

Multidimensional adaptive evolution of a feed-forward network and
the illusion of compensation

Kevin Bullaughey

Dept. of Ecology & Evolution, University of Chicago
920 E. 58th Street, CLSC 413
Chicago, IL 60637
773-702-2750 (ph.) 773-834-0505 (fax)
`bullaugh@uchicago.edu`

Running Head: Multidimensional adaptation

Key Words: adaptation, natural selection, gene regulation, epistasis, models/simulations

Abstract

When multiple substitutions affect a trait in opposing ways, they are often assumed to be compensatory, not only with respect to the trait, but also with respect to fitness. This type of compensatory evolution has been suggested to underlie the evolution of protein structures and interactions, RNA secondary structures, and gene regulatory modules and networks. The possibility for compensatory evolution results from epistasis. Yet if epistasis is widespread, then it is also possible that the opposing substitutions are individually adaptive. I term this possibility an adaptive reversal. Although possible for arbitrary phenotype-fitness mappings, it has not yet been investigated whether such epistasis is prevalent in a biologically-realistic setting. I investigate a particular regulatory circuit, the type I coherent feed-forward loop, which is ubiquitous in natural systems and is accurately described by a simple mathematical model. I show that such reversals are common during adaptive evolution, can result solely from the topology of the fitness landscape, and can occur even when adaptation follows a modest environmental change and the network was well adapted to the original environment. The possibility of adaptive reversals warrants a systems perspective when interpreting substitution patterns in gene regulatory networks.

Introduction

Advances in the functional annotation of DNA, combined with comparative genomic methods, allow for increasingly detailed characterizations of the evolution of biological function (Andolfatto 2005; Tuch *et al.* 2008; Gerke *et al.* 2009; Field *et al.* 2009; Chen *et al.* 2010). Explanations for observed sequence changes fall into several broad categories including random genetic drift, numerous types of adaptive change, compensatory evolution, relaxation of constraint, and others. Often several of these processes may be happening simultaneously, complicating efforts to understand the evolutionary forces behind the changes (Kreitman 1996; Lynch 2007; Barrett and Hoekstra 2011). The process of compensatory evolution is particularly compelling because it offers a means for the underlying encoding of a trait to evolve even as the correct functional output is conserved (Ancel and Fontana 2000; Moses *et al.* 2006).

Compensatory evolution has been invoked to explain the evolution of transcription factor binding sites (Gibson 1996; Moses *et al.* 2006; Ludwig *et al.* 2000; Lusk and Eisen 2010), RNA secondary structure (Stephan 1996; Chen *et al.* 1999; Ancel and Fontana 2000), protein structure and fold stability (DePristo *et al.* 2005; Callahan *et al.* 2011), nucleosome positioning (Kenigsberg *et al.* 2010), and *cis/trans* interactions (Landry *et al.* 2005; Tuch *et al.* 2008; Kuo *et al.* 2010). There are two commonly invoked modes of compensatory evolution. In the first, mutations are compensatory with respect to fitness, meaning that one or both of the mutations are individually deleterious, but jointly they are neutral or even advantageous. The functional effect of the alleles is not of primary relevance. In the second mode, the focus is the functional effect of the mutations. Jointly, the mutations preserve proper functioning of the system, but individually, one or both of the mutations would disrupt function. These two modes overlap when there is a direct relationship between function and fitness and often, such a relationship is implicitly assumed.

In the case of *cis*-regulatory modules, compensatory evolution is commonly offered as an explanation for functional reorganization, with nearly all such comments referring to the work of Ludwig *et al.* (2000). By making and expressing chimeric enhancer constructs,

Ludwig *et al.* (2000) showed that compensatory, epistatic changes must exist in the *even-skipped* stripe 2 enhancer despite conservation of functional output. In this setting, it is intuitive to imagine a direct relationship between function and fitness, particularly given the expression of important developmental enhancers is most certainly under stabilizing selection, and any functional mutation would likely disrupt proper expression.

However, when one assumes a direct link between functional changes and fitness consequences, several difficulties arise. Even mutations of very small deleterious effect are usually eventually removed from the population (Ewens 2004), such that an initial deleterious substitution is unlikely and the process is likely to proceed very slowly. Even if compensatory evolution occurs via the joint fixation of conditionally neutral mutations, it is also likely to proceed very slowly (Kimura 1985; Stephan 1996; Iwasa *et al.* 2004; Durrett and Schmidt 2008), and thus may not account for the bulk of apparent compensatory evolution. Ludwig *et al.* (2000) had argued that nearly-neutral changes might underlie the compensatory evolution. This requires either very many small-effect mutations or a function-fitness mapping in which mutations of large functional effect can nonetheless be nearly neutral. This latter explanation seems plausible based on a model of enhancer function (Bullaughey 2011), and has been termed pseudo-compensatory change (Haag 2007).

While there is a tendency to assume stabilizing selection on the trait, and conclude that compensatory evolution explains turnover, adaptation is also a possibility. He *et al.* (2011) examined nucleotide substitutions in *Drosophila* affecting known transcription factor binding sites in developmentally-important enhancers. The authors characterized each substitution with respect to whether it increases or decreases the site's affinity to the transcription factor. They found an excess of affinity-decreasing mutations in *Drosophila simulans*, suggesting adaptive loss of transcription factor binding sites. In addition, there were some affinity-increasing mutations, but not more than expected relative to the benchmark (substitutions that did not alter affinity). How should one interpret such changes? Are some of the affinity-decreasing mutations compensated by affinity-increasing

mutations? What about the excess of affinity-decreasing mutations? If adaptation is really driving the loss of binding sites, how do we interpret the affinity-increasing mutations? Would these have been deleterious?

In answering these questions, the distinction between compensation with respect to function and compensation with respect to fitness becomes important. Two mutations that appear compensatory with respect to function need not be compensatory with respect to fitness. In particular, when a system is evolving adaptively, mutations that have opposing effects with respect to a function, and thus appear compensatory, may be individually adaptive due to other epistatic changes that are occurring in the genetic background. I term this phenomenon an adaptive reversal. In an adaptive reversal, the sign of the fitness effect of altering a phenotype in a particular direction flips, as changes affecting other traits occur. In this sense, it is similar to sign epistasis, in which the sign of the fitness effect of a mutation depends on the genetic background (Weinreich *et al.* 2005).

At its simplest, epistasis is a relationship among multiple loci or traits whereby the joint effect of several changes is not the outcome expected of the individual changes (see Phillips 2008 for a good review). Epistasis can exist for many reasons, including structural considerations, molecular interactions, relationships mediated through the concentrations of molecules (e.g., transcriptional regulation), and even host-pathogen interactions. Despite its obvious importance (Wolf *et al.* 2000), we presently understand few of the seemingly infinite array of epistatic interactions, partly because it is challenging to detect (Whitlock *et al.* 1995). Our limited knowledge of epistasis has prompted many investigators to consider flexible models of two- or three-locus interactions, in which the epistatic effects are parameters, and not based on assumptions regarding specific biological interactions (Cheverud and Routman 1996; Wolf *et al.* 2000; Takahasi and Tajima 2005; Gandon and Otto 2007). Another approach is to consider complex models that capture epistasis among many loci (Gavrilets *et al.* 1998; Weinreich *et al.* 2005; Yukilevich *et al.* 2008), at many nucleotides within a locus (Kauffman and Levin 1987; Orr 2006) or in many dimensions

(Martin and Lenormand 2006), but again the epistasis present in these models is not based on explicit biological mechanisms.

One common framework for considering arbitrary epistatic relationships is the fitness landscape, with valleys and peaks representing regions of lower or higher fitness. Popularized early on by Wright (1932), fitness landscapes have continued to provide a fertile analogy for discussing evolutionary processes (Gavrilets 2004; Skipper 2004; Orr 2005; Weinreich *et al.* 2005, although see Kaplan 2008). A fitness landscape prescribes a fitness to each multilocus genotype (or combination of phenotypes), allowing for any type of epistasis. The simplest fitness landscape model is Fisher’s Geometrical model (Fisher 1930), which despite its simplicity has yielded important theoretical predictions (Orr 2005; Martin and Lenormand 2006) and a framework to test empirical measurements (Martin *et al.* 2007).

Yet like the models that directly prescribe epistatic interactions, most studies invoking fitness landscapes do not rely on known biological functional relationships to specify the landscape, with some notable exceptions (e.g., Poelwijk *et al.* 2006; Ancel and Fontana 2000; Perfeito *et al.* 2011). Instead, as a means of increasing generality and tractability at the cost of realism, landscapes are mainly parameterized by their dimensionality (Orr 2000), covariance matrix (Martin and Lenormand 2006), connectivity (Yukilevich *et al.* 2008; Macía *et al.* 2012), or degree of ruggedness (Orr 2006).

To evaluate whether adaptive reversals are plausible, in particular whether they stem naturally from the epistasis intrinsic to a gene network, I investigate them using an explicit model of a small regulatory network. This three-gene regulatory network, the type I coherent feed-forward loop, is a pathway that is ubiquitous in nature (Shen-Orr *et al.* 2002; Mangan *et al.* 2003), has been well-studied in *Escherichia coli*, has behavior that is accurately captured by a simple mathematical model (Mangan *et al.* 2003), and offers an interpretable basis for fitness (Dekel *et al.* 2005).

Methods

Dynamic model of a feed-forward gene circuit: A broad overview of the model and the rationale is given in the Results; what follows here are the modeling details. I closely follow the modeling conventions and notation of Alon and colleagues (Shen-Orr *et al.* 2002; Alon 2006), modeling network dynamics with a system of differential equations and fitness as a cost-benefit trade-off. The topology of the type I coherent feed-forward loop is given in Figure 1A and the network dynamics in Figure 1B. The dynamics of genes Y and Z are modeled as a pair of differential equations:

$$\frac{dY}{dt} = \beta_y[X^*] - Y\alpha_y$$

$$\frac{dZ}{dt} = \beta_z[X^* \cap Y > k] - Z\alpha_z$$

β_y and β_z are parameters (or traits) giving the production rates of Y and Z respectively. In each case X needs to be in its active form, X^* , and in order for Z to be induced, Y must be above a threshold k (Iverson square brackets indicate boolean functions). The genes are degraded/diluted at rates α_y and α_z , proportional to the current protein concentrations. The threshold, k , is an approximation to nonlinear activation dynamics of Z by Y (Alon 2006). Differential equations of this form have two solutions depending on whether production is turned on or not. For gene Y , the expression when the gene has been turned on for time t is:

$$Y_{\text{ON}}(t) = \frac{\beta_y}{\alpha_y}(1 - e^{-\alpha_y t})$$

Time t' after getting turned off having reached expression level Y_0 , the dynamics follow an exponential decay:

$$Y_{\text{OFF}}(t'|Y_0) = Y_0 e^{-\alpha_y t'}$$

Expression of Z is similar, except that in addition to activation of X , Y must surpass the threshold k before Z is expressed. This occurs at some delay, τ , that is a function of the production and degradation rates associated with Y :

$$\tau = -\frac{1}{\alpha_y} \log(1 - k \frac{\alpha_y}{\beta_y})$$

Costs associated with a network may come in the form of explicit costs, such as energetic costs (i.e., transcriptional and translational costs), and in the form of implicit costs, such as the opportunity cost of deploying limited resources. For simplicity, I assume that all these costs are proportional to the expression level of the protein with constant of proportionality, η_y , for gene Y and η_z for gene Z . The cost of Y for a particular signal of duration, d , is thus:

$$\text{COST}_Y(d) = \eta_y \int_0^d Y_{\text{ON}}(t) dt + \eta_y \int_0^\infty Y_{\text{OFF}}(t' | Y_{\text{ON}}(d)) dt' = \frac{\eta_y \beta_y d}{\alpha_y}$$

The first term in the sum is the cost while X is actively transcribing Y and the second term is the cost after X is deactivated and Y expression is subsequently shut down. These two terms are unlikely to have the same constant, η_y , but as a first-order approximation this may not be unreasonable.

The cost of Z expression is defined similarly, although upon activation of X , expression of Z starts at a delay, τ , meaning Z is only expressed for a duration, $d - \tau$:

$$\text{COST}_Z(d) = \begin{cases} \eta_z \int_0^{d-\tau} Z_{\text{ON}}(t) dt + \eta_z \int_0^\infty Z_{\text{OFF}}(t' | Z_{\text{ON}}(d - \tau)) dt' = \eta_z \beta_z (d - \tau) / \alpha_z & : \text{if } d > \tau \\ 0 & : \text{if } d \leq \tau \end{cases}$$

Similarly, I model the benefit, which captures the contribution of the network to fitness (e.g., by increasing growth rate), purely as a result of Z expression and proportional to it at rate δ . The benefit only applies while the signal, S_X is active (i.e., when the

metabolizable substrate is present), and when the signal duration is long enough for Y to pass the threshold, k , which happens at time τ :

$$\text{BENEFIT}_Z(d) = \begin{cases} \delta \int_0^{d-\tau} Z_{\text{ON}}(t) dt & \text{if } d > \tau \\ 0 & \text{if } d \leq \tau \end{cases}$$

These cost-benefit assumptions are oversimplifications for several reasons. Opportunity costs likely increase faster than linearly with expression and the benefit likely shows diminishing returns (Alon 2006). Also, not all costs are proportional to expression, some are proportional to the production rates, β_y and β_z ; while others are proportional to the degradation rate $Y\alpha_y$ and $Z\alpha_z$. Nonetheless, these simple cost-benefit relationships capture some of the epistasis among the evolvable traits.

Environment-specific fitness: Fitness is environment-specific, with the environment entirely described by the duration distributions of true and false signals. I desired the false signals to be primarily short, while allowing the true signals to be substantially longer. This pattern justifies the coherent type I feed-forward network, which functions as a filter of short signals (Alon 2006). I consider two environments, low- and high-noise environments. In order to maintain a closed form for fitness, I use simple shapes for these distributions. The distribution of false signals, $f_F(t)$, is quadratic in the range $(0, t_F)$ and satisfies the following criteria:

$$f_F(0) = 0, f_F(t_F) = 0, \int_0^{t_F} f_F(t) dt = 1, t_F = \frac{1}{2}$$

making $f_F(t)$:

$$f_F(t) = \begin{cases} \frac{6}{t_F^2}t - \frac{6}{t_F^3}t^2 & : \text{if } 0 < t < t_F \\ 0 & : \text{else} \end{cases}$$

The distribution of true signals is uniform with durations spanning the range $(0, t_T)$, and also integrates to unity. Thus $f_T(t) = \frac{1}{t_T}$ in the interval $(0, t_T)$ and zero elsewhere. For a given environment, the frequency of false and true signals are c_F and c_T respectively. For the simulations presented, $c_T = 0.15$. The only difference between the low- and high-noise environments is that the abundance of false signals in the high-noise environment is 6-fold higher ($c_F = 0.3$) than in the low-noise environment ($c_F = 0.05$).

The total fitness component attributable to this network motif is then the sum of two integrals over the true, $f_T(s)$, and false, $f_F(s)$, signal duration distributions:

$$\text{FITNESS} = \sum_{i \in \{T, F\}} \int_S c_i (\text{BENEFIT}_Z(s)[i = T] - \text{COST}_Y(s) - \text{COST}_Z(s)) f_i(s) ds - \text{PENALTY}_L - \text{PENALTY}_H$$

Iverson square brackets indicate a boolean function (i.e., there is no benefit for a false signal). The penalties are introduced in order to constrain the evolvable traits from becoming too high or too low (see below).

Putting everything together, the closed form for the fitness function, integrating over the distributions of true and false signals, is the following:

$$\begin{aligned} \text{FITNESS}(\alpha_y, \beta_y, \alpha_z, \beta_z, k | \delta, \eta_y, \eta_z, c_F, c_T, t_F, t_T) = & \\ & - \frac{\beta_y \eta_y (c_F t_F + c_T t_T)}{2\alpha_y} \\ & - c_F [\tau < t_F] \frac{\beta_z \eta_z}{2\alpha_z t_F^3} (t_F + \tau)(t_F - \tau)^3 \\ & - c_T [\tau < t_T] \frac{\beta_z \eta_z}{2\alpha_z t_T} (t_T - \tau)^2 \\ & + c_T [\tau < t_T] \frac{\beta_z \delta}{\alpha_z^3 t_T} \left(1 - \alpha_z t_T - e^{-\alpha_z(t_T - \tau)} + \alpha_z \tau + \frac{1}{2} \alpha_z^2 (t_T - \tau)^2 \right) \\ & - \text{PENALTY}_L - \text{PENALTY}_H \end{aligned}$$

The first five parameters of $\text{FITNESS}(\cdot)$ are the evolvable traits and the remaining are for the cost and benefit functions and the signal duration distributions (see Table 1). On the

right side of the equality are six terms: (i) the cost of Y expression, (ii) the cost of Z expression due to false signals, (iii) the cost of Z expression due to true signals, (iv) the benefit due to Z expression, and (v-vi) the penalties.

Penalty: The model described thus far suffers from several unrealistic behaviors that arise when the evolvable traits reach very high or low values. There are two strategies to circumvent these problems. One strategy is to make the model more realistic, so that it captures the non-linear consequences of very small or large trait values. The alternative is to simply introduce a fitness penalty that ensures that the system will not evolve into these implausible domains. I chose the latter strategy in order to keep the model as simple as possible while avoiding unrealistic outcomes.

As an example, the cost of protein expression should be a convex function of protein concentration (Dekel and Alon 2005) because, at very high expression levels, additional translated molecules deprive core cellular functions of needed protein, or become toxic. Rather than explicitly incorporate a non-linear cost, I achieve essentially the same constraint by penalizing high values of β . As another example, consider two aspects of the model: First, for an arbitrarily low steady-state expression level of Y , an appropriately small value of k can be chosen to achieve any desired delay, τ . Second, Z production is independent of Y expression conditional on a given τ . These two properties lead to the unlikely behavior that Y expression can be tuned to be arbitrarily low while maintaining the same profile of Z expression. However, at very low expression levels, the concentration is no longer well-approximated by a continuous, smooth curve, and instead is highly stochastic, depending on chance molecular interactions. Thus, a network relying on very low expression levels to activate a downstream pathway would not reliably maintain the pathway in either the on or off state. This would be costly, a feature not captured by the model. I prevent this from occurring by penalizing very small trait values.

I chose the following arbitrary penalties for penalizing high and low values of the traits respectively:

$$\text{PENALTY}_H = \exp \left\{ \frac{1}{25} \sum_j q_j \right\}$$

$$\text{PENALTY}_L = \exp \left\{ - \sum_j \log(q_j) \right\},$$

where q_j is the value of the j 'th evolvable trait. By design, the penalty for low values is much more abrupt than the penalty for high values, reflecting the non-linearities missing from the simplified model. For what I consider the meaningful parameter range, these penalties have little effect on fitness, as desired.

Numerical evaluation of adaptation using the small mutation size

approximation: In order to examine whether adaptive reversals are an intrinsic result of the topology of the fitness landscape, I consider an approximation to the adaptive process that is equivalent to computing the adaptive trajectory as the mutation step size becomes infinitesimally small. I refer to this as the small mutation size (SMS) approximation.

Consider the evolutionary trajectory of the population on a relatively smooth, continuous fitness landscape. The phenotype, $\mathbf{x} = (\alpha_y, \beta_y, \alpha_z, \beta_z, k)$, starts at coordinate \mathbf{x}_0 at time $t = 0$, and proceeds to evolve adaptively. The path traced out by evolution, $\mathbf{x}(t, \mathbf{x}_0)$, is thus a function of time and is dependent on the initial starting point. When the mutation rate is sufficiently high, and the mutational size is sufficiently small, adaptation will be nearly deterministic.

Another interpretation of this trajectory is as an expectation, $\bar{\mathbf{x}}(t, \mathbf{x}_0)$. If the fitness landscape is sufficiently smooth, then in the current local vicinity, $\mathbf{x} + \delta\mathbf{x}$, the landscape is a hyperplane with a slope in the orientation of the gradient of the fitness function. The expected phenotype some time δt in the future is simply a small step in the most up-hill direction (under the assumption of a symmetrical mutation kernel). Thus in this deterministic approximation, adaptation proceeds by marching smoothly up the landscape

in the most up-hill direction. Since future evolution in a constant environment depends only on the current phenotype and is independent of the history, one can think of the expected adaptive trajectory as satisfying a system of time-homogeneous ordinary differential equations:

$$\frac{dx_i}{dt} = h_i(\bar{\mathbf{x}}(t, \mathbf{x}_0)) \quad \forall x_i \in \{\alpha_y, \beta_y, \alpha_z, \beta_z, k\},$$

where $h_i(\cdot)$ is some function of the current phenotype. Knowing each $h_i(\cdot)$ for some starting trait vector, \mathbf{x}_0 , would then allow computation of the path using standard tools. Note that “time” in the SMS approximation relates to the evaluation of the path as a system of differential equations and does not necessary correspond directly to evolutionary time. Under the simplifying assumptions that I have described, the system of differential equations, $h_i(\cdot)$, is simply the gradient of the fitness function, F , evaluated at $\bar{\mathbf{x}}(t, \mathbf{x}_0)$:

$$h_i(\bar{\mathbf{x}}(t, \mathbf{x}_0)) = \left. \frac{\partial F}{\partial x_i} \right|_{\bar{\mathbf{x}}(t, \mathbf{x}_0)} \quad \forall x_i \in \{\alpha_y, \beta_y, \alpha_z, \beta_z, k\}$$

or alternatively:

$$\nabla F(\bar{\mathbf{x}}(t, \mathbf{x}_0)) = \left(\frac{\partial F}{\partial \alpha_y}, \frac{\partial F}{\partial \beta_y}, \frac{\partial F}{\partial \alpha_z}, \frac{\partial F}{\partial \beta_z}, \frac{\partial F}{\partial k} \right) \bigg|_{\bar{\mathbf{x}}(t, \mathbf{x}_0)}$$

I perform the numerical evaluation using the classic Runge-Kutta method (RK4) for solving systems of differential equations. In order to ensure that numerical evaluation of the gradient-ascent path is accurate given the curvature of the surface, I shrink the move at any step of the algorithm so that it corresponds to at most a 10% change in any single dimension. This has the effect of limiting the rate at which a parameter changes and serves to increase the accuracy of the computed path.

Adaptive walk simulations with discrete mutational effects: I simulate an adaptive walk using a standard technique in which candidate mutations are sampled from a mutation kernel and accepted as a substitution or rejected based on the fixation probability (Gillespie 1994). I compute fixation probabilities using the standard diffusion result for a haploid population (Ewens 2004):

$$\text{Prob}(\text{fix}) = \frac{1 - e^{-s}}{1 - e^{-Ns}}$$

where $N = 1000$ and s is the selection coefficient for the proposed mutation relative to the current fitness:

$$s = \frac{w_{\text{proposed}}}{w_{\text{current}}} - 1$$

A mutant network is arrived at as follows. First a trait is chosen among the evolvable traits with equal probability; each mutation affects only a single trait. Second, a mutational effect size is sampled. I chose to represent mutations on a multiplicative scale so that effect sizes are scale-invariant; in other words, a trait always has the same probability of changing by a given fraction regardless of the current trait value. Thus, for a mutational effect u and an original trait value of X , the new trait value is $X' = Xe^u$. Had I assumed a mutation has an additive effect drawn from a fixed distribution, then traits that evolved to large trait values would have also evolved smaller relative mutational effect sizes.

For the simulations presented, I draw u from a two-component mixture of zero-centered Gaussian distributions with $\sigma_1 = \frac{1}{20}$ and $\sigma_2 = \frac{1}{5}$ and mixture proportions of 0.7 and 0.3 respectively. Although arbitrary, the mutation kernel allows for a small number of large effect mutations and a larger number of relatively modest effect ones.

The adaptive walk simulations were run for 10,000 mutational proposals, resulting in an average of 24 substitutions. Given $N = 1000$, this represents a 2.4-fold increase above what would be expected under neutrality.

Results

Limitations of the single gene view of adaptation: Consider a single gene with substitutions in its promoter on a particular lineage. If the promoter is otherwise conserved, then the clustering of substitutions might suggest adaptation (Pollard *et al.* 2006; Holloway *et al.* 2008; Kim and Pritchard 2007). Further suppose that these substitutions are known to have sizeable functional effects. A cartoon illustrating such a scenario (Figure 2) shows three substitutions that decrease and one that increases expression. How does one interpret that one substitution is of opposite effect relative to the others? For a single evolving trait and a constant environment, there are three basic alternatives.

The first alternative is that these substitutions are neutral with respect to fitness, despite their demonstrable functional effects. I do not further discuss this possibility, as the present study is concerned with adaptation. In the second scenario, one mutation is deleterious and another compensatory (over-compensation is shown in Figure 3A). However, only deleterious mutations of very small effect have any appreciable chance of substituting, unless the substitution occurs jointly with compensatory ones (Kimura 1985), and in either case the rate of such substitutions will likely be very low (Stephan 1996; Ewens 2004; Iwasa *et al.* 2004; Durrett and Schmidt 2008).

In the third scenario, the functional effects are opposing, but both mutations are individually adaptive. Overshooting an optimum is an example of this (Figure 3B), which requires that the curvature (or ruggedness) of the landscape be on the same scale as the mutational effect sizes. In the case when small mutations are more likely than larger ones (a common assumption), the overshooting alternative is less likely than simply ascending the fitness landscape.

All of the above involves consideration of only a single trait—no epistatically-interacting traits are simultaneously evolving. This severe assumption limits the possible interpretations of opposing substitutions that affect a single trait. If there are epistatic interactions, more complex behaviors can arise. In particular, it is possible for

mutations to be of opposite effect with respect to a trait but be individually adaptive (Figure 3C).

An ideal case study—the type I coherent feed-forward loop: In the preceding discussion, the relationships between evolvable traits and fitness were arbitrary. The main aim of the present study, however, is to investigate a more concrete, biologically realistic fitness landscape. While little is known about the topology of fitness landscapes, biological systems are highly modular (Hartwell *et al.* 1999; Schlosser and Wagner 2004; Wagner *et al.* 2007), suggesting progress might be made by considering a small, tractable module, the function of which is relatively well understood and isolated from the many other subsystems of the organism (Alon 2006).

One such module that may be instructive is the type I coherent feed-forward loop (Shen-Orr *et al.* 2002). In its simplest incarnation, this small genetic network consists of only three genes (shown in Figure 1A). Gene X , upon responding to a signal, induces gene Z both directly and indirectly via gene Y with the two inputs to gene Z integrated through an AND operation. This regulatory module occurs in many organisms including *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and humans (see Kalir *et al.* 2005 for citations), and in the best characterized regulatory network of *E. coli*, it is highly over-represented (Shen-Orr *et al.* 2002). One reason this motif may be so common is that it performs an important biological function; namely, it acts as a sign-sensitive delay element (Mangan *et al.* 2003), in which induction of Z is delayed upon activation of the signal, S_X , but turned off immediately upon cessation of the signal (Shen-Orr *et al.* 2002). As a result of this sign-sensitive delay, it functions to filter out short pulses in the signal, inducing whatever pathway is downstream of Z only when the signal is sustained. Figure 1B illustrates the temporal dynamics of this feed-forward loop. Fitness is modeled as a cost-benefit trade-off, closely following Dekel *et al.* (2005) and Alon (2006) and motivated by optimality principles (Rosen 1967). In the case of the most

well-studied instantiation of this motif—the L-arabinose utilization system of *E. coli*—the benefit is primarily the ATP derived from metabolism of arabinose (Mangan *et al.* 2003).

Finally, and importantly here, the dynamics of this network motif are well-understood and accurately captured by a simple mathematical model, and they appear accurate even when embedded in the full gene regulatory network of an organism (Mangan *et al.* 2003). For these reasons, the type I coherent feed-forward loop is an ideal test case for adaptation in multiple dimensions, having the property that epistasis present among the functional dimensions emerges from a biologically realistic set of interactions.

The evolutionary maintenance of this motif rather than direct regulation of Z or no regulation of Z may occur in a number of circumstances. One scenario involves a particular cost-benefit trade-off, in which there is a fixed cost of expression, independent of level, and a benefit proportional to the expression level (Dekel *et al.* 2005). Another circumstance in which a sign-sensitive delay element will be favored is when the signal is noisy and the distribution of true and false signals differ sufficiently in their mean durations. I model this latter scenario and assume a heterogeneous mixture of true and false signals; when the signal is true, the benefit ensues, but a false signal has no benefit. When false signals are generally short, relative to true signals, this network will act as an effective filter.

Adaptive reversals stem directly from the fitness landscape topology: I

investigate the prevalence of adaptive reversals using the small feed-forward network I described earlier. I consider only five traits to be evolvable: the degradation and production rates of Y (α_y and β_y), the corresponding rates for Z (α_z and β_z), and the expression threshold, k . I assume that each of these is continuous, which may not be a poor approximation if sufficiently many nucleotides encode each of these traits.

I first consider an idealized approximation to the adaptive process that I refer to as the small mutation size (SMS) approximation (see Methods). If mutations are tiny relative to the curvature of the landscape, and the mutation rate is sufficiently high, then the adaptive trajectory will track the gradient of the fitness landscape (i.e., it will climb to a

local optimum via the most uphill path). I have formulated the fitness function in such a way that it has a closed form and known partial derivatives with respect to each evolvable trait. I am thus able to numerically evaluate the adaptive trajectory by taking the gradient of fitness and treating it as a system of differential equations formed by these partial derivatives, starting from some initial point in the five-dimensional parameter space.

In this idealized adaptive process, both compensatory mutation and over-shooting the optimum are impossible, and so by projecting the adaptive trajectory onto one trait axis, any observed reversal will be a direct result of the topology and the epistasis encoded in the genetic network. An adaptive trajectory on this fitness landscape is a path in six dimensions—five trait dimensions and fitness. In order to investigate and visualize the prevalence of adaptive reversals, I computed 200 SMS adaptive paths from random starting points in the five-dimensional hypercube. Figure 4 shows several of these trajectories projected onto two-dimensional slices of the six-dimensional fitness landscape, each slice corresponding to one evolvable trait dimension and fitness (also see Table S1 for a summary).

There are several qualitative behaviors apparent in these projections: (i) four of the five traits exhibit reversals, but not all traits do, (ii) some traits, (e.g., k) show much larger, more dramatic reversals than other traits, and (iii) the reversals do not happen strictly on fitness plateaus, i.e., fitness can increase substantially even during the return portion. Thus, even in this very low-dimensional, continuous trait setting, adaptation of some traits may often reverse course as a result of adaptive changes elsewhere.

Thus far, I considered adaptation from random starting points. However, evolution does not proceed in this way, but rather the ascent starts when a relatively well-adapted species encounters a new environment or the prevailing environment changes.

In a feed-forward loop, the major influence of the environment is the signal S_X , which in the case of the L-arabinose utilization network of *E. coli*, represents the presence or absence of L-arabinose. I therefore define the environment by the duration distributions of true and false signals. I then consider the case of adaptation to a new environment with a

higher proportion of false signals (high-noise), starting from a population well-adapted to a low-noise environment (see Figure 5A).

Upon introduction to the high-noise environment, the low-noise-adapted network has a much lower fitness than it had in the low-noise environment. In response, the network evolves towards a solution with higher fitness (Figure 5B). The most pronounced reversal that occurs in this particular bout of adaptation affects trait k (Figure 5C). Adaptation to the high-noise environment and the reversal of trait k can be interpreted as follows. In the low-noise environment, false signals were comparatively rare, and thus there was little need to delay activation of Z via a high threshold k . Thus the low-noise-adapted network exhibits a low value for trait k . However, upon introduction to the high-noise environment, the most expeditious way to increase the network’s fitness was to increase k , as this prevents Z from unnecessary induction. As a result, k quickly evolved to a higher value, and the delay in Z induction, τ , correspondingly increased (Figure 5D). However, after τ became long enough to filter out the false signals, k , in conjunction with α_y and β_y , evolved in a way that preserved τ , but reduced the level of Y expression, thereby lowering costs and increasing fitness. As a result, k exhibited an adaptive reversal.

Adaptive reversals also occur with discrete mutational effects: While the SMS approximation is useful for showing that reversals can result solely from the topology of the fitness landscape, there are a number of unrealistic aspects to this model. First, real mutations are likely of some discrete, potentially large effect, and thus adaptation is not a smooth optimal ascent up the landscape. Second, separate biological functions are often encoded in separate locations on the DNA (e.g., Ptashne 1988; Kirchhamer *et al.* 1996; Barrière *et al.* 2011), and thus mutations may generally affect only one of the various functions.

In order to demonstrate adaptive reversals when mutational effects-sizes are discrete and affect only single traits, I simulate an adaptive walk consisting of a succession of such mutations (Gillespie 1994). I use rejection sampling, with the acceptance probability of a

proposed mutation equal to the fixation probability (see Methods). In this approximation to a population process, drift and selection are both accounted for. However, the joint fixation of mutations that would not substitute individually is not possible, removing a confounding type of compensatory change. This modeling decision also means that linkage relationships among traits can be ignored.

In order to characterize the propensity for reversals when mutational effects are discrete, I initiate 1000 simulations from random starting points and evolve the networks in the high-noise environment of Figure 5. As under the SMS approximation, adaptive reversals are common; among these simulations, there are adaptive reversals affecting 30% of the simulations and all the traits (see Table 2). By dividing the adaptive reversals into two components, the initial portion and reverse portion (the portion in the opposing direction), it becomes apparent that the typical substitution on the initial portion is of much larger selective effect than the typical substitution on the reverse portion, despite that mutations on the initial and reverse portions often have similar effects on the traits (Table 2). This finding underscores the need to distinguish between functional effects and selective effects and it is consistent with adaptive walks on a fixed landscape offering diminishing returns (Orr 2005).

Unlike in the SMS approximation, the discrete mutational effects in these simulations introduce the possibility of overshooting an optimum. To determine if this was the case for each substitution, I retrospectively assess fitness at a grid of effect sizes smaller than the observed substitution. An intervening optimum was overshoot when a smaller effect size leads to a greater fitness than that associated with the observed substitution. I only count adaptive reversals that do not overshoot an intervening optimum.

As in the case of the SMS approximation, it may be more meaningful to consider adaptation to a new environment starting from a network adapted to a distinct but similar environment, rather than from arbitrary points in the fitness landscape. Therefore, I also conduct 100 replicate simulations starting from a network that is initially well-adapted to the low-noise environment of Figure 5, and evolve the network in the high-noise

environment (all replicates start from the same point). Two examples, shown in Figure 6, exhibit adaptive reversals affecting the production and degradation rates of Z (α_z and β_z). One advantage of a model in which each substitution only affects a single trait is that the beneficial (or deleterious) effect is clearly attributable to that trait. In other words, substitutions that have opposite effects with respect to the trait are individually adaptive solely as a result of the mutation’s effect on that trait, and not because of a pleiotropic effect of the mutation.

The replicate simulations are often very similar, as is expected when adaptation is largely determined by the topology of the landscape and when the starting point is the same for all replicates. Nonetheless, the replicate simulations differ in whether an adaptive reversal is observed or not. So although adaptive reversals are observed in 70% of the replicate simulations, others converge at near optimal solutions without exhibiting any adaptive reversals. In this way adaptive reversals, as with many features of evolution, exhibit a high degree of stochasticity and historical contingency (Lenormand *et al.* 2009; Bullaughey 2011; Salverda *et al.* 2011).

Traversal of a fitness landscape with discrete mutational effects is not limited to exactly ascending the fitness gradient, and thus it is even possible to have adaptive reversals involving traits that do not generally show adaptive reversals in the SMS approximation. For example, under the SMS approximation when starting from arbitrary parameterizations of the feed-forward network, adaptive reversals were not found to affect the production rate of Z , β_z (see Figure 4). Yet in the simulations detailed in Table 2, the most commonly affected trait is β_z and one example is shown in Figure 6 (right panel). Conversely, the threshold k does not show as many reversals in the discrete effect model as in the SMS approximation. These discrepancies appear to result from the SMS approximation allowing multiple parameters to change simultaneously. This suggests that discrete mutational effects may actually increase the opportunities for adaptive reversals.

Mutational target size and dimensionality: Adaptive reversals in the SMS approximation are impossible in one dimension, while they appear common in five dimensions (Figure 4). Clearly, the dimensionality of the landscape plays a role in whether adaptive reversals occur. But dimensionality is not a smoothly tunable knob; instead it is the combination of particular epistatic relationships intrinsic to the network construction, dynamics, and fitness model.

In order to gain insight into whether adaptive reversals are primarily the result of epistasis between certain subsets of the traits and how the number of evolving traits affects the prevalence of adaptive reversals, I again evolved networks using the SMS approximation starting from random points. However, I constrained evolution to a subset of the evolvable traits.

I considered all 31 non-empty subsets of the five evolvable traits. Fixing a trait—preventing it from evolving—has several interpretations. First, it removes this trait and all associated epistatic interactions with other traits from the model, thereby reducing the degree of epistasis. Second, the fixed traits can be thought of as having a mutational target that is so much smaller than the other traits that they rarely, or in this case never, evolve. Thus, this exercise is informative in that, in real biological systems, traits may have very different mutational target sizes, which in turn may shape which adaptive trajectories are most likely and whether adaptive reversals occur. The results of this analysis are presented in Figure 7.

As expected, adaptive reversals are absent when only one trait is allowed to evolve. In general, as the number of evolvable traits increases, there is a trend toward more reversals (Figure 7, top) and reversals of larger effect (Figure S1). But this trend is not perfect; some combinations of parameters are much more likely to exhibit adaptive reversals than others. This pattern reflects the nature of the epistasis. For example, in landscapes with only β_y and one other trait free to evolve, which trait exhibits reversals varies dramatically. When β_y evolves in conjunction with α_y , the former exhibits many reversals and the latter does not. When β_y evolves with α_z , the pattern is reversed, with β_y

not exhibiting many reversals while α_z does. When β_y and k evolve together, both show many reversals, and when β_y and β_z evolve together, neither show many.

These patterns align with the general intuition that some traits are coupled more tightly by epistatic interactions than others. How the propensity to undergo adaptive reversals scales to fitness landscapes of more than five dimensions remains to be seen, and will require biologically plausible higher-dimension models.

Discussion

In principle, epistasis allows for what I call an adaptive reversal—a trait adaptively evolving in one direction and then, after a change elsewhere in the system, evolving adaptively in the reverse direction. In such adaptive reversals, mutations of opposing functional effect are individually adaptive, which is in contrast to compensatory evolution, in which changes of opposing functional effect also have opposing fitness effects (i.e., a damaging and a restorative mutation). Whether regulatory networks exhibit an epistatic structure that can lead to adaptive reversals is not immediately clear, however, as the epistasis is shaped by functional properties of the system.

There are a number of *a priori* reasons why one might not expect adaptive reversals to be a common feature of regulatory evolution. First, epistasis relates to the dimensionality of the biological subsystem under consideration. For example, epistasis internal to a protein may relate to the catalytic properties of the active domains, the shapes of the interaction surfaces, the structural scaffold, the stability of the protein, and the folding efficiency. The protein, in these regards, is a high-dimensional object. In contrast, interactions in regulatory networks are largely mediated through expression levels, which are continuous and perhaps of lower dimensionality, varying only over time and space. Furthermore, the bulk of epistatic interactions may be among nucleotides involved in encoding a single biological function. For example, epistasis may be common within a protein, where there may be many interacting residues (Lunzer *et al.* 2010), but less common between proteins, because interaction surfaces may be limited to a subset of

domains and residues (Heger and Holm 2003). Similarly, epistasis in regulatory networks may be limited by the modularity characteristic of networks (Hartwell *et al.* 1999; Alon 2003); genes are encoded separately, transcriptional control is encoded separately from the determinants of transcript and protein stability, and the *cis*-regulatory encoding is itself modular. On the other hand, in a multi-component, adaptively evolving system, adaptation may be an especially strong driver of epistatic changes in the genetic background, making the assumption of no epistasis less plausible. So while adaptive reversals may not be surprising when one considers the dizzying possibilities for arbitrary high-dimensional fitness landscapes, that adaptive reversals occur in a relatively low-dimensional, biologically plausible model for regulatory function, is important.

In order to investigate the prevalence and nature of adaptive reversals in a setting in which the epistatic properties are determined by a biologically-plausible functional model, I employed a ubiquitous and well-understood regulatory network: the type I coherent feed-forward loop (Shen-Orr *et al.* 2002; Mangan *et al.* 2003; Alon 2007). This network motif offers an interpretable, low-dimensional model for a component of fitness, and thus may be informative for considering the relationships among functional effects and fitness effects of mutations. The fitness model is based on a cost-benefit trade-off relating the benefit of pathway activation and the costs associated with gene expression. Adaptation is generally imagined to proceed in response to environmental changes (Takahata *et al.* 1975; Gillespie 1993), and thus by using a model with multiple environments—a low-noise environment and a high-noise environment—I was able to consider adaptation of a network that is already well-adapted to its original environment, but which can be improved by new mutations upon its introduction to a new environment. Considering only a modest environmental change and a network that was well-adapted to a previous environment seems more relevant than considering adaptation from an arbitrary starting point in a multi-dimensional fitness landscape, as real populations are always reasonably well-adapted to some environment, and successive environments are probably auto-correlated. Together, the model of a small network, the cost-benefit fitness function, and the influence of the

environment comprise the necessary components in order to consider whether adaptive reversals are plausible in a regulatory network. Using this model, I have shown that substitutions that have opposing effects with respect to a trait need not be compensatory; instead there is good reason to expect such substitutions may often be individually adaptive and constitute adaptive reversals.

There are a number of predictions based on experimental data suggesting that adaptive reversals involving protein evolution or protein-DNA co-evolution may be important. One such example involves the evolution of resistance to the antibiotic cefotaxime in *Escherichia coli*. Weinreich *et al.* (2006) determined that 18 of the 120 possible 5-step paths that confer resistance are strictly increasing in fitness. When paths are extended to include all 18 billion acyclic paths involving reversions, only nine of these are found to be strictly increasing in fitness (Depristo *et al.* 2007). These particular paths each involve at least two adaptive reversals at the nucleotide level because the start and end sequences differ at all five considered nucleotides, and thus each site that experiences a reversion must mutate a second time. As the authors note, these nine paths are a tiny fraction of the 18 billion possible; nonetheless, these additional paths involving adaptive reversals increase by half the total number of strictly increasing adaptive trajectories. Another example involves the co-evolution of *Escherichia coli lac* repressor and its operator binding site. Using a model for fitness based in part on a large experimental dataset of reporter-operator affinities, Poelwijk *et al.* (2006) find that if one restricts evolution to adaptive trajectories that maintain a minimum level of affinity between repressor and operator, many of the strictly adaptive trajectories involve what they refer to as diversions, which are actually adaptive reversals at the nucleotide level. In both of these studies, adaptive reversals are at the nucleotide level, whereas in this study, I consider adaptive reversals affecting higher-order functions, and thus the constituent substitutions may be distributed among sites. Further progress in understanding the functional basis of adaptation should shed more light on the prevalence of adaptive reversals in a variety of systems (Dean and Thornton 2007).

The existence of adaptive reversals has important implications for inferring the regime of selection under which a certain function or system has evolved. Focusing on a single trait at a time rather than the system in which it is embedded could mistakenly lead one to infer the trait to be evolving under stabilizing selection (via compensatory evolution), when it is actually one component of an adaptively-evolving system. Although it remains to be shown, the substitutions of opposing functional effect may even be helpful in detecting adaptation, when combined with data from other co-evolving traits. For example, when assessing whether an excess of functionally relevant substitutions have occurred on a particular lineage, rather than viewing functionally opposing substitutions as evidence of stabilizing selection on the one trait, one could incorporate them in the pool of evidence consistent with directional adaptive change of the overall system.

There are a number of ways in which the present model may depart from the biological reality. First, the mutational process employed in this study operates at the level of the trait rather than at the level of the sequence. This property is shared by Fisher’s Geometrical Model and related models (Fisher 1930; Martin and Lenormand 2006). While this mutational process has the advantage that assumptions need not be made about the functional encoding, it implies that increases and decreases of any effect size are always possible. For quantitative traits, these may not be unrealistic approximations (Hansen 2006); however, the traits modeled here may be more granular. Implicit in this mutational scheme is the absence of pleiotropy. That the production and degradation rates are separately encoded and thus not pleiotropic is likely a reasonable assumption; most regulatory elements are outside the processed transcript (Maston *et al.* 2006) and it is the transcript and resultant protein that most determine stability and degradation (Ross 1996; Garneau *et al.* 2007), although exceptions to this have been recently uncovered in yeast (Bregman *et al.* 2011; Depristo *et al.* 2007; Shalem *et al.* 2011; Trcek *et al.* 2011). It is less clear whether the activation threshold, k , is independent from the production rate of Y , β_y , as these both may be encoded in the *cis*-regulatory elements associated with Y . Pleiotropy could also exist with traits not modeled. For example, the three transcription factors may

participate in other regulatory circuits or have other functions, making otherwise adaptive mutations deleterious. Such pleiotropy can restrict the additive genetic variance available to selection and thereby limit adaptation (Walsh and Blows 2009). Although pleiotropy is clearly biologically important, excluding it in the present study is helpful for keeping the model simple.

I also assume the environment is constant after the initial change, and although also unrealistic, it is nonetheless a common assumption when modeling adaptation (Orr 2005), and it seems like a useful starting point. By assuming constancy of the environment, I ensure that adaptive reversals resulted from other, epistatic changes to the network, and not simply because the environment changed again. While it is certainly possible that selective pressures change rapidly, the present study is concerned with the plausibility of adaptive reversals due solely to the fitness landscape topology. In fully specifying the fitness function for the feed-forward loops considered, I needed to make some assumptions regarding the exact nature of the cost functions. Finally, in the adaptive walk simulations, I accept or reject each mutation based on the fixation probability, and thus successive substitutions are independent. This approach does not model certain population dynamics including the joint substitution of mutations or interference among simultaneously segregating mutations. These dynamics are likely to be of very minor importance here, and by deliberately excluding them, I ensure that the substitutions comprising the adaptive reversal are indeed individually beneficial.

Conclusions

I have shown how adaptive reversals can occur during the evolution of a particular regulatory motif, the type I coherent feed-forward loop. I chose to investigate this motif because its dynamics are accurately captured by a simple mathematical model (Mangan *et al.* 2003), it is prevalent in nature (Shen-Orr *et al.* 2002; Alon 2007), has an interpretable basis for fitness, and is relatively low-dimensional. I did not expect this particular model to have an unusual propensity to exhibit adaptive reversals. Additional

work will be needed to show whether adaptive reversals are a common feature of multidimensional evolution in other systems, and this will require appropriate, biologically-accurate models, which unfortunately remain rather few.

A necessary requirement for adaptive reversals is the existence of epistasis among beneficial mutations, which experimental data suggest may be common (Weinreich *et al.* 2006; MacLean *et al.* 2010). Furthermore, experimental evolution (e.g., Barrick *et al.* 2009; MacLean *et al.* 2010) may soon offer opportunities to test the predictions of this study. Several criteria must be met. First, the relevant pathways must be sufficiently understood and potential epistatic interactions identified. Second, the functional effects of observed substitutions must be measurable. Third, the fitness effects of individual and combinations of mutations must be measurable. Fourth, the population must evolve in an appropriate environment to drive adaptation of the pathway of interest. For select cases, all of these criteria can already be met or will likely be met imminently.

Acknowledgements

I would like to thank Molly Przeworski, Bin He, Ilya Ruvinsky, Hideki Innan, Hisashi Ohtsuki, Richard Hudson, Martin Kreitman, and John Reinitz for helpful discussions. I would also like to thank Yongtao Guan, B.H. and M.P. for suggestions on the manuscript. This work was supported by a William Rainey Harper Fellowship to K.B., an NSF EAPSI and JSPS Summer Fellowship to K.B., and an HHMI Early Career award to M.P.. The author declares that he has no conflicts of interest.

Table 1**Notation used for fitness function**

Trait	Description	Can Evolve	Low-noise Optimum^a	High-noise Optimum^{a,b}
α_y	Degradation rate of gene Y	Yes	1.31	0.96
β_y	Production rate of gene Y	Yes	1.31	0.68
α_z	Degradation rate of gene Z	Yes	1.18	1.26
β_z	Production rate of gene Z	Yes	45.6	44.5
k	Y threshold required to express Z	Yes	0.04	0.16
δ	Benefit per unit of Z	No	5	-
η_y	Cost per unit of Y	No	0.4	-
η_z	Cost per unit of Z	No	2	-
c_F	Frequency of false signals	No	0.05	0.3
c_T	Frequency of true signals	No	0.15	-
t_F	Maximum false signal duration	No	0.5	-
t_T	Maximum true signal duration	No	4	-
τ	Delay before Z expression ^c			

^aFor the evolvable traits, values given are those producing the highest fitness network in the given environment. For the other parameters (i.e., the non-evolvable ones), the values given define the environment and cost function.

^bA dash indicates that the value of the low-noise environment is also used for the high-noise environment.

^c τ is a function of α_y , β_y , and k (see Methods).

Table 2

Reversals during adaptation from random starting points with discrete-sized mutational effects

Trait	Reversals ^a	Forward portion			Reversal portion		
		Steps ^b	median Ns^c	median Δ^d	Steps ^b	median Ns^c	median Δ^d
α_y	6	1.0	94.4	0.567	4.0	7.1	1.032
β_y	93	1.9	64.6	1.574	2.1	6.2	0.750
α_z	82	1.9	105.2	1.809	1.9	8.4	0.510
β_z	136	1.6	100.4	1.882	1.7	9.5	0.527
k	38	1.9	31.3	0.831	1.4	7.5	0.404

^aReversals observed out of 1000 simulations from random starting points conditional on fitness > 0 . Adaptation is in the high-noise environment of Figure 5. Only certain reversals are counted: ones not including deleterious mutations, not overshooting an intervening optimum, and including substitutions with $Ns > 5$ on both initial and reverse portions.

^bAverage number of mutations affecting the trait on the indicated portion of the adaptive reversal.

^cMedian population selection coefficient for all substitutions affecting the trait on the indicated portion of the adaptive reversal.

^dMedian change in the trait observed over the reversal portion, normalized by average change across all runs.

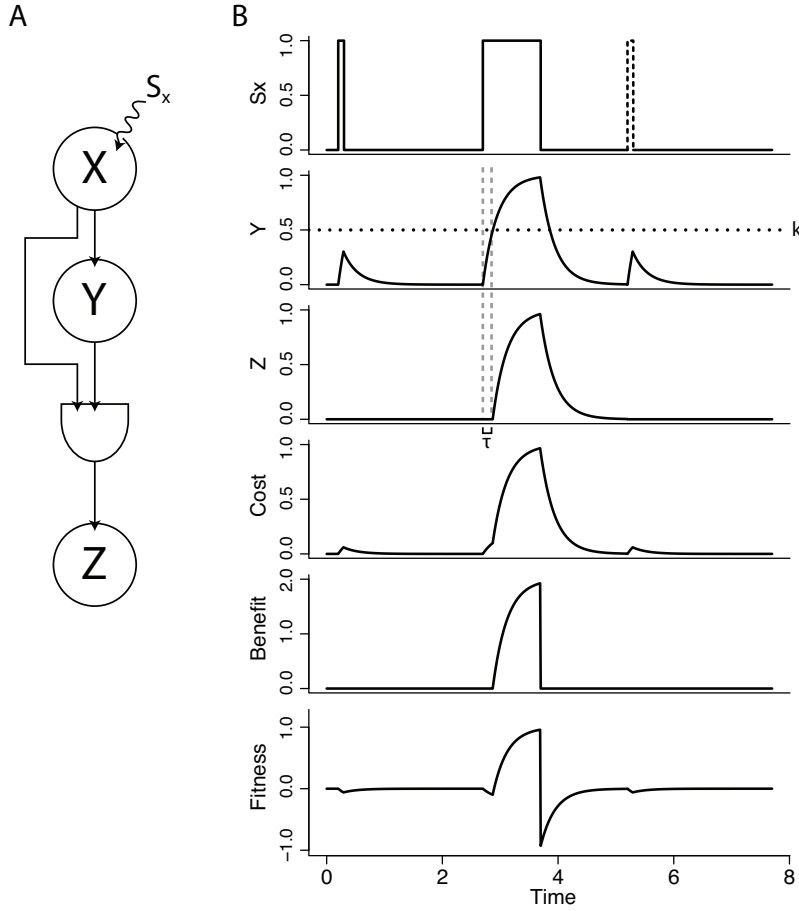


Figure 1.—Layout and dynamic behavior of a type I coherent feed-forward loop. (A) Layout of the feed-forward loop. Gene X is activated by an external signal, S_X , and in its active state induces genes Y and Z . Gene Y also drives the expression of Z . The inputs to Z are combined via an AND-type logical operation whereby Z is only induced when X is active and Y is sufficiently expressed. (B) Dynamical behavior and cost-benefit basis for fitness. Each plot shows responses to three signals as a function of time (horizontal axis). Top row: There are two true signals and one false signal (dashed). Second row: Gene Y is induced when X is active, but it must reach a threshold k before it is sufficiently expressed to activate Z . Third row: Gene Z is only induced when Y is above k , although at a delay τ relative to Y . Fourth row: Both Y and Z have costs proportional to their expression, with Z having a much larger cost, as it encapsulates the cost of activating the pathway downstream of Z . Fifth row: A benefit proportional to Z expression is derived from the pathway downstream of Z , but only in the presence of a true signal. Bottom row: The fitness rate is: benefit - cost. Total fitness (see the Methods), is an integral over time for each signal and also an integral over the signal duration distributions for true and false signals. This figure follows the illustration conventions of Alon (2006), adjusted to reflect minor modeling differences.

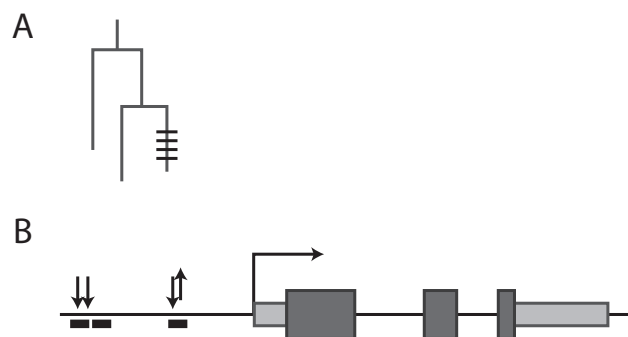


Figure 2.—Single-gene viewpoint of nucleotide substitutions. (A) Cartoon phylogeny showing a cluster of four substitutions on a single lineage. (B) Cartoon of a gene illustrating the same four substitutions as in (A). Arrows indicate the direction of the marginal functional effect of each substitution on some aspect of gene expression. Three substitutions reduce expression and one substitution increases it. In this cartoon, all four substitutions affect regions bound by transcription factors (black boxes). Despite detailed functional knowledge, observing only this one gene in isolation without the ability to consider epistatic interactions with other genes, limits the interpretations regarding mutations of opposing functional effect.

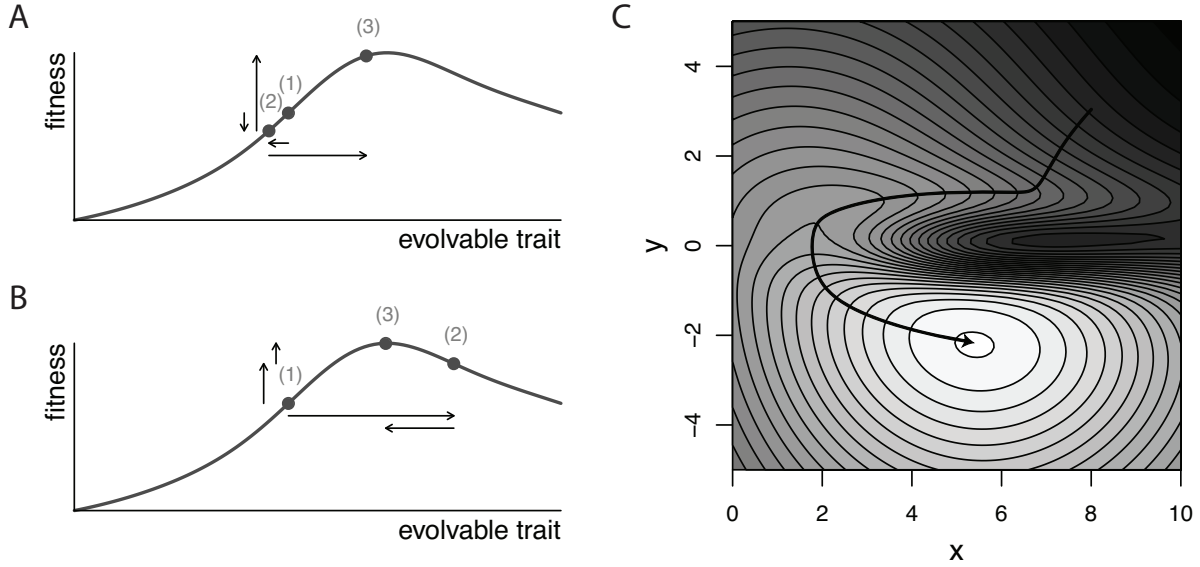


Figure 3.—Illustration of scenarios showing how mutations of opposing functional effect can be related to smooth fitness landscapes in one (A-B) and two (C) dimensions. (A) Compensatory evolution on a fitness landscape involving one evolvable trait. Points indicate three allelic values with the starting point of the adaptive walk at (1) followed by a deleterious substitution to allele (2) and a beneficial (over-compensating) substitution to allele (3). Horizontal arrows indicate the opposing functional changes while vertical arrows indicate the opposing changes in fitness. (B) Overshooting the optimum. Notation is as in (A), however in this example, despite opposing functional effects, the substitutions are individually adaptive. (C) Adaptation on a two-dimensional fitness landscape. Two continuous traits, x and y , comprise the plane, while fitness is shown as a shaded topology with contour lines. The adaptive trajectory in the small mutation size (SMS) approximation, i.e., optimal hill-climbing path, is shown as the black curved arrow. Notice that with respect to trait x , the adaptive trajectory bends back such that initially decreasing x is adaptive and later increasing x is adaptive. Fitness is a carefully selected, but arbitrary function chosen for illustrative purposes: $\text{fitness}(x, y) = -\frac{y}{4} + e^{-\frac{(x-4)^2 + y^2}{50}} \left(10 + \frac{x(1-e^y)}{1+e^y} - 10e^{-2y^2} \left(\frac{e^{-\frac{(x-4.5)^2}{100}}}{1+e^{4-x}} \right) \right)$.

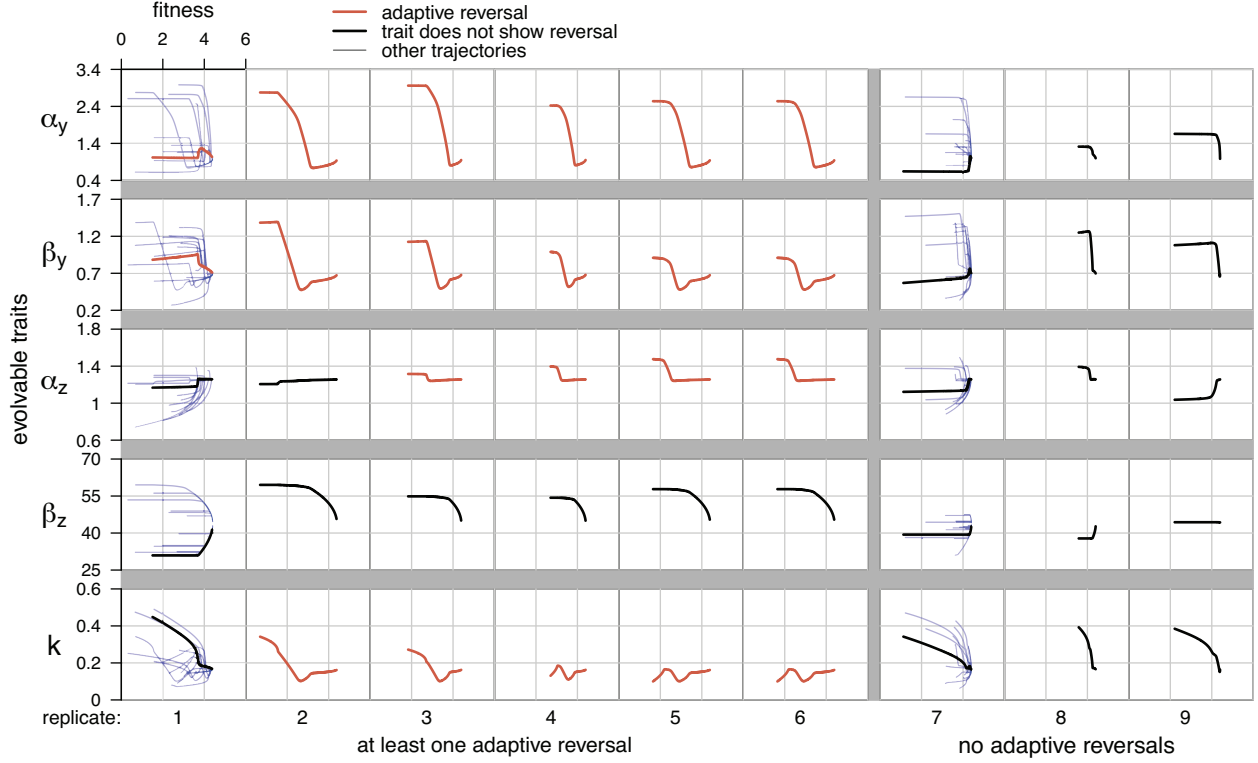


Figure 4.—Adaptive reversals occur solely as a result of the fitness landscape topology. Each column highlights one adaptive trajectory (thick lines) projected onto each of the five trait dimension (rows). The adaptive trajectory is a function of the evolvable trait (vertical axis) and fitness (horizontal axis). Given fitness is guaranteed to increase with time, adaptation always proceeds from left to right within each small plot. The adaptive trajectories shown correspond to evolution starting from points sampled uniformly from a five-dimensional hypercube. Several other adaptive trajectories are shown in the background (blue lines) for replicates 1 and 7 for reference and to illustrate how adaptation always converges to the same optimal point in the fitness landscape. The first six columns (from the left) exemplify adaptive trajectories with at least one adaptive reversal (red curves). The last three columns show adaptive trajectories in which none of the traits undergo adaptive reversals. All trajectories were numerically evaluated according to the small mutation size (SMS) approximation (see Methods).

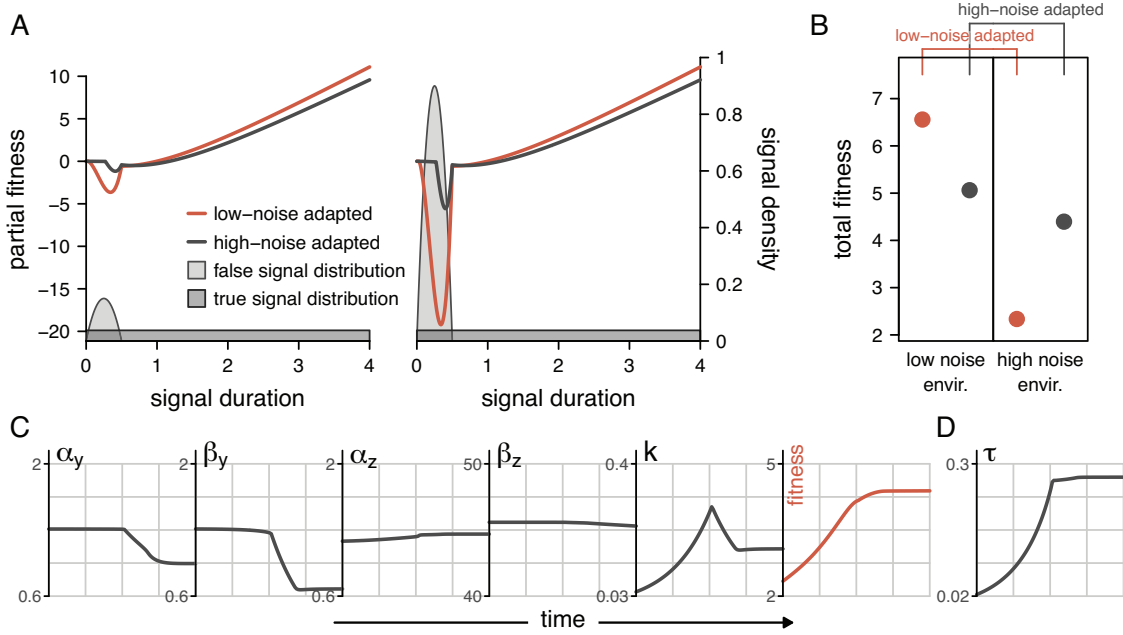


Figure 5.—Adaptation to a high-noise environment of a network originally adapted to a low-noise environment. (A) The two environments are defined by the signal distributions (shaded gray curves, right vertical axis). False signals are generally short (quadratic, light gray curve), with none longer than 0.5. True signals have a uniform distribution (dark gray curve). The low-noise environment (left plot) has a much smaller density of false signals than does the high-noise environment (right plot). Partial fitness (left vertical axis), defined as the net fitness for a single signal, is shown as a function of the duration of the signal in each of two environments for each of two networks: the one initially adapted to the low-noise environment (red curve), and the one that evolves as a result of adaptation to the high-noise environment (gray curve). As can be seen, in the high-noise environment (right plot), the network adapted to the low-noise environment (red curve) suffers a major fitness cost for short signal durations, as these tend to be false signals, which this network fails to filter as effectively as the high-noise-adapted network. (B) Total fitnesses of each network in each environment. (C) Evolution of a trait (vertical axis) as a function of time (horizontal axis, same for each plot) for each of the five evolvable traits (first five plots). See the Methods for the interpretation of time. The sixth plot shows the increase in fitness during adaptation to the high-noise environment. (D) Evolution of the delay, τ . The delay is a function of other parameters: $\tau = -\frac{1}{\alpha_y} \log(1 - k \frac{\alpha_y}{\beta_y})$.

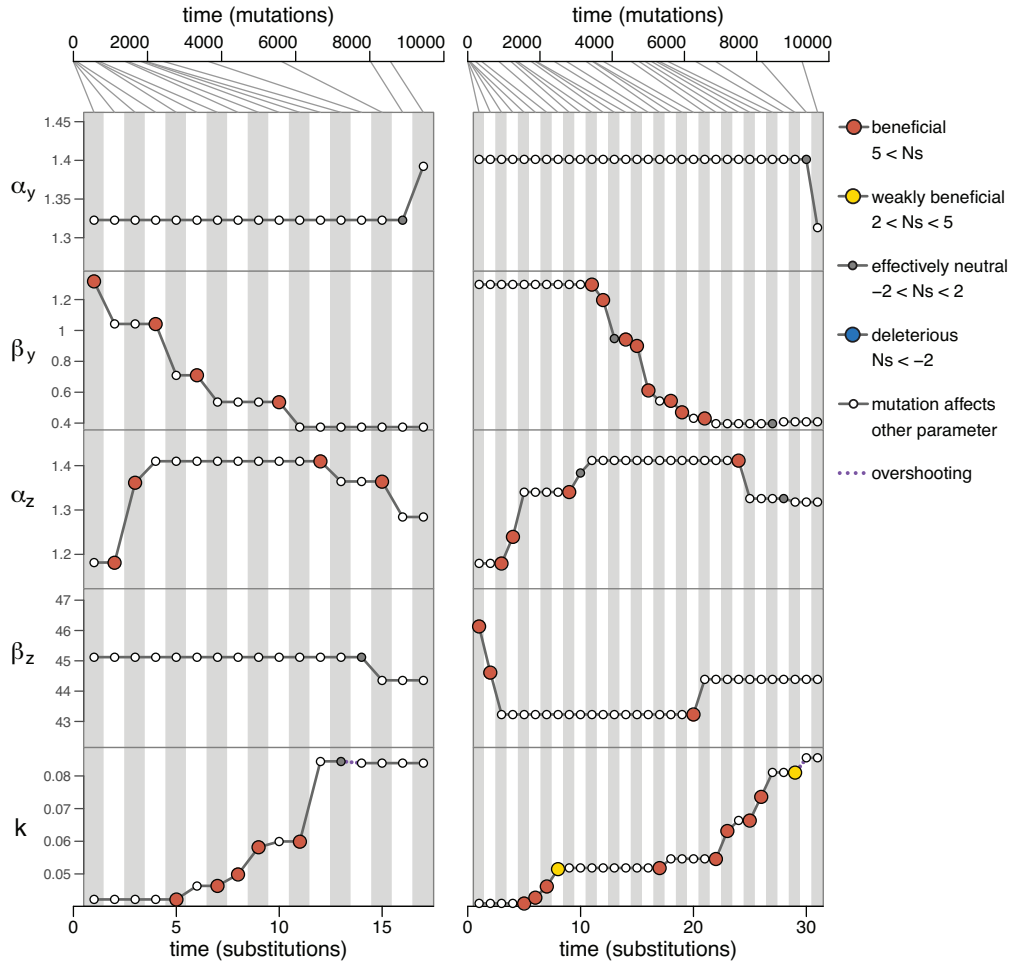


Figure 6.—Adaptation to an environmental shift with discrete mutational effects. Two replicate bouts of adaptation are shown, starting from a network adapted to a low-noise environment, and evolving in response to a shift to a high-noise environment. Each small plot shows the evolution of one trait (vertical axes) as a function of time (horizontal axes). Vertical stripes indicate the accumulation of substitutions with the bottom axis giving the number of accumulated substitutions and the top axis indicating when the substituted mutation occurred among all mutations (most of which are lost). A mutation only affects a single trait. All substitutions are indicated on every plot, but only the colored circles correspond substitutions that affect the given trait. Thus each column features exactly one colored circle. Substitutions are colored according to the population selection coefficient, Ns (see legend; $N = 1000$). Purple dotted lines connecting substitutions indicate that the substitution overshoot an intervening optimum. Solid lines indicate that overshooting an optimum did not occur. Adaptive reversals not overshooting an optimum can be seen in both simulations for trait α_z and in the right-hand simulation for β_z .

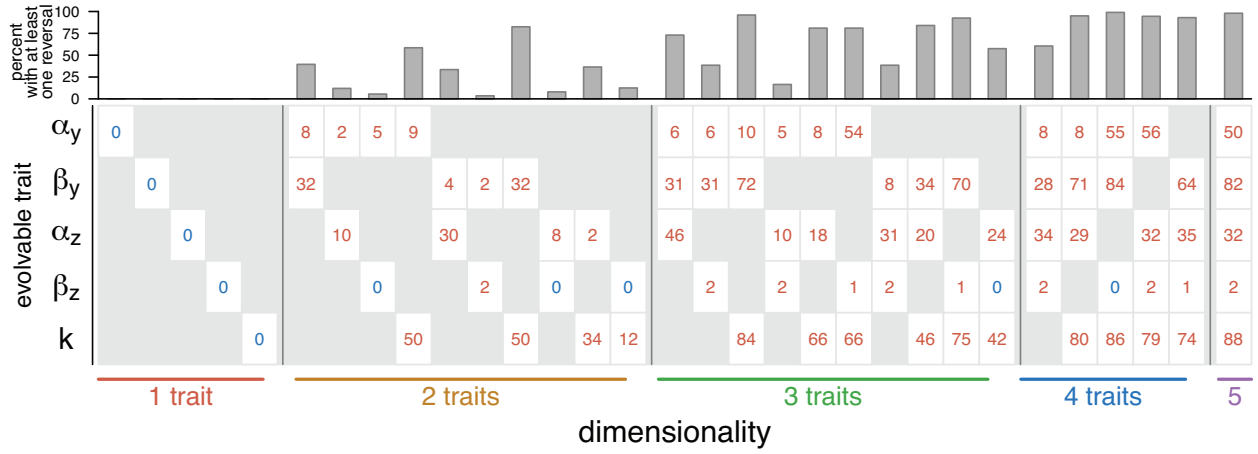


Figure 7.—Prevalence of reversals depends on what traits are evolvable and scales with dimensionality. In addition to allowing all five traits to evolve (right-most column, purple label), evolution is restricted to all possible non-empty subsets of the five traits. There are five possible one-trait subsets (left set, red label), ten possible two-trait subsets (second to left, orange label), and so on, arrayed along the horizontal axis. This allows consideration of how the prevalence of reversals scales with dimensionality or how evolution would proceed if the mutation target for one or more traits is nearly zero relative to the other traits. 200 random starting points were sampled from the same hypercube as used in Figure 4. Numbers indicate percentages of adaptive trajectories showing a reversal with respect to the trait. The fixed traits are set to the trait value that is optimal in the high noise environment. The bar plot above (gray bars) indicates the percentage of replicate simulations in which there is at least one reversal observed for that trait subset. All tests for reversals are done as in Figure 4—the SMS gradient ascent path is computed numerically for each starting point and the high-noise environment is used.

LITERATURE CITED

- Alon, U., 2003 Biological networks: the tinkerer as an engineer. *Science* *301*(5641): 1866–1867.
- Alon, U., 2006 *An Introduction to Systems Biology: Design Principles of Biological Circuits* (1 ed.). Chapman and Hall/CRC.
- Alon, U., 2007 Network motifs: theory and experimental approaches. *Nat Rev Genet* *8*(6): 450–461.
- Ancel, L. W. and W. Fontana, 2000 Plasticity, evolvability, and modularity in RNA. *J Exp Zool* *288*(3): 242–283.
- Andolfatto, P., 2005 Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* *437*(7062): 1149–1152.
- Barrett, R. D. H. and H. E. Hoekstra, 2011 Molecular spandrels: tests of adaptation at the genetic level. *Nat Rev Genet* *12*(11): 767–780.
- Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh et al., 2009 Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* *461*(7268): 1243–1247.
- Barrière, A., K. L. Gordon, and I. Ruvinsky, 2011 Distinct functional constraints partition sequence conservation in a cis-regulatory element. *PLoS Genet* *7*(6): e1002095.
- Bregman, A., M. Avraham-Kelbert, O. Barkai, L. Duek, A. Guterman et al., 2011 Promoter elements regulate cytoplasmic mRNA decay. *Cell* *147*(7): 1473–1483.
- Bullaughey, K., 2011 Changes in selective effects over time facilitate turnover of enhancer sequences. *Genetics* *187*(2): 567–582.
- Callahan, B., R. A. Neher, D. Bachtrog, P. Andolfatto, and B. I. Shraiman, 2011 Correlated evolution of nearby residues in *Drosophilid* proteins. *PLoS Genet* *7*(2): e1001315.

- Chen, S., Y. E. Zhang, and M. Long, 2010 New genes in *Drosophila* quickly become essential. *Science* *330*(6011): 1682–1685.
- Chen, Y., D. B. Carlini, J. F. Baines, J. Parsch, J. M. Braverman et al., 1999 RNA secondary structure and compensatory evolution. *Genes & Genetic Systems* *74*(6): 271–286.
- Cheverud, J. M. and E. J. Routman, 1996 Epistasis as a Source of Increased Additive Genetic Variance at Population Bottlenecks. *Evolution* *50*(3): 1042–1051.
- Dean, A. M. and J. W. Thornton, 2007 Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* *8*(9): 675–688.
- Dekel, E. and U. Alon, 2005 Optimality and evolutionary tuning of the expression level of a protein. *Nature* *436*(7050): 588–592.
- Dekel, E., S. Mangan, and U. Alon, 2005 Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Physical biology* *2*(2): 81–88.
- DePristo, M., D. Weinreich, and D. Hartl, 2005 Missense meanderings in sequence space: A biophysical view of protein evolution. *Nat Rev Genet* *6*(9): 678–687.
- Depristo, M. A., D. L. Hartl, and D. M. Weinreich, 2007 Mutational reversions during adaptive protein evolution. *Mol Biol Evol* *24*(8): 1608–1610.
- Durrett, R. and D. Schmidt, 2008 Waiting for two mutations: with applications to regulatory sequence evolution and the limits of Darwinian evolution. *Genetics* *180*(3): 1501–1509.
- Ewens, W. J., 2004 *Mathematical population genetics* (2nd ed.). New York: Springer.
- Field, Y., Y. Fondufe-Mittendorf, I. K. Moore, P. Mieczkowski, N. Kaplan et al., 2009 Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nat Genet* *41*(4): 438–445.
- Fisher, R. A., 1930 *The genetical theory of natural selection*. Oxford: The Clarendon Press.

- Gandon, S. and S. P. Otto, 2007 The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis. *Genetics* **175**(4): 1835–1853.
- Garneau, N. L., J. Wilusz, and C. J. Wilusz, 2007 The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* **8**(2): 113–126.
- Gavrilets, S., 2004 *Fitness landscapes and the origin of species*. Princeton Univ Pr.
- Gavrilets, S., H. Li, and M. D. Vose, 1998 Rapid parapatric speciation on holey adaptive landscapes. *Proc Biol Sci* **265**(1405): 1483–1489.
- Gerke, J., K. Lorenz, and B. Cohen, 2009 Genetic interactions between transcription factors cause natural variation in yeast. *Science* **323**(5913): 498–501.
- Gibson, G., 1996 Epistasis and pleiotropy as natural properties of transcriptional regulation. *Theor Popul Biol* **49**(1): 58–89.
- Gillespie, J. H., 1993 Substitution processes in molecular evolution. I. Uniform and clustered substitutions in a haploid model. *Genetics* **134**(3): 971–981.
- Gillespie, J. H., 1994 *The Causes of Molecular Evolution*. Oxford University Press, USA.
- Haag, E. S., 2007 Compensatory vs. pseudocompensatory evolution in molecular and developmental interactions. *Genetica* **129**(1): 45–55.
- Hansen, T. F., 2006 The evolution of genetic architecture. *Annual Review of Ecology Evolution and Systematics* **37**: 123–157.
- Hartwell, L. H., J. J. Hopfield, S. Leibler, and A. W. Murray, 1999 From molecular to modular cell biology. *Nature* **402**(6761 Suppl): C47–52.
- He, B. Z., A. K. Holloway, S. J. Maerkl, and M. Kreitman, 2011 Does positive selection drive transcription factor binding site turnover? A test with *Drosophila* cis-regulatory modules. *PLoS Genet* **7**(4): e1002053.
- Heger, A. and L. Holm, 2003 Exhaustive enumeration of protein domain families. *J Mol Biol* **328**(3): 749–767.

- Holloway, A. K., D. J. Begun, A. Siepel, and K. S. Pollard, 2008 Accelerated sequence divergence of conserved genomic elements in *Drosophila melanogaster*. *Genome Res* **18**(10): 1592–1601.
- Iwasa, Y., F. Michor, and M. A. Nowak, 2004 Stochastic tunnels in evolutionary dynamics. *Genetics* **166**(3): 1571–1579.
- Kalir, S., S. Mangan, and U. Alon, 2005 A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. *Mol Syst Biol* **1**: 2005.0006.
- Kaplan, J., 2008 The end of the adaptive landscape metaphor? *Biology and Philosophy* **23**: 625–638.
- Kauffman, S. and S. Levin, 1987 Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol* **128**(1): 11–45.
- Kenigsberg, E., A. Bar, E. Segal, and A. Tanay, 2010 Widespread compensatory evolution conserves DNA-encoded nucleosome organization in yeast. *PLoS Comput Biol* **6**(12): e1001039.
- Kim, S. Y. and J. K. Pritchard, 2007 Adaptive evolution of conserved noncoding elements in mammals. *PLoS Genet* **3**(9): 1572–1586.
- Kimura, M., 1985 The role of compensatory neutral mutations in molecular evolution. *Journal of Genetics* **64**(1): 7–19.
- Kirchhamer, C. V., C. H. Yuh, and E. H. Davidson, 1996 Modular cis-regulatory organization of developmentally expressed genes: two genes transcribed territorially in the sea urchin embryo, and additional examples. *Proc Natl Acad Sci USA* **93**(18): 9322–9328.
- Kreitman, M., 1996 The neutral theory is dead. Long live the neutral theory. *BioEssays* **18**(8): 678–83; discussion 683.

- Kuo, D., K. Licon, S. Bandyopadhyay, R. Chuang, C. Luo et al., 2010 Coevolution within a transcriptional network by compensatory trans and cis mutations. *Genome Res* *20*(12): 1672–1678.
- Landry, C. R., P. J. Wittkopp, C. H. Taubes, J. M. Ranz, A. G. Clark et al., 2005 Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics* *171*(4): 1813–1822.
- Lenormand, T., D. Roze, and F. Rousset, 2009 Stochasticity in evolution. *Trends Ecol Evol (Amst)* *24*(3): 157–165.
- Ludwig, M. Z., C. Bergman, N. H. Patel, and M. Kreitman, 2000 Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* *403*(6769): 564–567.
- Lunzer, M., G. B. Golding, and A. M. Dean, 2010 Pervasive cryptic epistasis in molecular evolution. *PLoS Genet* *6*(10): e1001162.
- Lusk, R. W. and M. B. Eisen, 2010 Evolutionary Mirages: Selection on Binding Site Composition Creates the Illusion of Conserved Grammars in *Drosophila* Enhancers. *PLoS Genet* *6*(1): e1000829.
- Lynch, M., 2007 The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci USA* **104 Suppl 1**: 8597–8604.
- Macía, J., R. V. Solé, and S. F. Elena, 2012 The causes of epistasis in genetic networks. *Evolution* *66*(2): 586–596.
- MacLean, R. C., G. G. Perron, and A. Gardner, 2010 Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics* *186*(4): 1345–1354.
- Mangan, S., A. Zaslaver, and U. Alon, 2003 The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol* *334*(2): 197–204.
- Martin, G., S. F. Elena, and T. Lenormand, 2007 Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat Genet* *39*(4): 555–560.

- Martin, G. and T. Lenormand, 2006 A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* *60*(5): 893–907.
- Maston, G. A., S. K. Evans, and M. R. Green, 2006 Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* *7*: 29–59.
- Moses, A. M., D. A. Pollard, D. A. Nix, V. N. Iyer, X.-Y. Li et al., 2006 Large-scale turnover of functional transcription factor binding sites in *Drosophila*. *PLoS Comput Biol* *2*(10): e130.
- Orr, H. A., 2000 Adaptation and the cost of complexity. *Evolution* *54*(1): 13–20.
- Orr, H. A., 2005 The genetic theory of adaptation: a brief history. *Nat Rev Genet* *6*(2): 119–127.
- Orr, H. A., 2006 The population genetics of adaptation on correlated fitness landscapes: the block model. *Evolution* *60*(6): 1113–1124.
- Perfeito, L., S. Ghozzi, J. Berg, K. Schnetz, and M. Lässig, 2011 Nonlinear fitness landscape of a molecular pathway. *PLoS Genet* *7*(7): e1002160.
- Phillips, P. C., 2008 Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* *9*(11): 855–867.
- Poelwijk, F. J., D. J. Kiviet, and S. J. Tans, 2006 Evolutionary potential of a duplicated repressor-operator pair: simulating pathways using mutation data. *PLoS Comput Biol* *2*(5): e58.
- Pollard, K. S., S. R. Salama, B. King, A. D. Kern, T. Dreszer et al., 2006 Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* *2*(10): e168.
- Ptashne, M., 1988 How eukaryotic transcriptional activators work. *Nature* *335*(6192): 683–689.
- Rosen, R., 1967 Optimality principles in biology. Butterworth.

- Ross, J., 1996 Control of messenger RNA stability in higher eukaryotes. *Trends in Genetics* *12*(5): 171–175.
- Salverda, M. L. M., E. Dellus, F. A. Gorter, A. J. M. Debets, J. van der Oost et al., 2011 Initial mutations direct alternative pathways of protein evolution. *PLoS Genet* *7*(3): e1001321.
- Schlosser, G. and G. P. Wagner, 2004 *Modularity in development and evolution*. University Of Chicago Press.
- Shalem, O., B. Groisman, M. Choder, O. Dahan, and Y. Pilpel, 2011 Transcriptome kinetics is governed by a genome-wide coupling of mRNA production and degradation: a role for RNA Pol II. *PLoS Genet* *7*(9): e1002273.
- Shen-Orr, S. S., R. Milo, S. Mangan, and U. Alon, 2002 Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* *31*(1): 64–68.
- Skipper, R. A. J., 2004 The Heuristic Role of Sewall Wright’s 1932 Adaptive Landscape Diagram. *Philosophy of Science* *71*(5): 1176–1188.
- Stephan, W., 1996 The rate of compensatory evolution. *Genetics* *144*(1): 419–426.
- Takahasi, K. R. and F. Tajima, 2005 Evolution of coadaptation in a two-locus epistatic system. *Evolution* *59*(11): 2324–2332.
- Takahata, N., K. Ishii, and H. Matsuda, 1975 Effect of temporal fluctuation of selection coefficient on gene frequency in a population. *Proc Natl Acad Sci USA* *72*(11): 4541–4545.
- Trcek, T., D. R. Larson, A. Moldón, C. C. Query, and R. H. Singer, 2011 Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. *Cell* *147*(7): 1484–1497.
- Tuch, B. B., D. J. Galgoczy, A. D. Hernday, H. Li, and A. D. Johnson, 2008 The evolution of combinatorial gene regulation in fungi. *Plos Biol* *6*(2): e38.

- Wagner, G. P., M. Pavlicev, and J. M. Cheverud, 2007 The road to modularity. *Nat Rev Genet* **8**(12): 921–931.
- Walsh, B. and M. W. Blows, 2009 Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology Evolution and Systematics* **40**(1): 41–59.
- Weinreich, D. M., N. F. Delaney, M. A. Depristo, and D. L. Hartl, 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**(5770): 111–114.
- Weinreich, D. M., R. A. Watson, and L. Chao, 2005 Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**(6): 1165–1174.
- Whitlock, M. C., P. C. Phillips, F. B.-G. Moore, and S. J. T. Tonsor, 1995 Multiple Fitness Peaks and Epistasis. *Annual Review of Ecology and Systematics* **26**: 601–629.
- Wolf, J. B., E. D. Brodie, and M. J. Wade, 2000 *Epistasis and the evolutionary process*. Oxford: Oxford University Press, USA.
- Wright, S., 1932 The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress on Genetics* **1**: 356–366.
- Yukilevich, R., J. Lachance, F. Aoki, and J. R. True, 2008 Long-term adaptation of epistatic genetic networks. *Evolution* **62**(9): 2215–2235.