat. Commun.* **2017**, *8*, 216. [[CrossRef](#)]
74. Cuypers, T.D.; Rutten, J.P.; Hogeweg, P. Evolution of evolvability and phenotypic plasticity in virtual cells. *BMC Evol. Biol.* **2017**, *17*, 60. [[CrossRef](#)]
75. Yubero, P.; Manrubia, S.; Aguirre, J. The space of genotypes is a network of networks: Implications for evolutionary and extinction dynamics. *Sci. Rep.* **2017**, *7*, 13813. [[CrossRef](#)]
76. Harmand, N.; Gallet, R.; Jabbour-Zahab, R.; Martin, G.; Lenormand, T. Fisher's geometrical model and the mutational patterns of antibiotic resistance across dose gradients. *Evolution* **2017**, *71*, 23–37. [[CrossRef](#)] [[PubMed](#)]
77. Yang, C.-H.; Scarpino, S.V. Reproductive barriers as a byproduct of gene network evolution. *bioRxiv* **2020**. [[CrossRef](#)]
78. Yang, C.-H.; Scarpino, S.V. The ensemble of gene regulatory networks at mutation-selection balance. *bioRxiv* **2021**. [[CrossRef](#)]
79. Ciliberti, S.; Martin, O.C.; Wagner, A. Innovation and robustness in complex regulatory gene networks. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13591–13596. [[CrossRef](#)] [[PubMed](#)]
80. Payne, J.L.; Wagner, A. Latent phenotypes pervade gene regulatory circuits. *BMC Syst. Biol.* **2014**, *8*, 64. [[CrossRef](#)]
81. Godsil, C.D. Compact graphs and equitable partitions. *Linear Algebra Its Appl.* **1997**, *255*, 259–266. [[CrossRef](#)]
82. Das, S.G.; Direito, S.O.L.; Waclaw, B.; Allen, R.J.; Krug, J. Predictable properties of fitness landscapes induced by adaptational tradeoffs. *eLife* **2020**, *9*, e55155. [[CrossRef](#)]
83. Wagner, A. Neutralism and selectionism: A network-based reconciliation. *Nat. Rev. Genet.* **2008**, *9*, 965–974. [[CrossRef](#)]
84. Bendixsen, D.P.; Collet, J.; Østman, B.; Hayden, E.J. Genotype network intersections promote evolutionary innovation. *PLoS Biol.* **2019**, *17*, e3000300. [[CrossRef](#)]
85. Hunt, G.; Hopkins, M.J.; Lidgard, S. Simple versus complex models of trait evolution and stasis as a response to environmental change. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4885–4890. [[CrossRef](#)]
86. Heasley, L.R.; Sampaio, N.M.V.; Argueso, J.L. Systemic and rapid restructuring of the genome: A new perspective on punctuated equilibrium. *Curr. Genet.* **2020**, *67*, 57–63. [[CrossRef](#)] [[PubMed](#)]
87. Aguilar-Rodríguez, J.; Peel, L.; Stella, M.; Wagner, A.; Payne, J.L. The architecture of an empirical genotype-phenotype map. *Evolution* **2018**, *72*, 1242–1260. [[CrossRef](#)] [[PubMed](#)]
88. López-Presa, J.L.; Chiroque, L.F.; Anta, A.F. Novel techniques to speed up the computation of the automorphism group of a graph. *J. Appl. Math.* **2014**, *2014*, 934637. [[CrossRef](#)]
89. Stoichev, S.D. New exact and heuristic algorithms for graph automorphism group and graph isomorphism. *J. Exp. Algorithmics (JEA)* **2019**, *24*, 1–27. [[CrossRef](#)]
90. Battiston, F.; Cencetti, G.; Iacopini, I.; Latora, V.; Lucas, M.; Patania, A.; Young, J.-G.; Petri, G. Networks beyond pairwise interactions: Structure and dynamics. *Phys. Rep.* **2020**, *874*, 1–92. [[CrossRef](#)]
91. Maheshwari, P.; Albert, R. A framework to find the logic backbone of a biological network. *BMC Syst. Biol.* **2017**, *11*, 122. [[CrossRef](#)]
92. Yang, G.; Gómez Tejada Zañudo, J.; Albert, R. Target control in logical models using the domain of influence of nodes. *Front. Physiol.* **2018**, *9*, 454. [[CrossRef](#)]
93. Hu, T.; Tomassini, M.; Banzhaf, W. A network perspective on genotype-phenotype mapping in genetic programming. *Genet. Program. Evolvable Mach.* **2020**, *21*, 375–397. [[CrossRef](#)]
94. Greenbury, S.F.; Louis, A.A.; Ahnert, S.E. The structure of genotype-phenotype maps makes fitness landscapes navigable. *bioRxiv* **2021**. [[CrossRef](#)]