Patterns of selection across gene regulatory networks

Jeanne M. C. McDonald and Robert D. Reed

Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, United States

Abstract

Gene regulatory networks (GRNs) are the core engine of organismal development. If we would like to understand the origin and diversification of phenotypes, it is necessary to consider the structure of GRNs in order to reconstruct the links between genetic mutations and phenotypic change. Much of the progress in evolutionary developmental biology, however, has occurred without a nuanced consideration of the evolution of functional relationships between genes, especially in the context of their broader network interactions. Characterizing and comparing GRNs across traits and species in a more detailed way will allow us to determine how network position influences what genes drive adaptive evolution. In this perspective paper, we consider the architecture of developmental GRNs and how positive selection strength may vary across a GRN. We then propose several testable models for these patterns of selection and experimental approaches to test these models.

Keywords

Development, gene regulatory networks, natural selection, gene regulation, protein evolution

1. Introduction

Organisms assemble themselves through an orchestrated sequence of genes being expressed in different combinations, at different times, in different cells. The logic underlying this orchestration emerges largely from interactions between the genes themselves, and these interactions comprise vast and complex regulatory networks capable of allowing single cells to construct things like mushrooms or hedgehogs. Accordingly, in 2007 Wilkins [1] argued that a gene network-based approach was necessary to advance the field of evolutionary developmental biology. At that time, evo-devo was largely focused on studies showing changes in the regulation of individual genes associated with the evolution, and often convergent evolution, of morphological traits [2–4]. While this is still largely the state of the field, a larger philosophical question continues to crystallize and become more urgent: Why do some genes seem to be more likely to facilitate morphological evolution than others? Drawing on concepts of gene regulatory networks (GRNs) [5], Stern and Orgogozo [6] proposed that these genes occupy unique positions within developmental networks such that they integrate many inputs and regulate many outputs.

Few studies have explicitly tested this idea, however [7], and the evolutionary consequences of many other features of GRNs have also yet to be explored [5]. These include the idea that some highly essential subnetworks, or network 'kernels', are evolutionarily constrained, while other subnetworks that can be co-opted for different functions, or network 'plug-ins', are more evolutionarily labile [5]. The type of gene regulation circuitry could also indicate the degree of evolutionary constraint on different genes [8]. A larger body of evo-devo research has instead focused on other questions concerning the genetics of adaptation, such as whether adaptive evolution is occurring primarily in *cis* vs. *trans* sequences or via *de novo* mutations vs. standing variation [6,9]. The literature on adaptive trait evolution still remains relatively separate from the growing body of literature on network evolution in other fields of biology. These literatures include the study of network evolution *in silico* [10,11], as well as the wealth of information on protein-protein interaction (PPI) networks and the distribution of evolutionary rates across these PPI networks [12–14]. We propose that research in these fields can help inform our predictions for the evolution of GRNs.

Over a decade after Wilkins' essay [1], his proposed GRN-oriented reframing of evo-devo still eludes us. The developmental GRNs for some traits have been described in great detail, such as the GRN for sea urchin embryogenesis, yet we still have little understanding of the role of selection in shaping such networks [15]. Some studies have begun to describe the distribution of selection using networks constructed from gene co-expression correlation matrices [16]. However, the conclusions we can make from these types of transcriptomic studies are limited by our lack of knowledge of gene regulatory

70 interactions. In this perspective paper, we will discuss patterns in GRN structure and
71 key case studies of GRNs for adaptive traits before proposing several testable
72 hypotheses for how positive selection pressure could vary across this GRN topology.
73 We then consider how generalizable these predictions are across different types of
74 GRNs and recommend approaches to test these predictions.
75
76 **2. GRN structure and gene connectivity**
77 There are two primary ways a GRN can evolve. First, a network can gain or lose
78 components, such as by cis-regulatory elements (CREs) gaining or losing new binding
79 sites or proteins changing regulatory targets[1]. Second, the timing, location, or level of
80 expression of genes within a network can evolve via changes to either component
81 proteins or CREs. For example, a common hypothesis in studies of co-option is that
82 complete or partial networks are simply re-activated and redeployed at a different time
83 or location, without many changes to their components, to drive the development of new
84 traits [19]. This evolution may occur at some positions in a network moreso than others,
85 so to understand gene evolution we must first characterize the structure of a network.
86
87 As discussed in the introduction, so-called input-output genes are well-known for their
88 proposed role in driving morphological evolution due to their distinct network positions
89 [20]. Input-output genes are identified as switch genes in a GRN, where they integrate
90 the inputs of many upstream patterning genes to control the activation of many
91 downstream cell differentiation genes. Many input-outputs are characterized by their

[1] GRNs can also expand in other ways, see: [17,18] for examples of how gene duplication and transposable element domestication can also drive GRN evolution.

strong phenotypic effects, where they are both necessary and sufficient for determining a trait of interest. For example, changes to any single known gene downstream of the input-output gene *shavenbaby* (*svb*) are not sufficient to promote or inhibit trichome development, but changes to *svb* expression itself are sufficient to alter trichome development [21]. *svb* is also necessary for trichome development [21]. To understand the network context such input-output genes inhabit, and the common properties of these networks, we can draw from research on other biological networks.

The condition of some genes having more interactions than other genes, just as the input-output gene is connected to many more genes than others, has been well-explored in other areas of network biology. Many networks in biology are considered using the graph model of the scale-free network [22]. These networks are composed of nodes (in this case, genes) and edges (regulatory connections between genes). A few nodes are connected to many other nodes ('hubs'), while most nodes have few interactions. This distribution of connections can be described by the power law function.

Recent work from Ouma et al. [23] using GRNs derived from protein-DNA interaction databases across four organisms found that these global GRNs fit the scale-free model. They found that most transcription factors only interacted with a few genes, while only a few transcription factors interacted with many genes, following the predicted power law distribution with different scaling exponents for different species. While they found that

subnetworks of these GRNs also fit the scale-free model, it remains to be tested whether specific developmental GRNs are truly scale-free [24].

This general principle of a few genes with many connections and many genes with few connections will likely hold true. Research on PPI networks can help us assess this prediction and its implications. There are typically a few high-connectivity proteins and many low-connectivity proteins in a network, with connectivity defined as the number of interactions per protein. These few high connectivity proteins are more likely to interact with low connectivity proteins and less likely to interact with each other than expected by chance, forming networks that have many peripheral interacting genes and a few central genes with many interactions [25]. Networks with this asymmetric distribution of connectivity are generally highly robust to random errors but are extremely vulnerable to the removal of the high connectivity nodes [26]. Consistent with this predicted robustness, evolved protein interaction networks are more resilient to the removal of random nodes than randomized networks [27].

One network structure that can account for this variation in connectivity is the bow-tie structure. A bow-tie refers to a structure where there are two layers composed of many nodes and an intermediate layer that is composed of very few nodes that connects these two layers [28]. This central layer forms the core or ‘knot’ of the bow-tie (Fig. 1). The nodes at the core of the bow-tie have the highest number of connections [10]. Many types of networks, including metabolic and signaling pathways, can be characterized by this bow-tie structure [28]. Bow-ties are thought to be common across

biological systems because they facilitate both robustness and evolvability of the system [10].

A directed bow-tie structure is composed of many inputs which are integrated by the few nodes at the central core. These core nodes then regulate many outputs. This concept can also be applied to developmental GRNs, where many upstream genes are inputs to the input-output gene(s), which then targets many downstream genes to regulate cellular differentiation [6,7]. Bow-tie networks can be distinguished from the hierarchical null model by demonstrating that a gene (or genes) is connected to more genes both upstream and downstream than others [28].

3. The evolution of GRNs for rapidly-evolving morphological traits

3.1 Two case studies

The *svb* GRN fits the bow-tie architecture [7]. This GRN controlling larval trichome pattern in *Drosophila* is composed of many upstream gene inputs, an input-output gene (*svb*), and many output genes[2]. Evolutionary divergence at the CREs controlling expression of *svb* has repeatedly driven morphological change [7]. The higher substitution rate in the *svb* regulatory region compared to neighboring regions indicates

[2] This structure is also commonly described as an hourglass-shaped network [7]. Here we refer to it as a bow-tie structure to connect this concept from GRN studies with the literature on other types of biological networks [28] and to avoid confusion with the developmental hourglass model [29].

that it is the target of positive selection or is under relaxed constraint [30]. These CRE mutations have resulted in parallel losses of trichomes in multiple *Drosophila* species [2]. Thus, *svb* is considered a hotspot gene for morphological evolution.

Another example of a hotspot gene for morphological evolution is *optix,* which is a proposed input-output gene for wing patterning across butterflies [31]. There are many known downstream genes of *optix*, as well as many candidate upstream genes [31–34]. Therefore, the *optix* butterfly wing color pattern GRN most likely fits the bow-tie structure (Fig. 2). The adaptive convergent evolution of red wing color pattern mimicry in *Heliconius* butterflies is due to selection on *optix* CREs [4,35,36]. We have evidence that GRNs for rapidly-evolving morphological traits are evolving primarily by positive selection acting on the CREs of the input-output genes from the *svb* and *optix* networks. We still have little information, however, on how positive selection acts on the broader networks that host these genes.

A recent study on the *optix* GRN was able to shed some light on the distribution of selection throughout downstream elements of the network. Lewis et al. [34] combined methods to detect selective sweeps with molecular approaches to characterize genes regulated by *optix* in order to identify genes under selection in the *optix* GRN that may also be involved in adaptive wing pattern evolution. By identifying binding sites of the optix protein, and then determining which genes optix-bound CREs were regulating, they were able to identify numerous direct targets of optix. Notably, optix-bound CREs showed significantly elevated signals of selection compared to randomly-selected

CREs, although, interestingly, few of these genes showed nearly as great a signal of selection as *optix* itself. This suggests that these directly downstream genes are targets of positive selection but are less strongly selected upon than the regulatory region of the input-output gene itself.

3.2 GRN structure and the strength of positive selection

Using the *optix* GRN as a case study (Fig. 2), we can predict how different levels within a GRN for a rapidly-evolving adaptive trait may be more or less likely to be targets of positive selection.

3.2.1 Key predictions for the evolution of different levels of GRNs

Prediction I: CREs of input-output genes are more likely to be under strong positive selection than CREs of other genes in a GRN, while input-output gene protein-coding regions are more likely to be constrained.

An important prediction in modern evo-devo is that CRE sequences should drive trait evolution more frequently than coding regions because they make up a much larger percentage of the genome, and expected to have more trait-specific (and less pleiotropic) effects on phenotypes [8,20]. Following this, we would further predict that input-output gene CREs are more likely to be under positive selection than genes at

other positions in a GRN because the handful of input-output gene case studies, such as *optix*, show these loci can have strong signatures of selection and population structure compared to the rest of the genome.

Conversely, the protein-coding regions for input-output genes may be more constrained due to these transcription factors' involvement in other more ancestral developmental processes. For example, *svb* is required for the production of all trichomes of *Drosophila* larvae and adults, and an isoform of *svb* is required for oogenesis [6,37], and *optix* is known to be essential for eye morphogenesis in *Drosophila* and may have been co-opted to regulate red color pattern in butterflies relatively recently [31,38]. There is considerable study on how proteins with a higher number of interaction partners are more constrained and more likely to be under negative selection [13,39,40]. In contrast, the idea that the protein-coding sequence for a gene connected to more genes through *cis*- interactions is more constrained is, to our knowledge, largely untested. One study investigated this question by measuring natural variation in gene expression level in the plant *Capsella grandiflora* to infer gene co-expression networks [16]. This study determined gene connectivity by measuring the sum of correlations with other genes, weighted by the strengths of correlations. The genes with higher connectivity scores were more likely to be under negative selection, but the level of gene connectivity had no detectable correlation with rate of fixations driven by positive selection. However, interpretation of this result is limited by the fact that it is based on networks inferred from gene expression and not functionally validated regulatory relationships.

Prediction II: Input gene protein-coding sequences are more likely to be under stronger stabilizing selection due to pleiotropy than those of output genes.

Proteins that are on the periphery of a PPI network, with the fewest interaction partners, are more likely to be targets of positive selection [13,14]. We may predict a similar pattern for genes with fewer connections to other genes in a network. Similar to input-output genes, upstream transcription factors are more likely to be involved in essential developmental processes and to be more constrained than peripheral genes [8]. We may expect an increase in pleiotropy in a protein's function to correlate with an increase in constraint on the amino acid sequence. Likewise, this constraint does not necessarily extend to the CREs of these genes [41].

This pleiotropy may also potentiate adaptive evolution in other ways. More pleiotropic proteins could have more binding domains and more opportunities to interact with new partners, so we might expect stronger positive selection on their regulation than less pleiotropic proteins, although this prediction has not been tested. In other cases, we might expect more pleiotropic genes to be regulated by more pleiotropic CREs, so the evolution of these CREs may or may not also be constrained [8,42]. Therefore, unlike for protein-coding sequences, it is difficult to predict whether there is a difference in selection strength on upstream vs. downstream gene CREs. Future work on cis-regulatory grammar and interaction dynamics will help resolve this [43,44].

Prediction III: Traits evolving rapidly under positive selection are controlled by more fragile GRNs.

For a robust GRN, mutations and genetic variation will generate less phenotypic variation that can be subject to selection. By contrast, we might expect traits that are rapidly-evolving under positive selection to be controlled by more fragile networks. In this case, fragility meaning that minor mutations, such as in individual CREs, are likely to have substantive phenotypic effects [35]. A trade-off between robustness and innovation has been predicted on short time scales, and recent empirical work shows GRNs for rapidly-evolving adaptive traits are more fragile than previously thought [35,45]. However, the extent of this trade-off is still an active area of investigation. Robustness can also increase later opportunities for selection on a GRN in the long-term [45].

Prediction IV: Evolutionary drift is likely to be more prominent than positive selection in robust GRNs.

Often, networks that are observed to be under developmental systems drift (DSD) – the process by which homologous traits diverge in their genetic mechanism via neutral evolution – are thought to be more robust [46]. This is because this drift suggests that there is some level of functional redundancy among nodes in the network. DSD can occur at different positions in a network. Nahmad et al. [47] found that neutral evolution in the regulation of genes at different positions in the GRN that controls ant wing

polyphenism can result in similar effects in wing size. It is still unclear, however, if there is any predictability in how robustness and redundancy are distributed across different aspects of GRNs. Robustness can be an emergent property under long periods of stabilizing selection or it can be selected for when there are many perturbations to a trait [48,49]. Whatever the origin of robustness may be, we would expect GRNs for older homologous traits and early developmental stages to be more robust than younger and later-acting GRNs. This idea is supported by gene expression and modeling data comparing early and late networks [50]. Furthermore, older traits also simply have had longer to evolve robustness, and therefore, by extension, we would expect DSD to occur more often in older GRNs.

3.2.2 Models for strength of positive selection across a GRN

Given the predictions above, we can construct several models of positive selection pressure across a GRN. These models are neither comprehensive nor mutually exclusive, but they provide several testable hypotheses for how network positionality can affect the rate of fixation driven by positive selection in genes and CREs at different levels of the network.

Based on the case studies of *svb* and *optix,* all models for cis-regulatory evolution assume strong positive selection at the input-output genes [30,35,36]. Beyond this, one possible model is that whether a gene is upstream or downstream of the input-output

gene has little effect on the rate and strength of positive selection on that gene's CREs (Fig. 3a). Changes in both upstream and downstream gene CREs may result in expression in a new spatiotemporal domain and changes in the trait. For one well-studied trait – abdominal pigmentation in *Drosophila* – it appears there is change occurring both upstream and downstream in the network that can explain pigmentation variation within and between species [51]. These genes also show some evidence of selection [52,53]. It is challenging, however, to differentiate which genes may truly be input-output genes until the network is better characterized. Abdominal pigmentation is an excellent target for future work given the many genes associated with variation in this trait that can be evaluated further to compare the frequency of selective sweeps on different types of genes [54].

Another model is that CREs of upstream genes are under pleiotropic constraint while CREs of downstream genes are under positive selection (Fig. 3b). This model may be more likely if the input genes' CREs are all also shared (possibly through co-option) as part of more ancient, essential GRNs [8,55]. We may expect that input genes are more likely to be involved in network kernels that have dense circuitry. These input genes' CREs are thus more likely to be constrained, such as by requiring a precise order of cooperatively-binding transcription factors to activate an essential function [55]. How widespread this type of constraint is on the regulation of upstream genes is unclear. Some ancient CREs have been found to drive adaptive trait evolution, and some upstream genes with constrained CREs can also gain new, possibly more evolutionarily

317 labile CREs [35,56]. More research is needed to determine whether the regulation of
318 upstream genes is more often constrained than downstream genes.
319
320 A third alternative model posits that some traits may be evolving rapidly, primarily by
321 changes in upstream patterning, so the CREs of upstream genes may be under positive
322 selection while the regulatory architecture of the downstream genes is functionally
323 conserved (Fig. 3c). We expect this to occur in cases where a GRN was co-opted to
324 reproduce a structure at a new location or timepoint. For example, the development of
325 the novel adult male-specific posterior lobe in *Drosophila melanogaster* is driven by a
326 GRN co-opted from the development of the larval posterior spiracle. This co-opted GRN
327 shares many of its downstream genes and enhancers with its ancestral GRN [57]. The
328 origin of the novel trait is most likely due to changes in upstream patterning.
329 Downstream terminal effectors may also be highly conserved such that upstream genes
330 are evolving more by contrast. An interesting observation consistent with this model
331 comes out of the many studies of adaptive wing patterning evolution in Lepidoptera,
332 where selection on a pigmentation gene has never been found to be the primary driver
333 of wing color pattern evolution in nature, even for simple color switches [4,58,59].
334
335 In terms of selection on protein sequences, downstream proteins may be the least
336 constrained and most likely to be under positive selection (Fig. 3d). There are many
337 examples of downstream protein structural changes involved in adaptive evolution of
338 melanism, for example [60–62]. Interestingly, these genes tend to be receptors or
339 signaling proteins in the melanin pathway, not the terminal effectors. It has been

proposed that further downstream genes evolve more slowly because they occupy a more stable cellular environment [63]. These cases suggest that the downstream proteins for this melanic trait are generally much more evolutionarily labile than upstream transcription factors. This is consistent with the research on PPIs that proteins on the periphery of a network should be under the strongest positive selection compared to other proteins, but we need more comprehensive and comparative studies to determine whether selection is indeed mainly targeting coding regions of these downstream genes [13].

We could also observe positive selection on an upstream gene or genes (Fig. 3e). While we would expect upstream transcription factors to be more evolutionarily constrained due to pleiotropy, there could be positive selection for transcription factor modularity by evolving additional DNA binding or protein binding domains [64]. We might expect this for younger transcription factors that do not have many essential roles and are less constrained in their structure. This upstream protein evolution could also occur after a gene duplication event, which could release this gene from constraint and allow for the duplicate gene to diverge and gain a new role in regulating the input-output gene or other upstream genes [65].

It is also worth considering that a cofactor for the input-output gene could be under positive selection to interact with the input-output gene and activate different suites of genes (Fig. 3f). Cofactors can increase the capacity for the network core to activate modules of differentiation genes in specific spatial contexts and are critical for the

development of specific tissues and cell types [66]. We also expect core proteins to be evolutionarily constrained because changes to their binding domains would affect many processes at once. However, the less conserved regions of the protein structure can evolve more easily and allow new protein-protein interactions. This can avoid the potential pleiotropic costs of changes to the binding domains themselves [67].

4. How generalizable are these predictions across developmental GRNs?

Our predictions – and much of our understanding of GRNs – come from study of the development of rapidly-evolving, adaptive morphological traits. However, GRNs can be considered at many spatial and temporal scales, from the set of genes that underlies an entire developmental stage to the set of genes responsible for a specific discrete trait. Whether our predictions can be applied across developmental GRNs is unclear. There are some cases where GRNs are not under positive selection. These may include highly-conserved, essential GRNs [5]. There are also some specific developmental stages where the networks are much more constrained given the high degree of conservation across taxa, such as the genes underlying the midembryogenesis period of development [8,29,68].

Further, we assume that the networks controlling the development of these morphological traits fit a bow-tie structure, with a distinct input-output gene or genes that are much more connected to other genes than these other genes are connected to each

other in the network. This assumption has not been rigorously tested. With more research on gene regulatory relationships, we can better model the structures of GRNs and how these structures can vary. Perhaps, for example, bow-tie GRNs are more commonly seen as a feature of more rapidly-evolving traits (e.g. color patterns), while more deeply conserved traits (e.g. embryonic patterning) tend towards different structures. Presently, however, we cannot say how generalizable these ideas are beyond that they are almost certainly not universally applicable – there are simply too few case studies.

5. Experimental methods for GRN evolution

5.1 Inferring networks and patterns of selection

To test whether the proposed models (or, more likely, combinations of models) of positive selection across GRNs hold for adaptive morphological traits, and whether these patterns are found more broadly across developmental GRNs, we need two types of information. We need first to characterize the GRN for traits of interest, and then we need to determine the patterns of selection across the genome. Experimental methods for the latter have been well-developed: We know that the selection across genomes is not evenly distributed, and many studies have extensively investigated individual loci that show strong signals of selection and are involved in morphological evolution [7,69]. Our knowledge of developmental GRNs is comparatively lacking. Most GRNs are

409 inferred from co-expression correlation matrices generated from bulk RNA-seq data.
410 While these data can be very informative, the actual regulatory relationships between
411 genes remain unknown [70]. Here, we discuss first how these networks can be
412 described in more detail, and then how these data can be integrated with tests for
413 selective sweeps to relate network position to gene evolution.
414
415 **5.1.1 Characterizing GRNs**
416
417 There are many tools that can help improve our understanding of regulatory interactions
418 and confirm causality between interactions. One of the most critical pieces of
419 information is to understand where key transcription factors are binding in the genome,
420 and to infer their target genes. Analyzing transcription factor genes that have been
421 associated with trait evolution by using chromatin immunoprecipitation and sequencing
422 (ChIP-seq), or similar methods, is a key step in characterizing GRNs [71]. For binding
423 sites that are not located at the promoter of a gene, the target gene can be identified
424 using chromosome conformation capture methods (e.g., Hi-C, 4C, etc.) to determine
425 whether the bound DNA region physically interacts with the promoter of a gene [72].
426 These inferences can be further supported using gene expression data [72,73]. Many
427 methods have been developed for network inference from single cell RNA sequencing
428 (scRNA-seq) data that leverage analysis across cell types and timepoints [74]. scRNA-
429 seq data can also be integrated with analysis of chromatin accessibility [72]. For
430 humans, yeast, and other organisms with large amounts of pre-existing molecular data,
431 GRNs can be predicted by integrating known protein-protein interactions, gene

expression, and binding motif data [75,76]. These data can further expand our knowledge of upstream and downstream genes in the network that can be later confirmed using functional tests.

Functional tests of candidate genes can confirm not only that the gene is involved in the trait of interest, but also the direction of regulation. We can knock out, knock down, or drive expression of a key transcription factor and assay for changes in the expression of candidate downstream genes. Alternatively, we can use genetic tools to manipulate the expression of multiple genes in a hypothesized network to test whether they are in the same network and to determine the relative position of these genes. Reporter constructs can also assist in validating the role of particular CREs in driving expression in a particular region. CRISPR/Cas9 technology has made all of these approaches much more accessible in emerging model systems [77].

While inferring GRNs requires a lot of experiments, some of this work has already been completed in a handful of study systems. We suggest that GRNs that have been studied in depth in various model systems are ripe to be used in comparative evolutionary studies by extending work into related species. Comparative analysis of these GRNs could then shed light on the patterns in evolution across different levels of the network. For example, comparative work on neural crest cell development in other vertebrates in addition to chicks has illuminated the evolution of the cranial neural crest by successive additions of components to the network from an ancestral trunk-like lineage [78]. Another recent study compared the well-characterized sea urchin endomesoderm GRN

455 with a newly constructed sea star GRN for the same trait, finding both shared and
456 unique modules [79]. Thus, there are quite a few promising systems for exploring GRN
457 evolution.
458
459 **5.1.2 Detecting positive selection**
460
461 It will be exciting to combine functional GRN models with tests for signals of positive
462 selection. There are a number of methods to detect positive selection using variation
463 within and between species [80–82]. Since selection can be tested at both micro- and
464 macroevolutionary scales, we can also compare the patterns of selection across
465 networks that may emerge at different time scales. Testing for positive selection can
466 also be useful for building the GRN for a particular trait since regions under selection
467 will have some functional role in a phenotype.
468
469 Many tests for selection on genes are based on the ratio of the rate of nonsynonymous
470 substitutions to the rate of synonymous substitutions (d_N/d_S). There is no equivalent to
471 this statistic for CREs. Positive selection in CREs has been identified using tests for
472 selective sweeps and divergence in substitution rate in specific regions across taxa
473 [36,83,84]. In principle, future work on CRE evolution could also leverage analysis of
474 motif composition in a similar way to synonymous and nonsynonymous changes to
475 genes. These tests would require a sophisticated understanding of what affects a
476 motif's affinity for specific transcription factors and how transcription factors' binding
477 sites differ from their canonical motifs in different taxa. Despite these complications, it is

478 worthwhile to analyze the motifs of a CRE in the event that transcription factor binding is
479 conserved despite sequence divergence. These functionally-conserved CREs have
480 been identified at deep evolutionary time scales [85]. Understanding what changes to
481 CREs are meaningful and are more likely to be the result of positive selection and what
482 changes are due to drift can be aided by characterizing motifs.
483
484 **5.2 Limitations and challenges**
485
486 There are several common limitations and biases to studies of the type mentioned
487 above. The main challenge moving forward will be scaling up experiments to sufficiently
488 characterize a GRN, or many GRNs, to answer questions of network position and
489 selection. Choosing a few transcription factors that are well-described and known to be
490 under strong selection can help focus this research, but it also introduces bias in the
491 description of the network's structure. This streetlight effect is unavoidable unless we
492 endeavor to describe every unknown gene that is associated with a trait.
493
494 Necessarily, any description of a network for a specific character involves decisions of
495 what is and is not included as part of the network. No GRN is an island: The
496 development of a late-acting GRN for a trait will often be contingent upon proper early
497 development of the organism. How we should make these decisions of what is and is
498 not considered part of a trait's underlying network is an open question. Some suggest
499 that every gene expressed in the cells that give rise to a trait should be considered part
500 of the GRN for that trait – a viewpoint growing in popularity with respect to disease

states [86]. Most evo-devo studies include genes in the network for a trait if they have functional or other molecular evidence to support its inclusion. More data on the gene regulatory networks underlying traits will help us understand how best to characterize them and whether the bow-tie model fits or if a different structure is more representative.

It is also important to consider that often a gene may be located at different network positions depending on the trait or network scale considered. For example, different strains of *Drosophila melanogaster* have different patterns of trichomes on the legs. Initially, it seemed surprising that these differences were not facilitated by changes in CREs regulating *svb* expression, as was found for larval trichome pattern. Instead, differences in leg trichomes were mediated by changes to the CREs of a different gene, *miR-92a.* This finding could be explained by differences between the larval and leg trichome GRNs [87]. Thus, the selective pressure on any individual gene or CRE can be affected by its different network positions and roles for different traits. Further, even within the network for the same trait, a gene can also play multiple roles and occupy different network positions, such as both regulating (upstream of) and being regulated by (downstream of) the input-output gene.

Generally, to identify genes that underlie adaptive morphological evolution, they must meet two conditions: i) they have detectable effects on phenotype and ii) they have detectable signatures of selection. The literature reviewed in this paper is thus biased to focus on large- and intermediate-effect size genes with evidence of recent divergence.

These examples demonstrate that large- and intermediate-effect genes do in fact drive adaptation, as can be predicted under some evolutionary scenarios [88]. However, these data are likely not representative of the entire spectrum of genetic variation underlying trait evolution including all minor effect genes, especially for complex developmental traits [89]. More research aimed at detecting polygenic selection across networks can reveal whether gene network position is less important in this evolutionary regime [82].

Finally, complete knowledge of every GRN and every gene's regulation and function is still probably not sufficient to predict gene evolutionary rates at different network positions due to the potential effects of population size and structure [90]. In small populations, mutations that have a larger effect on the network structure may be more likely to be fixed, whereas in larger populations, we might expect this to occur less often. This is because small populations tend to accumulate deleterious mutations, and a mutation that significantly changes gene interactions is more likely to be deleterious compared to a mutation that slightly alters expression of a downstream gene [90].

6. Conclusion

Characterizing GRNs and patterns of selection across them is clearly not a small task, but it can lend great insight into the evolution of adaptive traits. Positionality within a network has long been proposed as an important factor in the evolution of genes within a regulatory network, and many studies have tested for similar patterns of selection

across different components of signaling and metabolic pathways [28]. Due to the paucity of thoroughly characterized developmental GRNs, especially for rapidly adapting traits, this question has still not been addressed. Open questions include whether evolution at the CREs of input-output genes is the primary driver of morphological evolution and whether there are common patterns in how selection varies across GRNs. We are well-positioned with molecular techniques available today to address these network-related gene evolution questions. As GRNs are the bridge between genotype and phenotype, the better we can understand regulatory networks, the better we can understand the mechanisms of adaptation.

Acknowledgments

The authors would like to thank Omid Saleh Ziabari, Sonia Messar, and members of the Reed lab for helpful comments and discussion. We are also grateful to Adam Wilkins and one anonymous reviewer for their thoughtful feedback. This work was supported by NSF GRFP DGE-1650441 to JMCM and NSF grant IOS-2128164 to RDR.

References

[1] A.S. Wilkins, Between “design” and “bricolage”: Genetic networks, levels of selection, and adaptive evolution, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 8590–8596.

[2] E. Sucena, I. Delon, I. Jones, F. Payre, D.L. Stern, Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism, Nature. 424 (2003) 935–938.

[3] M.D. Shapiro, M.A. Bell, D.M. Kingsley, Parallel genetic origins of pelvic reduction in vertebrates, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 13753–13758.

[4] R.D. Reed, R. Papa, A. Martin, H.M. Hines, B.A. Counterman, C. Pardo-Diaz, C.D. Jiggins, N.L. Chamberlain, M.R. Kronforst, R. Chen, G. Halder, H.F. Nijhout, W.O. McMillan, optix drives the repeated convergent evolution of butterfly wing pattern mimicry, Science. 333 (2011) 1137–1141.

[5] E.H. Davidson, D.H. Erwin, Gene regulatory networks and the evolution of animal body plans, Science. 311 (2006) 796–800.

[6] D.L. Stern, V. Orgogozo, Is Genetic Evolution Predictable?, Science. 323 (2009) 746–751. https://doi.org/10.1126/science.1158997.

[7] D.L. Stern, The genetic causes of convergent evolution, Nat. Rev. Genet. 14 (2013) 751–764.

[8] E.H. Davidson, I. Peter, Genomic Control Process, Academic Press, 2015.

[9] R.D.H. Barrett, D. Schluter, Adaptation from standing genetic variation, Trends Ecol. Evol. 23 (2008) 38–44.

[10] H. Kitano, Biological robustness, Nat. Rev. Genet. 5 (2004) 826–837.

[11] S. Cussat-Blanc, K. Harrington, W. Banzhaf, Artificial gene regulatory networks-A review, Artif. Life. 24 (2018) 296–328.

[12] B. Lehne, T. Schlitt, Protein-protein interaction databases: keeping up with growing interactomes, Hum. Genomics. 3 (2009) 291–297.

[13] P.M. Kim, J.O. Korbel, M.B. Gerstein, Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 20274–20279.

[14] H.B. Fraser, A.E. Hirsh, L.M. Steinmetz, C. Scharfe, M.W. Feldman, Evolutionary rate in the protein interaction network, Science. 296 (2002) 750–752.

[15] I.S. Peter, E.H. Davidson, The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage, Dev. Biol. 340 (2010) 188–199.

[16] E.B. Josephs, S.I. Wright, J.R. Stinchcombe, D.J. Schoen, The Relationship between Selection, Network Connectivity, and Regulatory Variation within a Population of Capsella grandiflora, Genome Biol. Evol. 9 (2017) 1099–1109.

[17] K. Voordeckers, K. Pougach, K.J. Verstrepen, How do regulatory networks evolve and expand throughout evolution?, Curr. Opin. Biotechnol. 34 (2015) 180–188.

[18] C. Feschotte, Transposable elements and the evolution of regulatory networks, Nat. Rev. Genet. 9 (2008) 397–405.

[19] E. McQueen, M. Rebeiz, On the specificity of gene regulatory networks: How does network co-option affect subsequent evolution?, Curr. Top. Dev. Biol. 139 (2020) 375–405.

[20] D.L. Stern, V. Orgogozo, The loci of evolution: how predictable is genetic evolution?, Evolution. 62 (2008) 2155–2177.

[21] A.P. McGregor, V. Orgogozo, I. Delon, J. Zanet, D.G. Srinivasan, F. Payre, D.L. Stern, Morphological evolution through multiple cis-regulatory mutations at a single gene, Nature. 448 (2007) 587–590.

[22] R. Albert, Scale-free networks in cell biology, J. Cell Sci. 118 (2005) 4947–4957.

[23] W.Z. Ouma, K. Pogacar, E. Grotewold, Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties, PLoS Comput. Biol. 14 (2018) e1006098.

[24] A.D. Broido, A. Clauset, Scale-free networks are rare, Nat. Commun. 10 (2019) 1017.

[25] S. Maslov, K. Sneppen, Specificity and stability in topology of protein networks, Science. 296 (2002) 910–913.

[26] R. Albert, H. Jeong, A.L. Barabasi, Error and attack tolerance of complex networks, Nature. 406 (2000) 378–382.

[27] R. Maddamsetti, Selection maintains protein interactome resilience in the long-term evolution experiment with Escherichia coli, Genome Biol. Evol. (2021). https://doi.org/10.1093/gbe/evab074.

[28] T. Friedlander, A.E. Mayo, T. Tlusty, U. Alon, Evolution of bow-tie architectures in biology, PLoS Comput. Biol. 11 (2015) e1004055.

[29] N. Irie, S. Kuratani, The developmental hourglass model: a predictor of the basic body plan?, Development. 141 (2014) 4649–4655.

[30] N. Frankel, D.F. Erezyilmaz, A.P. McGregor, S. Wang, F. Payre, D.L. Stern, Morphological evolution caused by many subtle-effect substitutions in regulatory DNA, Nature. 474 (2011) 598–603.

[31] L. Zhang, A. Mazo-Vargas, R.D. Reed, Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 10707–10712.

[32] H.M. Hines, R. Papa, M. Ruiz, A. Papanicolaou, C. Wang, H.F. Nijhout, W.O. McMillan, R.D. Reed, Transcriptome analysis reveals novel patterning and pigmentation genes underlying Heliconius butterfly wing pattern variation, BMC Genomics. 13 (2012) 288.

[33] J.J. Hanly, R.W.R. Wallbank, W.O. McMillan, C.D. Jiggins, Conservation and flexibility in the gene regulatory landscape of heliconiine butterfly wings, Evodevo. 10 (2019) 15.

[34] J.J. Lewis, S.M. Van Belleghem, R. Papa, C.G. Danko, R.D. Reed, Many functionally connected loci foster adaptive diversification along a neotropical hybrid zone, Sci Adv. 6 (2020). https://doi.org/10.1126/sciadv.abb8617.

[35] J.J. Lewis, R.C. Geltman, P.C. Pollak, K.E. Rondem, S.M. Van Belleghem, M.J. Hubisz, P.R. Munn, L. Zhang, C. Benson, A. Mazo-Vargas, C.G. Danko, B.A. Counterman, R. Papa, R.D. Reed, Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry, Proc. Natl. Acad. Sci. U. S. A. 116 (2019) 24174–24183.

[36] M. Moest, S.M. Van Belleghem, J.E. James, C. Salazar, S.H. Martin, S.L. Barker, G.R.P. Moreira, C. Mérot, M. Joron, N.J. Nadeau, F.M. Steiner, C.D. Jiggins, Selective sweeps on novel and introgressed variation shape mimicry loci in a butterfly adaptive radiation, PLoS Biol. 18 (2020) e3000597.

[37] C. Salles, M. Mével-Ninio, A. Vincent, F. Payre, A germline-specific splicing generates an extended ovo protein isoform required for Drosophila oogenesis, Dev. Biol. 246 (2002) 366–376.

[38] M. Seimiya, W.J. Gehring, The Drosophila homeobox gene optix is capable of inducing ectopic eyes by an eyeless-independent mechanism, Development. 127 (2000) 1879–1886.

[39] D.M. Krylov, Y.I. Wolf, I.B. Rogozin, E.V. Koonin, Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution, Genome Res. 13 (2003) 2229–2235.

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

[41] D. Molodtsova, B.A. Harpur, C.F. Kent, K. Seevananthan, A. Zayed, Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors, Front. Genet. 5 (2014) 431.

[42] G. Sabarís, I. Laiker, E. Preger-Ben Noon, N. Frankel, Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity, Trends Genet. 35 (2019) 423–433.

[43] G.A. Jindal, E.K. Farley, Enhancer grammar in development, evolution, and disease: dependencies and interplay, Dev. Cell. 56 (2021) 575–587.

[44] B. Lim, M.S. Levine, Enhancer-promoter communication: hubs or loops?, Curr. Opin. Genet. Dev. 67 (2021) 5–9.

[45] S. Ciliberti, O.C. Martin, A. Wagner, Innovation and robustness in complex regulatory gene networks, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 13591–13596.

679 [46] C.K. Ewe, Y.N. Torres Cleuren, J.H. Rothman, Evolution and developmental
680 system drift in the endoderm gene regulatory network of Caenorhabditis and other
681 nematodes, Front. Cell Dev. Biol. 8 (2020) 170.
682 [47] M. Nahmad, L. Glass, E. Abouheif, The dynamics of developmental system drift
683 in the gene network underlying wing polyphenism in ants: a mathematical model, Evol.
684 Dev. 10 (2008) 360–374.
685 [48] A. Wagner, Does evolutionary plasticity evolve?, Evolution. 50 (1996) 1008.
686 [49] M.-A. Félix, A. Wagner, Robustness and evolution: concepts, insights and
687 challenges from a developmental model system, Heredity (Edinb.). 100 (2008) 132–
688 140.
689 [50] K.E. Sears, J.A. Maier, M. Rivas-Astroza, R. Poe, S. Zhong, K. Kosog, J.D.
690 Marcot, R.R. Behringer, C.J. Cretekos, J.J. Rasweiler 4th, Z. Rapti, The relationship
691 between gene network structure and expression variation among individuals and
692 species, PLoS Genet. 11 (2015) e1005398.
693 [51] J.H. Massey, P.J. Wittkopp, The Genetic Basis of Pigmentation Differences
694 Within and Between Drosophila Species, Genes and Evolution. (2016) 27–61.
695 https://doi.org/10.1016/bs.ctdb.2016.03.004.
696 [52] J.E. Pool, C.F. Aquadro, The genetic basis of adaptive pigmentation variation in
697 Drosophila melanogaster, Molecular Ecology. 16 (2007) 2844–2851.
698 https://doi.org/10.1111/j.1365-294x.2007.03324.x.
699 [53] S. Jeong, M. Rebeiz, P. Andolfatto, T. Werner, J. True, S.B. Carroll, The
700 evolution of gene regulation underlies a morphological difference between two
701 Drosophila sister species, Cell. 132 (2008) 783–793.

[54] L.M. Dembeck, W. Huang, M.M. Magwire, F. Lawrence, R.F. Lyman, T.F.C. Mackay, Genetic Architecture of Abdominal Pigmentation in Drosophila melanogaster, PLoS Genet. 11 (2015) e1005163.

[55] M. Rebeiz, N.H. Patel, V.F. Hinman, Unraveling the tangled skein: The evolution of transcriptional regulatory networks in development, Annu. Rev. Genomics Hum. Genet. 16 (2015) 103–131.

[56] A.C. Thompson, T.D. Capellini, C.A. Guenther, Y.F. Chan, C.R. Infante, D.B. Menke, D.M. Kingsley, A novel enhancer near the Pitx1 gene influences development and evolution of pelvic appendages in vertebrates, Elife. 7 (2018). https://doi.org/10.7554/eLife.38555.

[57] W.J. Glassford, W.C. Johnson, N.R. Dall, S.J. Smith, Y. Liu, W. Boll, M. Noll, M. Rebeiz, Co-option of an Ancestral Hox-Regulated Network Underlies a Recently Evolved Morphological Novelty, Developmental Cell. 34 (2015) 520–531. https://doi.org/10.1016/j.devcel.2015.08.005.

[58] E.L. Westerman, N.W. VanKuren, D. Massardo, A. Tenger-Trolander, W. Zhang, R.I. Hill, M. Perry, E. Bayala, K. Barr, N. Chamberlain, T.E. Douglas, N. Buerkle, S.E. Palmer, M.R. Kronforst, Aristaless Controls Butterfly Wing Color Variation Used in Mimicry and Mate Choice, Curr. Biol. 28 (2018) 3469-3474.e4.

[59] K. Tunström, A. Woronik, J.J. Hanly, P. Rastas, A. Chichvarkhin, A.D. Warren, A. Kawahara, S.D. Schoville, V. Ficarrotta, A.H. Porter, W.B. Watt, A. Martin, C.W. Wheat, A complex interplay between balancing selection and introgression maintains a genus-wide alternative life history strategy, BioRxiv. (2021) 2021.05.20.445023. https://doi.org/10.1101/2021.05.20.445023.

[60] M.W. Nachman, H.E. Hoekstra, S.L. D'Agostino, The genetic basis of adaptive melanism in pocket mice, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 5268–5273.

[61] J.A.C. Uy, E.A. Cooper, S. Cutie, M.R. Concannon, J.W. Poelstra, R.G. Moyle, C.E. Filardi, Mutations in different pigmentation genes are associated with parallel melanism in island flycatchers, Proc. Biol. Sci. 283 (2016). https://doi.org/10.1098/rspb.2016.0731.

[62] H.R. McRobie, N.D. Moncrief, N.I. Mundy, Multiple origins of melanism in two species of North American tree squirrel (Sciurus), BMC Evol. Biol. 19 (2019) 140.

[63] K. Julenius, A.G. Pedersen, Protein evolution is faster outside the cell, Mol. Biol. Evol. 23 (2006) 2039–2048.

[64] A.M. Cheatle Jarvela, V.F. Hinman, Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks, Evodevo. 6 (2015) 3.

[65] H. Innan, F. Kondrashov, The evolution of gene duplications: classifying and distinguishing between models, Nat. Rev. Genet. 11 (2010) 97–108.

[66] B.M. Spiegelman, R. Heinrich, Biological control through regulated transcriptional coactivators, Cell. 119 (2004) 157–167.

[67] K.J. Brayer, V.J. Lynch, G.P. Wagner, Evolution of a derived protein–protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) E414–E420.

[68] F. Galis, T.J.M. van Dooren, J.A.J. Metz, Conservation of the segmented germband stage: robustness or pleiotropy?, Trends Genet. 18 (2002) 504–509.

[69] C.W. Ahrens, P.D. Rymer, A. Stow, J. Bragg, S. Dillon, K.D.L. Umbers, R.Y. Dudaniec, The search for loci under selection: trends, biases and progress, Mol. Ecol. 27 (2018) 1342–1356.

[70] R. De Smet, K. Marchal, Advantages and limitations of current network inference methods, Nat. Rev. Microbiol. 8 (2010) 717–729.

[71] A. Barski, K. Zhao, Genomic location analysis by ChIP-Seq, J. Cell. Biochem. 107 (2009) 11–18.

[72] S. Jiang, A. Mortazavi, Integrating ChIP-seq with other functional genomics data, Brief. Funct. Genomics. 17 (2018) 104–115.

[73] C. Angelini, V. Costa, Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems, Front. Cell Dev. Biol. 2 (2014) 51.

[74] H. Nguyen, D. Tran, B. Tran, B. Pehlivan, T. Nguyen, A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data, Brief. Bioinform. 22 (2021). https://doi.org/10.1093/bib/bbaa190.

[75] K. Glass, C. Huttenhower, J. Quackenbush, G.-C. Yuan, Passing messages between biological networks to refine predicted interactions, PLoS One. 8 (2013) e64832.

[76] S. Aibar, C.B. González-Blas, T. Moerman, V.A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z.K. Atak, J. Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering, Nat. Methods. 14 (2017) 1083–1086.

[77] A.F. Gilles, M. Averof, Functional genetics for all: engineered nucleases, CRISPR and the gene editing revolution, Evodevo. 5 (2014) 43.

[78] M.L. Martik, S. Gandhi, B.R. Uy, J.A. Gillis, S.A. Green, M. Simoes-Costa, M.E. Bronner, Evolution of the new head by gradual acquisition of neural crest regulatory circuits, Nature. 574 (2019) 675–678.

[79] G.A. Cary, B.S. McCauley, O. Zueva, J. Pattinato, W. Longabaugh, V.F. Hinman, Systematic comparison of sea urchin and sea star developmental gene regulatory networks explains how novelty is incorporated in early development, Nat. Commun. 11 (2020) 6235.

[80] T.R. Booker, B.C. Jackson, P.D. Keightley, Detecting positive selection in the genome, BMC Biology. 15 (2017). https://doi.org/10.1186/s12915-017-0434-y.

[81] P. Pavlidis, N. Alachiotis, A survey of methods and tools to detect recent and strong positive selection, J. Biol. Res. . 24 (2017) 7.

[82] M. Fagny, F. Austerlitz, Polygenic adaptation: Integrating population genetics and gene regulatory networks, Trends Genet. 37 (2021) 631–638.

[83] M.E. Mathyer, E.A. Brettmann, A.D. Schmidt, Z.A. Goodwin, I.Y. Oh, A.M. Quiggle, E. Tycksen, N. Ramakrishnan, S.J. Matkovich, E. Guttman-Yassky, J.R. Edwards, C. de Guzman Strong, Selective sweep for an enhancer involucrin allele identifies skin barrier adaptation out of Africa, Nat. Commun. 12 (2021). https://doi.org/10.1038/s41467-021-22821-w.

[84] A. Berrio, R. Haygood, G.A. Wray, Identifying branch-specific positive selection throughout the regulatory genome using an appropriate proxy neutral, BMC Genomics. 21 (2020) 359.

792 [85] E.S. Wong, D. Zheng, S.Z. Tan, N.L. Bower, V. Garside, G. Vanwalleghem, F.
793 Gaiti, E. Scott, B.M. Hogan, K. Kikuchi, E. McGlinn, M. Francois, B.M. Degnan, Deep
794 conservation of the enhancer regulatory code in animals, Science. 370 (2020)
795 eaax8137.
796 [86] E.A. Boyle, Y.I. Li, J.K. Pritchard, An expanded view of complex traits: From
797 polygenic to omnigenic, Cell. 169 (2017) 1177–1186.
798 [87] S. Kittelmann, A.D. Buffry, F.A. Franke, I. Almudi, M. Yoth, G. Sabaris, J.P.
799 Couso, M.D.S. Nunes, N. Frankel, J.L. Gómez-Skarmeta, J. Pueyo-Marques, S. Arif,
800 A.P. McGregor, Gene regulatory network architecture in different developmental
801 contexts influences the genetic basis of morphological evolution, PLoS Genet. 14
802 (2018) e1007375.
803 [88] H.A. Orr, The population genetics of adaptation: The distribution of factors fixed
804 during adaptive evolution, Evolution. 52 (1998) 935.
805 [89] M.V. Rockman, The QTN program and the alleles that matter for evolution: all
806 that's gold does not glitter, Evolution. 66 (2012) 1–17.
807 [90] D.L. Stern, Evolution, development, and the predictable genome, Roberts &
808 Company, Greenwood Village, CO, 2010.
809
810

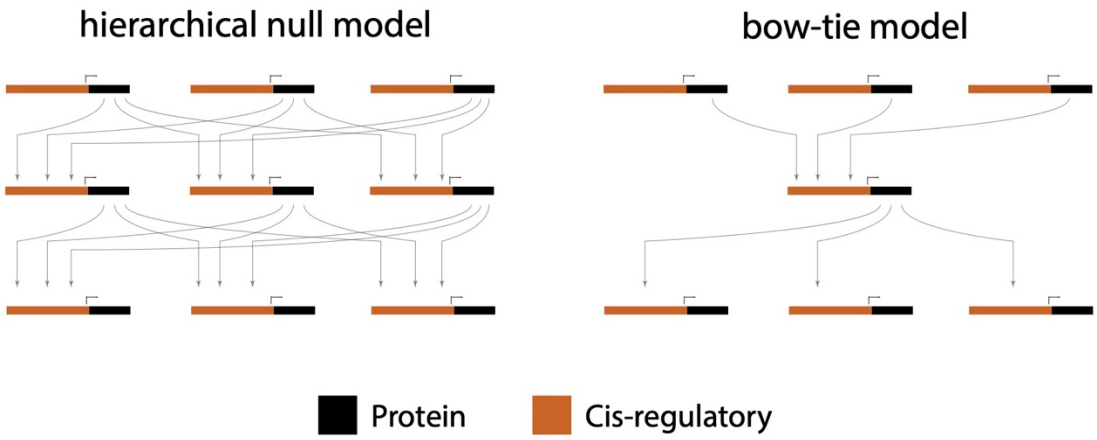
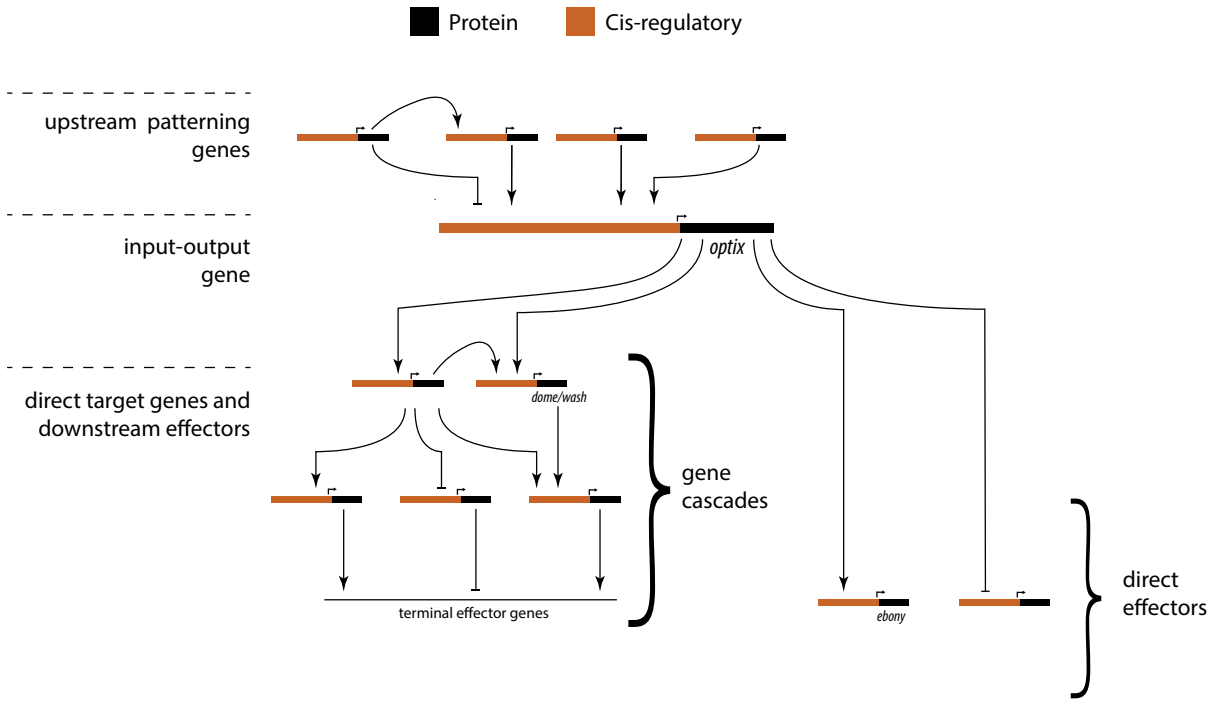


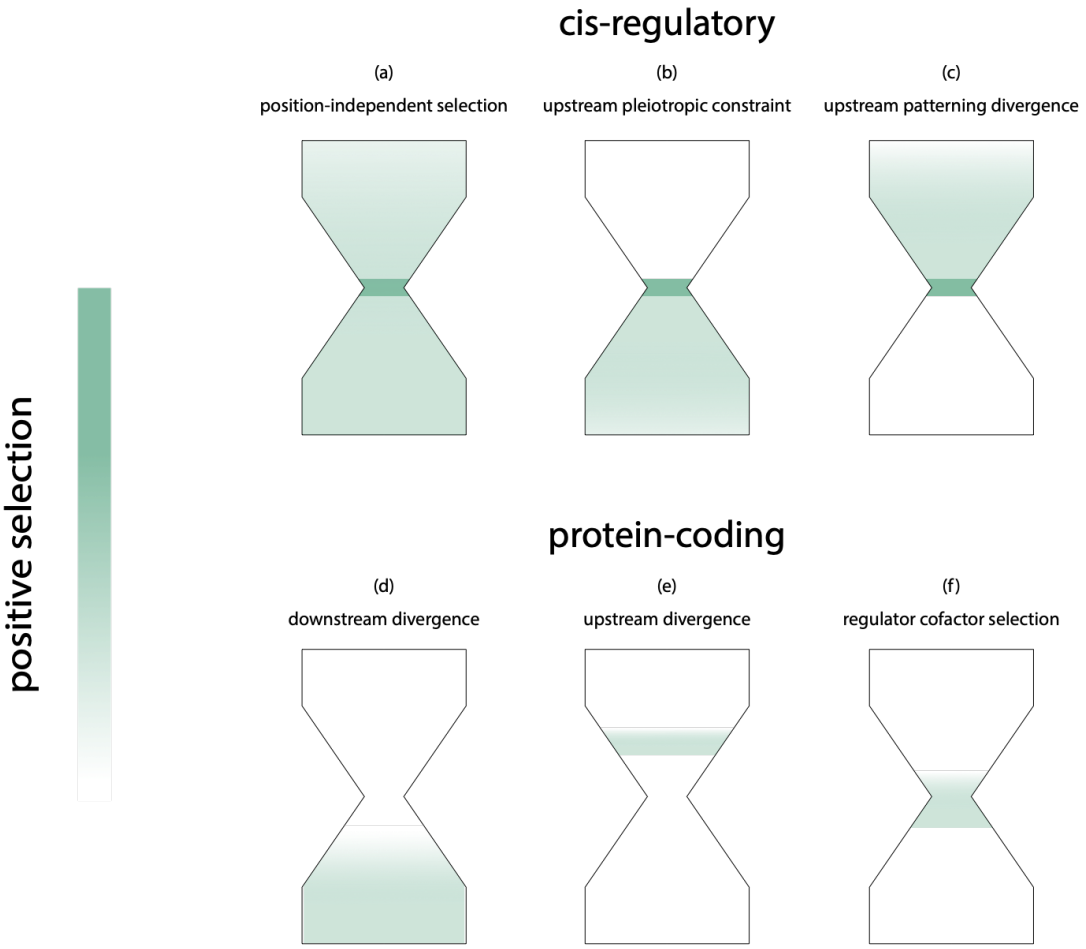
Figure 1: The bow-tie GRN consists of an input-output gene that is functionally connected to many genes upstream and downstream. The upstream and downstream genes can also be connected to other genes but not to nearly as many. Some developmental GRNs may fit the null hierarchal model, where there is little appreciable difference in connectivity between genes in the network.



819

820 **Figure 2: Levels of the *optix* GRN regulating wing color pattern in *Heliconius* butterflies.** The
821 GRN for wing color pattern is modeled as a bow-tie structure, with *optix* acting as the input-output
822 gene. *optix* is likely directly regulated by many upstream genes (inputs) and is known to directly
823 target many downstream genes (outputs). Direct targets of *optix* include (a) intermediate factors that
824 initiate downstream cascades that can be turned on or off, such as *dome/wash* as well as (b)
825 directly-targeted terminal effectors, such as the pigmentation enzyme *ebony*. The *optix* network
826 contains more regulatory relationships than shown here, and the number of inputs and outputs
827 involved in this GRN is likely much higher than illustrated [34,35].

828



829

830 **Figure 3: Distribution models of relative positive selection pressure across a bow-tie GRN.**

831 Patterns of positive selection on CREs (A-C) and protein-coding genes (D-F) at different positions in

832 a bow-tie GRN. Highlighted sections indicate the position within the network of the CRE(s) or

833 protein-coding gene(s) that selection is acting on and are not intended to indicate the number of

834 genes under selection (e.g. for a trait GRN to fit model A, CREs for genes upstream and

835 downstream are under selection, but not necessarily all CREs within the network).

836