On the incongruence of genotype-phenotype and fitness landscapes

Malvika Srivastava[1,2] **and Joshua L. Payne**[1,2,*]

[1]Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland

[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland

[*]joshua.payne@env.ethz.ch

ABSTRACT

The mapping from genotype to phenotype to fitness typically involves multiple nonlinearities that can transform the effects of mutations. For example, mutations may contribute additively to a phenotype, but their effects on fitness may combine non-additively because selection favors a low or intermediate value of that phenotype. This can cause incongruence between the topographical properties of a fitness landscape and its underlying genotype-phenotype landscape. Yet, genotype-phenotype landscapes are often used as a proxy for fitness landscapes to study the dynamics and predictability of evolution. Here, we use theoretical models and empirical data on transcription factor-DNA interactions to systematically study the incongruence of genotype-phenotype and fitness landscapes when selection favors a low or intermediate phenotypic value. Using the theoretical models, we prove a number of fundamental results. For example, selection for low or intermediate phenotypic values does not change simple sign epistasis into reciprocal sign epistasis, yet it changes reciprocal sign epistasis into simple sign epistasis and no sign epistasis with equal probability. More broadly, we show that such selection tends to create fitness landscapes that are more rugged than the underlying genotype-phenotype landscape, but this increased ruggedness typically does not frustrate adaptive evolution because the local adaptive peaks in the fitness landscape tend to be nearly as tall as the global peak. Many of these results carry forward to the empirical genotype-phenotype landscapes, which may help to explain why low- and intermediate-affinity transcription factor-DNA interactions are so prevalent in eukaryotic gene regulation.

AUTHOR SUMMARY

How do mutations change phenotypic traits and organismal fitness? This question is often addressed in the context of a classic metaphor of evolutionary theory — the fitness landscape. A fitness landscape is akin to a physical landscape, in which genotypes define spatial coordinates, and fitness defines the elevation of each coordinate. Evolution then acts like a hill-climbing process, in which populations ascend fitness peaks as a consequence of mutation and selection. It is becoming increasingly common to construct such landscapes using experimental data from high-throughput sequencing technologies and phenotypic assays, in systems such as macromolecules and gene regulatory circuits. Although these landscapes are typically defined by molecular phenotypes, and are therefore more appropriately referred to as genotype-phenotype landscapes, they are often used to study evolutionary dynamics. This requires the assumption that the molecular phenotype is a reasonable proxy for fitness, which need not be the case. For example, selection may favor a low or intermediate phenotypic value, causing incongruence between a fitness landscape and its underlying genotype-phenotype landscape. Here, we study such incongruence using a diversity of theoretical models and experimental data from gene regulatory systems. We regularly find incongruence, in that fitness landscapes tend to comprise more peaks than their underlying genotype-phenotype landscapes. However, using evolutionary simulations, we show that this increased ruggedness need not impede adaptation.

Introduction

Characterizing the relationship between genotype and phenotype is key to our understanding of evolution[1,2]. For quantitative phenotypes, such as the expression level of a gene or the enzymatic activity of a protein, this relationship can be formalized as a genotype-phenotype landscape[3]. In such a landscape, genotypes represent coordinates in an abstract genotype space and their phenotype defines the elevation of each coordinate in this space[4]. The topographical properties of genotype-phenotype landscapes, such as their ruggedness, are influenced by epistasis[5] — non-additive interactions between mutations in their contribution to phenotype. These topographical properties have important evolutionary consequences, because they determine how mutation brings forth the phenotypic variation upon which selection acts[6,7].

Technological advances are facilitating the construction and analysis of empirical genotype-phenotype landscapes at ever-increasing resolution, scale, and scope[8]. Example phenotypes include the enzymatic activity[9], binding affinity[10], allosteric profile[11], and fluorescence intensity of proteins[12], as well as exon inclusion levels[3], the expression levels of genes driven by regulatory elements[13], the expression patterns of gene regulatory circuits[14–16], and flux through metabolic pathways[17]. In analyzing the topographical properties of these landscapes and their evolutionary consequences, an assumption is often made that phenotype is a proxy for fitness[12,18–24], thus rendering genotype-phenotype landscapes equivalent to fitness landscapes[25]. While this assumption may be justified under certain conditions, such as in directed protein evolution experiments[26], it is often the case that the relationship between phenotype and fitness is not so straightforward. For example, fitness may depend upon more phenotypes than those being assayed[27] or the relationship between phenotype and fitness may be inherently nonlinear, for example reflecting a tradeoff between the costs and benefits associated with a phenotype[28]. In the latter case, selection may favor a low or intermediate phenotypic value[29,30]; e.g., an intermediate gene expression level,[31–33], enzyme efficiency[34] or protein production rate or activity[35]. Such non-linearities are a cause of epistasis[16,36–38], and they can transform the effects of mutations as they map onto phenotype and fitness[37], thus rendering the topographical properties of a fitness landscape qualitatively different from those of its underlying genotype-phenotype landscape (Fig. 1A).

Selection for low or intermediate phenotypic values is especially relevant to transcription factor-DNA interactions[39]. Transcription factors are sequence-specific DNA binding proteins that help regulate gene expression. They do so by binding DNA sequences (transcription factor binding sites) in regulatory regions such as promoters and enhancers to recruit or block the recruitment of RNA polymerase[40]. The regulatory effect of such a binding event depends in part on the affinity with which the DNA sequence is bound by the transcription factor[41,42]. As such, binding affinity is an important molecular phenotype of transcription factor binding sites, upon which selection acts[43,44]. While it is commonly assumed that selection increases binding affinity[21,43–47], several lines of evidence suggest that selection for low or intermediate binding affinity also influences the evolution of transcription factor binding sites. For example, paralogous transcription factors often bind the same DNA sequences with high affinity, but different DNA sequences with low affinity[48,49]. If an optimal gene expression pattern requires binding by just one of several transcription factor paralogs, such specificity can be achieved using low-affinity transcription

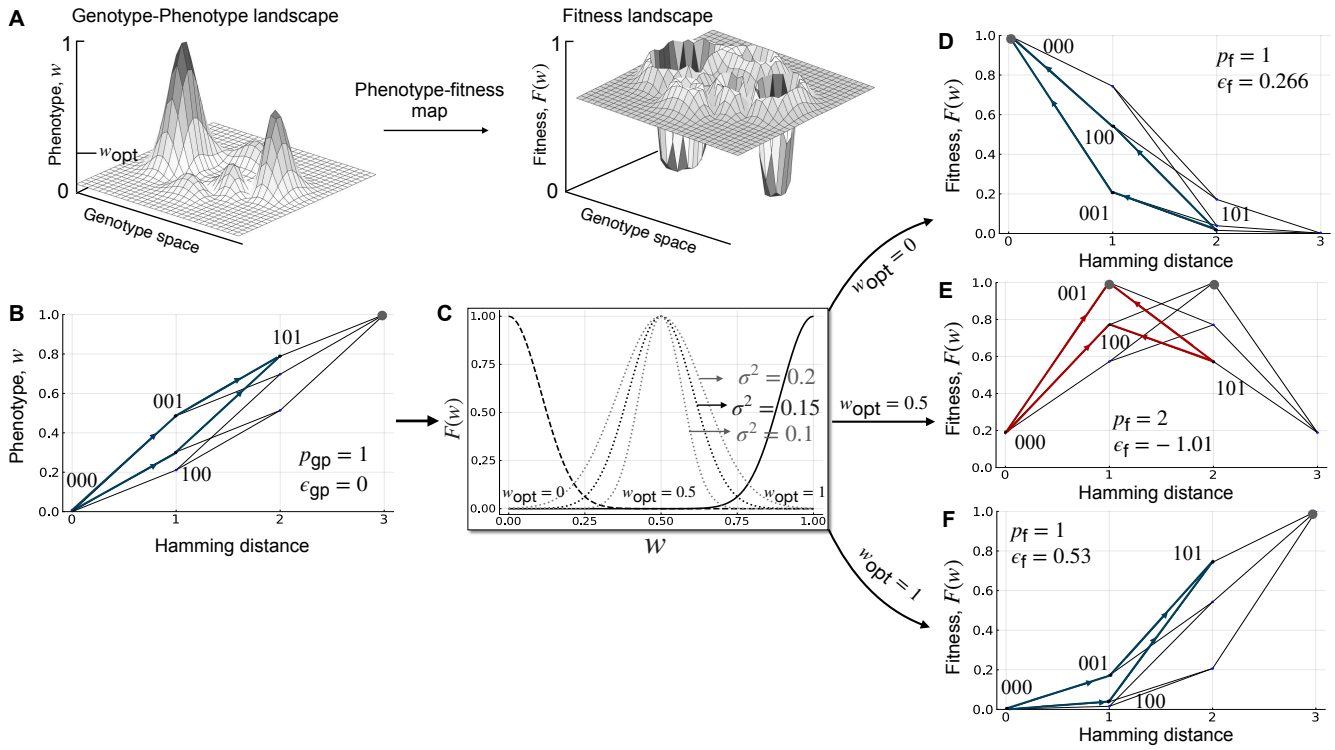


Figure 1. Incongruence. (A) Schematic illustration of how selection for an intermediate phenotypic value w_{opt} can make a genotype-phenotype landscape incongruent with the resulting fitness landscape. (B) An additive three-locus, biallelic genotype-phenotype landscape with a single peak (gray filled circle; $p_{\text{gp}} = 1$). One pairwise interaction is highlighted in blue. It exhibits no magnitude epistasis ($\varepsilon_{\text{gp}} = 0$) or sign epistasis. (C) The Gaussian phenotype-fitness map (Eq. 1) is shown for three values of w_{opt} (dashed line, $w_{\text{opt}} = 0$; dotted line, $w_{\text{opt}} = 0.5$; solid line, $w_{\text{opt}} = 1$), with three values of σ shown for $w_{\text{opt}} = 0.5$. (D) Applying the Gaussian phenotype-fitness map with $w_{\text{opt}} = 0$ to the genotype-phenotype landscape results in a single-peaked fitness landscape (gray filled circle; $p_{\text{f}} = 1$). The same pairwise interaction from (B) is highlighted in blue. It exhibits positive epistasis ($\varepsilon_{\text{gp}} = 0.266$), but no sign epistasis. (E) Applying the Gaussian phenotype-fitness map with $w_{\text{opt}} = 0.5$ to the genotype-phenotype landscape results in a multi-peaked fitness landscape (gray filled circles; $p_{\text{f}} = 2$). The same pairwise interaction from (B) is highlighted in red. It exhibits negative epistasis ($\varepsilon_{\text{gp}} = -1.01$), as well as reciprocal sign epistasis. (F) Applying the Gaussian phenotype-fitness map with $w_{\text{opt}} = 1$ to the genotype-phenotype landscape results in a single-peaked fitness landscape (gray filled circle; $p_{\text{f}} = 1$). The same pairwise interaction from (A) is highlighted in blue. It exhibits positive epistasis ($\varepsilon_{\text{gp}} = 0.53$), but no sign epistasis. In panels B,D-F, arrows point from genotypes with lower phenotypic or fitness values to genotypes with higher phenotypic or fitness values. The no sign epistasis motif is highlighted in blue and the reciprocal sign epistasis motif in red. Note the symmetry of landscapes in panels D and F.

factor binding sites, resulting in selection for low binding affinity[50]. Additional documented cases in which low-affinity binding sites play important regulatory roles include negative auto-regulation by high-copy number transcription factors in *Escherichia coli*[51], where high-affinity binding sites cause suboptimal noise suppression, and developmental patterning in *Ciona intestinalis* embryos[52,53], where high-affinity binding sites cause deleterious ectopic gene expression patterns. Moreover, low-affinity binding sites are commonly observed in the regulatory portfolios of a diversity of organisms, including bacteria[51], yeast[54], fly[49,55,56], sea stars and sea urchins[57], as well as humans[58].

How incongruent are the topographies of genotype-phenotype and fitness landscapes when selection favors a low or intermediate phenotypic value? How does this depend on the ruggedness of the genotype-phenotype landscape? These are important questions, because the topography of a fitness landscape has implications for several evolutionary phenomena,

including the evolution of genetic diversity[59], reproductive isolation[60], and sex[61], as well as the predictability of the evolutionary process itself[62]. How much we can learn about these phenomena from knowledge of a genotype-phenotype landscape depends on the genotype-phenotype landscape's congruence with the fitness landscape. Despite decades of research on fitness landscapes, these questions have not been addressed even in the context of classical theoretical models, such as Mt. Fuji[63], House-of-Cards[64], or NK landscapes[65]. They have also not been addressed in the context of biophysical models of genotype-phenotype landscapes or empirical genotype-phenotype landscapes. Here, we fill this knowledge gap by defining local and global measures of incongruence, which describe the topographical differences between a genotype-phenotype landscape and the corresponding fitness landscape when selection favors a low or intermediate phenotypic value. We use these measures to study incongruence in the context of the aforementioned theoretical models[63–65] and derive some fundamental results that are applicable to all empirical genotype-phenotype landscapes. We then consider the specific case of genotype-phenotype landscapes of transcription factor-DNA interactions, by first looking at an idealised biophysical model[66] and then taking a step further, by analysing 1,137 empirical genotype-phenotype landscapes, wherein genotypes are transcription factor binding sites and the phenotype is a measure of relative binding affinity[21]. We study transcription factor-DNA interactions because there is strong biological motivation for studying selection for low or intermediate binding affinity, as discussed above, and because a large number of empirical genotype-phenotype landscapes of transcription factor-DNA interactions are publicly available[21], thus facilitating the statistical analysis of their incongruence.

List of symbols

w, w_{opt}	Phenotypic value, Optimal phenotypic value
σ	Strength of selection
$F(w)$	Fitness value
ε_{gp}, ε_{f}	Epistasis in the genotype-phenotype landscape (gp) and fitness landscape (f) respectively
p_{gp}, p_{f}	Number of peaks in the genotype-phenotype landscape (gp) and fitness landscape (f) respectively
L	Length of the sequence
K	Ruggedness parameter of the NK model
a	Number of alleles at each locus
m	Number of mismatches in the mismatch model
$\langle l \rangle$	Average length of adaptive walk
$\langle f \rangle$	Mean fitness at equilibrium

Results

We first present our measures of incongruence and the phenotype-to-fitness map used to incorporate the effect of selection for a low or intermediate phenotypic value. We use these measures and this map to study the incongruence of randomly generated

genotype-phenotype landscapes, specifically two-locus biallelic landscapes and multi-locus biallelic landscapes. For the latter, we use NK landscapes[65], which include the corner cases of Mt. Fuji[63] ($K = 0$) and House-Of-Cards[64] ($K = N - 1$) landscapes. We then apply the principles learned from these model genotype-phenotype landscapes to genotype-phenotype landscapes of transcription factor-DNA interactions, first in the context of the mismatch model[66] (which enables us to study landscapes with more than two alleles per locus), and then in the context of empirical measurements of transcription factor-DNA interactions[21]. Finally, we study the evolutionary consequences of landscape incongruence.

Landscape incongruence

Our goal is to quantify the topographical differences between a fitness landscape and its underlying genotype-phenotype landscape, when selection favors a low or intermediate phenotypic value. To do so, we define measures of landscape incongruence, at both a local and a global scale. At a local scale, we quantify differences in pairwise epistasis amongst loci in the genotype-phenotype landscape relative to the same loci in the fitness landscape (Methods). We do so by classifying the type of magnitude epistasis (ε) as additive (i.e., no epistasis; $\varepsilon = 0$), positive ($\varepsilon > 0$), or negative ($\varepsilon < 0$). For a pair of loci, we then compare the type of epistasis in the genotype-phenotype landscape ($\varepsilon_{\mathrm{gp}}$) to the type of epistasis in the fitness landscape (ε_{f}), and report whether this is the same in the two landscapes. We do the same for an additional classification of epistasis based on the absence or presence of sign epistasis. This results in three categories – no sign epistasis, simple sign epistasis, or reciprocal sign epistasis[67]. At a global scale, we quantify incongruence as the difference in the number of peaks in the fitness landscape (p_{f}) relative to the genotype-phenotype landscape (p_{gp}). Taken together, these local and global measures allow us to determine the extent to which selection for a low or intermediate phenotypic value increases or decreases the ruggedness of the fitness landscape, relative to the genotype-phenotype landscape.

Phenotype-to-fitness map

To study the effect of selection for low or intermediate phenotypic values, we use a Gaussian phenotype-to-fitness map $F(w)$ centred around an optimal phenotypic value w_{opt},

$$F(w) = \exp\left[-\left(\frac{w - w_{\mathrm{opt}}}{\sigma}\right)^2\right]. \tag{1}$$

The parameter σ determines the strength of selection by controlling how rapidly fitness decreases as the phenotype w deviates from the optimal phenotype w_{opt}. Increasing σ decreases the strength of selection, because it broadens the fitness map around w_{opt} and thus decreases the fitness differences between similar phenotypes. Similar maps are commonly used in evolutionary modeling frameworks, such as Fisher's Geometric model[68,69] and models of speciation[70], as well as in biophysical models of intermolecular interactions[29], including transcription factor-DNA interactions[46]. However, we emphasize that most of our measures of incongruence depend only on the rank ordering of phenotypic or fitness values, and are therefore independent of the exact shape of the phenotype-fitness map, so long as it is symmetric.

Fig. 1B-F illustrates the application of the phenotype-to-fitness map to a simple three-locus, biallelic genotype-phenotype landscape. This landscape is purely additive, and as a consequence, it exhibits no epistasis and has only one peak (Fig. 1B). Applying the phenotype-to-fitness map (Fig. 1C) to this landscape can change the amount and type of epistasis, as well as the location and number of peaks in the resulting fitness landscape, relative to the genotype-phenotype landscape (Fig. 1D-F). It can therefore cause incongruence between genotype-phenotype and fitness landscapes. Whereas this schematic and our analyses below pertain to a single phenotype, extending our model to multiple phenotypes is straightforward, as we later discuss.

Two-locus biallelic genotype-phenotype landscapes

We first study incongruence using the simplest form of genotype-phenotype landscape that is capable of exhibiting epistasis: a two-locus biallelic landscape. We represent genotypes as binary strings of length $L = 2$ and randomly assign a phenotype w_i to each genotype i, which we draw from a uniform distribution between 0 and 1. We then apply the Gaussian phenotype-fitness map (Eq. 1) to generate the corresponding fitness landscape. We repeat this process 10,000 times for values of $w_{\text{opt}} \in [0,1]$ (in increments of 0.01), and report the probability that the type of epistasis in the genotype-phenotype landscape is the same as in the fitness landscape. We first differentiate between no magnitude epistasis, positive epistasis, and negative epistasis, and then between no sign epistasis, simple sign epistasis, and reciprocal sign epistasis.

Incongruence in magnitude epistasis is highest when selection favors low phenotypic values

To determine whether the type of magnitude epistasis is the same in the genotype-phenotype landscape and fitness landscape, we calculate the product $\varepsilon_{\text{gp}} \cdot \varepsilon_{\text{f}}$, which will be positive when the two landscapes have the same type of epistasis. We use this product, rather than a discrete categorization, to ensure analytical tractability. We obtain an analytical expression for $\varepsilon_{\text{gp}} \cdot \varepsilon_{\text{f}}$ by assuming σ to be large (Supplementary Material, Derivation 1):

$$\varepsilon_{\text{f}} \cdot \varepsilon_{\text{gp}} = \frac{1}{\sigma^2} \left[2\varepsilon_m \cdot \varepsilon_{\text{gp}} + (2w_{\text{opt}} - \Sigma_i w_i) \cdot \varepsilon_{\text{gp}}^2 \right], \tag{2}$$

where w_i is the phenotypic value of genotype i with $i \in \{0,1\}^2$ and $\varepsilon_m = w_{00}w_{11} - w_{01}w_{10}$, which is also known as multiplicative epistasis[71].

The probability that the type of epistasis is the same in the genotype-phenotype and fitness landscapes, $P(\varepsilon_{\text{gp}} \cdot \varepsilon_{\text{f}} > 0)$ can be computed using Monte-Carlo methods and yields results in good agreement with the randomly generated genotype-phenotype landscapes when σ is large. This is shown in Fig. 2, where three trends are immediately apparent. First, for large σ, the probability that the type of epistasis is the same in the genotype-phenotype landscape and the fitness landscape increases as the optimal phenotype w_{opt} increases, in agreement with the intuition that selection for large phenotypic values leaves the genotype-phenotype landscape mostly unchanged, except for the nonlinear rescaling introduced by the phenotype-fitness map. Second, even when $w_{\text{opt}} = 1$, the probability that the type of epistasis is the same in the genotype-phenotype landscape and the fitness landscape is less than one. The reason is the nonlinear rescaling introduced by the phenotype-to-fitness map does not

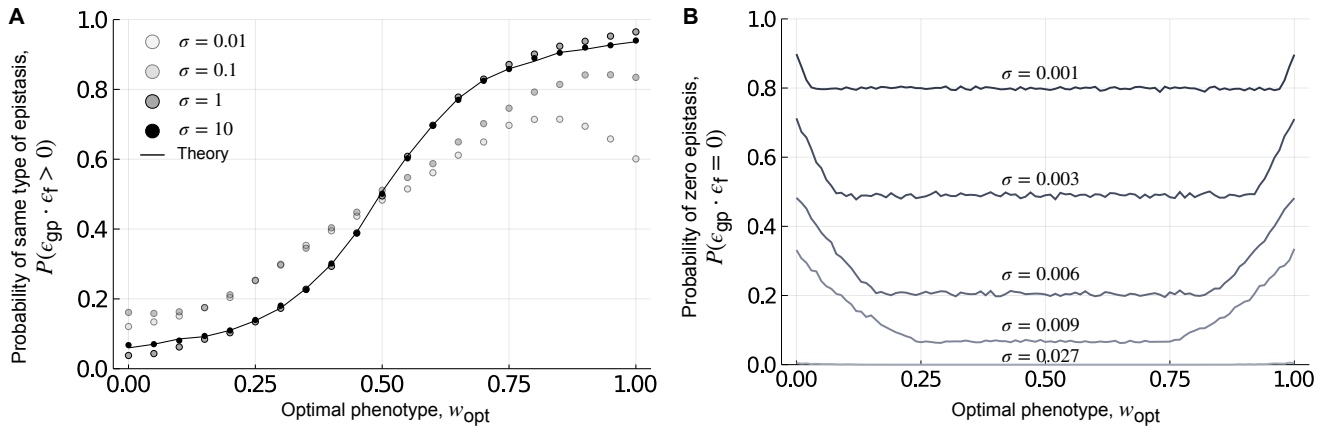


Figure 2. Local incongruence: magnitude epistasis. (A) The probability of retaining the type of epistasis, shown in relation to the optimal phenotype w_{opt}. The black line shows the theoretical prediction and the dots show the results from randomly generated genotype-phenotype landscapes for different values of σ. The theoretical approximation agrees well with the results from randomly generated genotype-phenotype landscapes for large σ (i.e., $\sigma \geq 1$). (B) The probability of observing zero magnitude epistasis in the fitness landscape, shown in relation to w_{opt}, for different values of σ, which we selected to show the range of variation in the probability of observing zero epistasis.

guarantee conservation of the type of magnitude epistasis, even though it does preserve the rank ordering of fitness values. Results obtained with the randomly generated genotype-phenotype landscapes show this effect becomes even more pronounced as σ decreases. This means that as the strength of selection for w_{opt} increases, so does the likelihood of landscape incongruence. Finally, as σ decreases, the probability of retaining the type of epistasis from the genotype-phenotype landscape in the fitness landscape is not maximized at $w_{\text{opt}} = 1$, but rather at a smaller w_{opt} (e.g., at $w_{\text{opt}} \approx 0.8$ when $\sigma = 0.01$, Fig. 2A). The reason is as σ decreases, the fitness function becomes extremely narrow and more phenotypic values are mapped to zero fitness (within computer precision), resulting in cases where the genotype-phenotype landscape exhibits epistasis (i.e., $\varepsilon_{\text{gp}} \neq 0$), but the fitness landscape does not (i.e., $\varepsilon_{\text{f}} = 0$), because all fitness values are zero. Fig. 2B shows this is more likely to occur when σ is small and as w_{opt} approaches its extreme values of 0 or 1. In these randomly generated landscapes, the probability of obtaining negative epistasis is the same as the probability of obtaining positive epistasis, and the probability of conserving the type of epistasis is independent of the type of epistasis in the genotype-phenotype landscape. In sum, these results show that selection for low or intermediate phenotypic values can modify the genotype-phenotype landscape, such that the resulting fitness landscape exhibits a different type of magnitude epistasis, and this effect is most pronounced when selection is strong and the optimal phenotypic value is low.

Incongruence in sign epistasis is highest when selection favors intermediate phenotypic values

Next, we categorized landscapes as exhibiting no sign epistasis, simple sign epistasis, or reciprocal sign epistasis. These three motifs are shown in the center panel of Fig. 3, where arrows point from genotypes with a lower phenotypic or fitness value to genotypes with a higher phenotypic or fitness value. Because the presence of sign epistasis only depends upon the partial ordering of the phenotypic values, we expect to retain the motif from the genotype-phenotype landscape in the fitness landscape

as $w_{\text{opt}} \to 1$. We also expect to retain the motif as $w_{\text{opt}} \to 0$, because selecting for $w_{\text{opt}} = 0$ simply flips all the arrows from the genotype-phenotype landscape in the fitness landscape, which does not change the categorization of the motif. Thus, we expect the probability of retaining the motif from the genotype-phenotype landscape in the fitness landscape to be "U" shaped and symmetric about $w_{\text{opt}} = 0.5$. Fig. 3 shows the probability of retaining or changing the motif from the genotype-phenotype landscape in the fitness landscape for 10,000 randomly generated two-locus biallelic landscapes, grouped according to the motif in the genotype-phenotype landscape. These results confirm the expected "U" shape of the probability of retaining the type of epistasis from the genotype-phenotype landscape in the fitness landscape as w_{opt} is varied from 0 to 1.

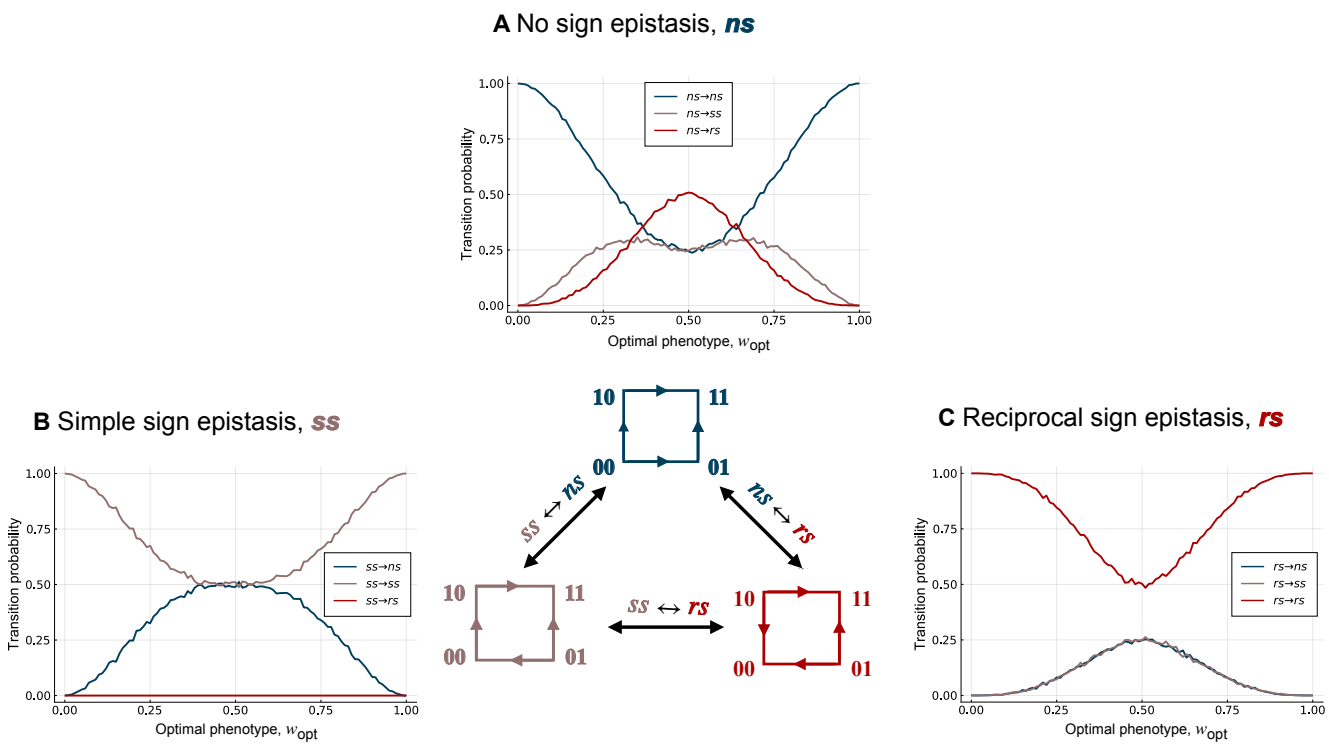


Figure 3. Local incongruence: sign epistasis. The probability of retaining or changing the type of epistasis in the genotype-phenotype landscape, relative to the fitness landscape, shown in relation to the optimal phenotypic value w_{opt}. Data are grouped based on whether the genotype-phenotype landscapes exhibits (A) no sign epistasis (blue), (B) simple sign epistasis (brown), or (C) reciprocal sign epistasis (red). The colours of the lines represent the type of epistasis in the resulting fitness landscape. These results are independent of σ, because they only depend on the rank ordering of fitness values. Notice the "U" shape of the probability of retaining the type of epistasis in each panel.

Simple sign epistasis cannot be modified into reciprocal sign epistasis

When the genotype-phenotype landscape has the no sign epistasis motif, selection for an intermediate phenotypic value can transform the landscape into the simple sign epistasis motif or the reciprocal sign epistasis motif (Fig. 3A), in line with recent results on Fisher's Geometric model[72]. When the genotype-phenotype landscape has the simple sign epistasis motif, selection for an intermediate phenotypic value can transform the landscape into a no sign epistasis motif, but not into a reciprocal sign epistasis motif (Fig. 3B; Supplementary Material, Proof 1). Finally, when the genotype-phenotype landscape has the reciprocal sign epistasis motif, selection for an intermediate phenotypic value transforms the landscape into the no sign epistasis motif or

the simple sign epistasis motif with equal probability (Fig. 3C; Supplementary Material, Proof 2). Moreover, the probability of retaining the motif from the genotype-phenotype landscape in the fitness landscape is always higher than the probability of changing it when the genotype-phenotype landscape has the simple sign epistasis motif (Fig. 3B) or the reciprocal sign epistasis motif (Fig. 3C). This is not always true when the genotype-phenotype landscape has the no sign epistasis motif (Fig. 3A), because at intermediate w_{opt} the landscape is most likely to transform into the reciprocal sign epistasis motif. Taken together, these results show that selection for intermediate phenotypic values can modify genotype-phenotype landscapes with no sign epistasis into fitness landscapes with sign epistasis and vice versa.

The inferences about pairwise interactions can be carried forward to multi-locus biallelic landscapes because their genotype spaces, which are L-dimensional hypercubes, are composed of two-dimensional squares. Due to the adjacency of squares, in the three-locus case, the motifs of four out of the six squares are sufficient to determine the motifs of the rest of the squares, and for any L, the motifs of only $2^{L-2} \cdot (L-1)$ squares are necessary to determine the motifs of all of the remaining squares. This is only a fraction $2/L$ of all the squares in the hypercube (because the total number of faces in an L-dimensional hypercube is $2^{L-2} \cdot \binom{L}{2}$), which is clearly minuscule for large L. However, pairwise interactions are not sufficient to predict peak patterns, which may result from higher-order interactions[73]. We study these in the next section.

Multi-locus biallelic genotype-phenotype landscapes

We use the NK model[65] to study multi-locus biallelic genotype-phenotype landscapes (Methods). In this model, each locus in a genotype of length L epistatically interacts with K other loci (whereas N is typically used to denote the number of loci in this model, we use L for consistency with the rest of our text). As corner cases, this model includes Mt. Fuji landscapes[63] when $K = 0$ and House-of-Cards landscapes[64] when $K = L-1$. For each combination of L and K, we use this model to randomly generate a genotype-phenotype landscape. We then apply the Gaussian phenotype-fitness map (Eq. 1) to generate a fitness landscape. We repeat this process 10,000 times for $w_{\text{opt}} \in [0,1]$ (in increments of 0.01), and report the average of the absolute change in the number of peaks, i.e., $\langle |p_{\text{f}} - p_{\text{gp}}| \rangle$, where p_{f} is the number of peaks in the fitness landscape and p_{gp} is the number of peaks in the genotype-phenotype landscape. This is our measure of global incongruence.

Mt. Fuji landscapes

We begin with Mt. Fuji genotype-phenotype landscapes. Because these are single-peaked, selection for a low or intermediate phenotypic value can only maintain or increase the number of peaks from the genotype-phenotype landscape in the fitness landscape. Fig. 4A shows this change in the number of peaks, and Fig. 4B shows the probability that the number of peaks changes, in relation to w_{opt} for landscapes with $L = 2$ to $L = 8$ loci. These trends are symmetric about $w_{\text{opt}} = 0.5$, because Mt. Fuji landscapes are additive, so selecting for $w_{\text{opt}} = 0$ is equivalent to selecting for $w_{\text{opt}} = 1$ with regard to the change in the number of peaks. The reason is that selecting for $w_{\text{opt}} = 0$ flips all of the arrows in the fitness landscape, relative to the genotype-phenotype landscape, which changes the location of the peak, but does not change the number of peaks. This is illustrated in Fig. 1B,D. In Fig. 4A, the plot of $\langle |p_{\text{f}} - p_{\text{gp}}| \rangle$ with respect to w_{opt} exhibits $L-1$ maxima, each separated

from one another by a distance $1/L$ in w_{opt} (Supplementary Material, Note 1). For $L \geq 4$, the maxima are difficult to visually discern. More obvious is the increase in the number of peaks, and the probability that the number of peaks increases, as L increases, the latter converging to one for all values of w_{opt} except the extreme cases of $w_{\text{opt}} = 0$ and $w_{\text{opt}} = 1$. Note, however, that a high probability of increase in the number of peaks does not necessarily correspond to a high increase in the number of peaks, as can be seen from the different positions of the maxima in Fig. 4A,B (Supplementary Material, Note 1). For large L (> 10), the expected number of peaks in the fitness landscape increases exponentially with L[72]. In sum, these results show that selection for intermediate phenotypic values readily transforms Mt. Fuji genotype-phenotype landscapes, which are smooth and single-peaked, into rugged fitness landscapes, and that this effect is most pronounced for large L and intermediate w_{opt}.

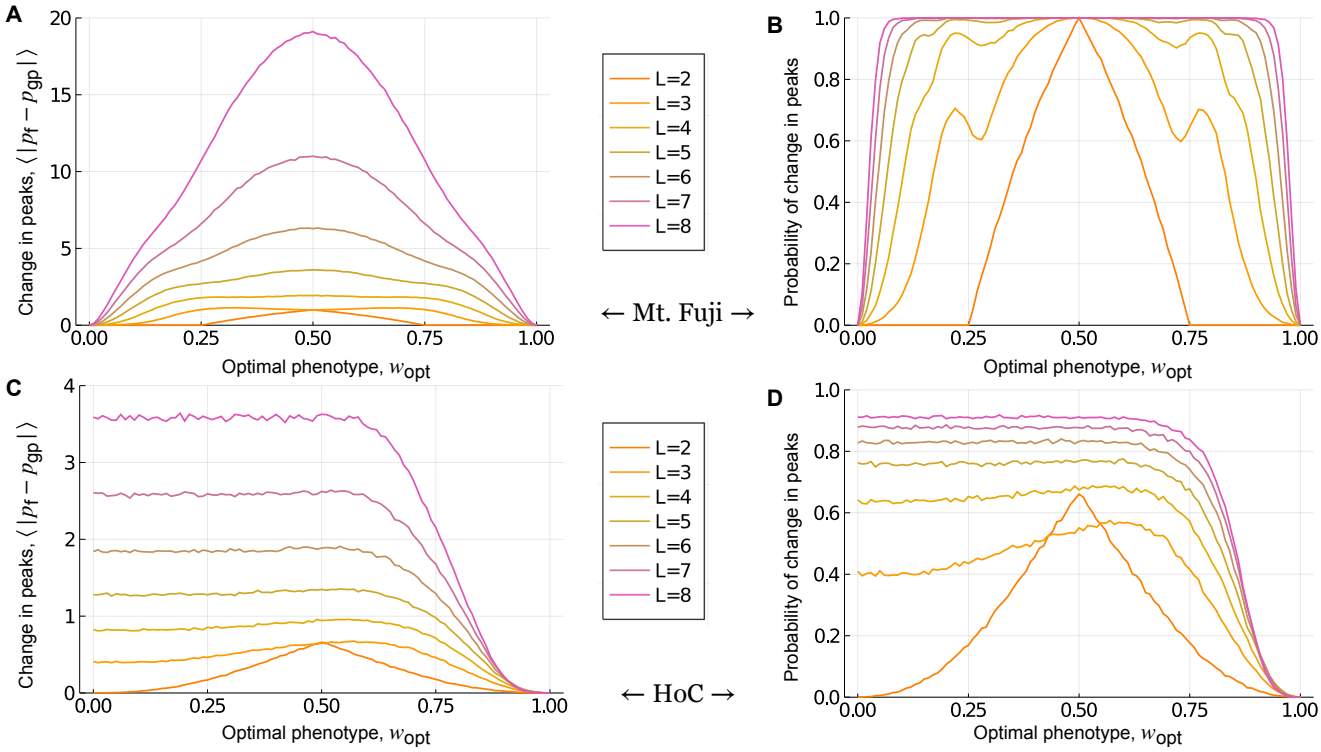


Figure 4. Global incongruence : Mt. Fuji and House-of-Cards genotype-phenotype landscapes. The absolute change in the number of peaks and the probability that the number of peaks changes in the fitness landscape, relative to (A,B) Mt. Fuji and (C,D) House-of-Cards genotype-phenotype landscapes, shown in relation to w_{opt} for $L \in \{2, 3..8\}$. These results are independent of σ, because they only depend on the rank ordering of fitness values.

House-of-Cards landscapes

We next study House-of-Cards genotype-phenotype landscapes. These landscapes are highly rugged, with an average of $\frac{2^L}{L+1}$ peaks[74], whereas the maximum possible number of peaks is 2^{L-1}. As such, selection for a low or intermediate phenotypic value can either increase or decrease the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape. Fig. 4C shows this change in the number of peaks, and Fig. 4D shows the probability that the number of peaks changes, in relation to w_{opt} for landscapes with $L = 2$ to $L = 8$ loci. The change in the number of peaks is symmetric about $w_{\text{opt}} = 0.5$ for the two-locus case, where the number of peaks does not change as $w_{\text{opt}} \to 0$ or $w_{\text{opt}} \to 1$. However, this symmetry is lost for $L > 2$.

The reason is that although the phenotype-fitness map flips all of the arrows in the fitness landscape when $w_{\text{opt}} = 0$, relative to the genotype-phenotype landscape, this does not guarantee conservation of the number of peaks. The number of peaks is jointly determined by the adjacent faces of the hypercube and thus, only very specific changes in the directions of arrows in the genotype-phenotype landscape guarantees conservation of the number of peaks in the fitness landscape (Supplementary Fig. 4). However, in contrast to Mt. Fuji genotype-phenotype landscapes, the magnitude of change in the number of peaks increases very little with L, despite an exponential increase in the maximum number of possible peaks. Moreover, the probability that the number of peaks changes is still less than one for large L. Finally, for large L, both the change in the number of peaks and the probability that the number of peaks changes are independent of w_{opt}, so long as w_{opt} is sufficiently less than one. This observation depends on the probability distribution used to generate these landscapes. Because the 2^L phenotypes in the NK model are drawn from a uniform distribution, nearly the same number of these phenotypes will be close to the optimal phenotype w_{opt}, so long as L is sufficiently large. Thus, averaging over all possible configurations of the genotype-phenotype landscape yields the same value of $\langle |p_{\text{f}} - p_{\text{gp}}| \rangle$ for every w_{opt}. For sufficiently large L, this value is given by (Supplementary Material, Derivation 2):

$$\lim_{L\to\infty} \langle |p_{\text{f}} - p_{\text{gp}}| \rangle = \sqrt{\frac{2^L \cdot (L-1)}{2\pi(L+1)^2}}. \tag{3}$$

So far we have focused on the absolute change $\langle |p_{\text{f}} - p_{\text{gp}}| \rangle$ in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape. For House-of-Cards genotype-phenotype landscapes, selection for a low or intermediate phenotypic value can either increase or decrease the number of peaks. We were therefore interested in finding out which outcome is more likely. While one might expect a decrease to be more likely, due to the extreme ruggedness of House-of-Cards genotype-phenotype landscapes, we find that the number of peaks is equally likely to increase or decrease (Supplementary Material, Proof 3). In sum, these results show that in House-of-Cards genotype-phenotype landscapes, selection for a low or intermediate phenotypic value increases or decreases the number of peaks in the fitness landscape with equal probability, and the severity as well as the probability of this change increases with L and is largely independent of w_{opt}.

Global incongruence decreases as the ruggedness of the genotype-phenotype landscape increases

Finally, we study NK genotype-phenotype landscapes, which bridge the gap between Mt. Fuji and House-of-Cards landscapes in terms of ruggedness, as K increases from 0 to $L-1$. Fig. 5 shows the absolute change in the number of peaks in the fitness landscape relative to the genotype-phenotype landscape for genotypes of length $L = 5$ and $L = 8$, as K is increased from 0 to $L-1$. Note the gradual transition from the trends observed for Mt. Fuji genotype-phenotype landscapes (Fig. 4A) to those observed for House-of-Cards genotype-phenotype landscapes (Fig. 4C) as K increases. From these trends, we conclude four principles of how the ruggedness of an NK genotype-phenotype landscape influences its incongruence with the fitness landscape. As an NK genotype-phenotype landscape becomes more rugged, incongruence (1) loses symmetry about $w_{\text{opt}} = 0.5$, (2) becomes less sensitive to w_{opt}, and (3) decreases in severity, at least in terms of the absolute change in the number of peaks.

Finally, the probability of increasing the number of peaks is always greater than or equal to the probability of decreasing the number of peaks, with the equality holding for House-of-Cards genotype-phenotype landscapes. This last principle is both intuitive and informative — it tells us that on average, selection for low or intermediate values is more likely to increase the ruggedness of a fitness landscape, relative to the genotype-phenotype landscape. Thus, selection for low or intermediate values is more likely to break than to create phenotypic correlations between mutationally similar genotypes as they map onto fitness, rendering fitness landscapes more rugged than their underlying genotype-phenotype landscapes. In the subsequent sections we address whether and how these results apply to genotype-phenotype landscapes with more than two alleles per locus, specifically in the context of a biophysical model and experimental measurements of transcription factor-DNA interactions.

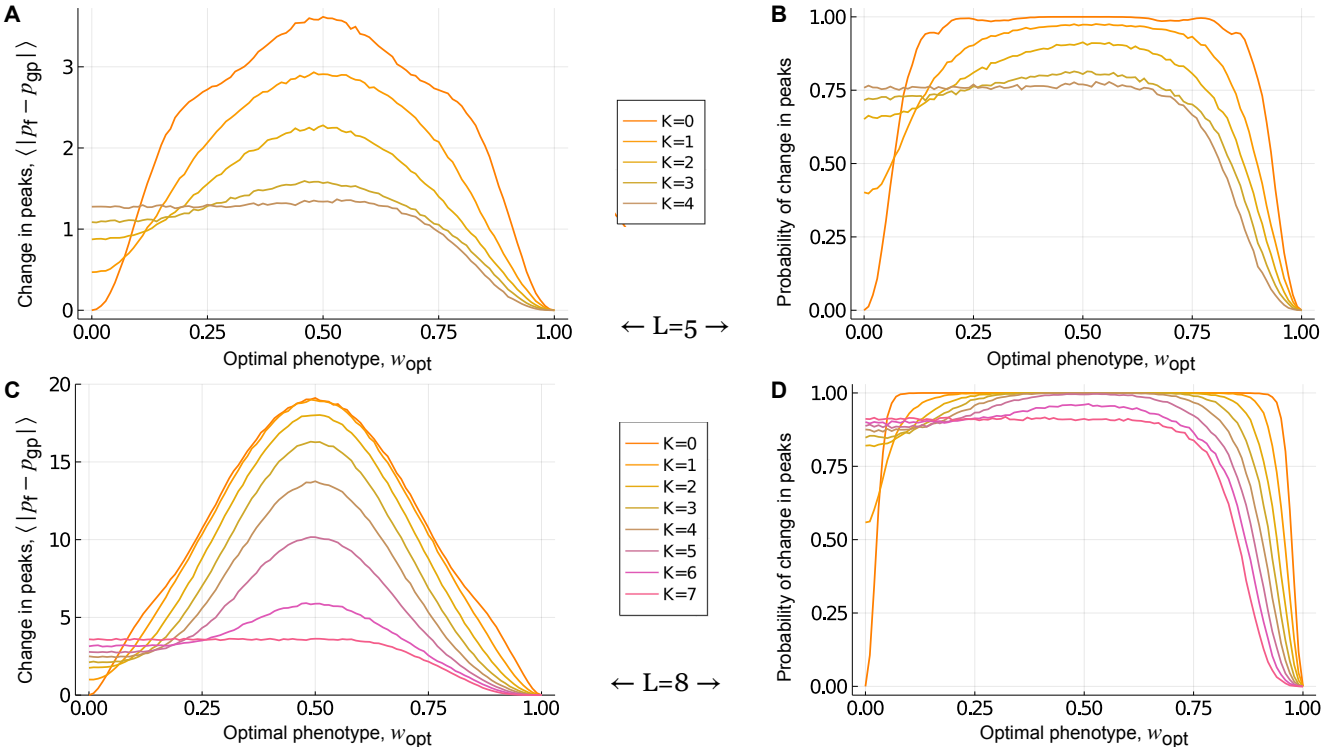


Figure 5. Global incongruence : NK genotype-phenotype landscapes. The absolute change in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape, is shown in relation to w_{opt} for genotypes of length (A) $L = 5$ and (C) $L = 8$, as K increases from zero to $L - 1$. The corresponding probability of change in the number of peaks is shown in relation to w_{opt} for genotypes of length (B) $L = 5$ and (D) $L = 8$, as K increases from zero to $L - 1$. These results are independent of σ, because they only depend on the rank ordering of fitness values.

Genotype-phenotype landscapes of transcription factor-DNA interactions

Motivated by the common usage of low- and intermediate-affinity transcription factor binding sites in the regulatory portfolios of a diversity of organisms[49,51,54–57], we now study the incongruence of genotype-phenotype landscapes of transcription factor-DNA interactions and the corresponding fitness landscapes generated after selection for low or intermediate phenotypic values. In these landscapes, genotypes represent DNA sequences — transcription factor binding sites — and the phenotype of a DNA sequence is the affinity with which it binds a transcription factor[21]. Because the regulatory effects of transcription

factor-DNA interactions are partly determined by binding affinity[41,42] and mutations to transcription factor binding sites can alter binding affinity[48,75], the topographies of genotype-phenotype landscapes of transcription factor-DNA interactions have important implications for the evolution of gene regulation[21]. We study these landscapes using both a biophysical model and experimental measurements of transcription factor-DNA interactions. We focus on transcription factor binding sites of length $L = 8$, because this is the length of the binding sites assayed by protein binding microarrays[75,76] — the data used to construct the empirical genotype-phenotype landscapes of transcription factor-DNA interactions[21].

The mismatch model

We first study genotype-phenotype landscapes of transcription factor-DNA interactions generated using the so-called mismatch model[46,47,66]. The key assumption of this model is that the binding energy of a DNA sequence is a linear function of the number of mismatches between the sequence (genotype) and a transcription factor's consensus sequence — the sequence it binds with the highest affinity. Further, each mismatch is assumed to have the same energetic cost and these costs combine additively to determine binding energy. This model results in a Mt. Fuji-like, permutation-invariant genotype-phenotype landscape, wherein the phenotype only depends on the number of differences between the genotype and the consensus sequence, but not on which loci in the genotype differ from the consensus sequence. Although this is a simplified model, it provides an opportunity to study the effects of having more than two alleles per locus and serves as a bridge to our analyses of empirical transcription factor-DNA interactions.

To ensure that these results are comparable to the results on the empirical landscapes, we consider the negative of the binding energy as our phenotype, such that sequences that bind more strongly are assigned higher phenotypic values. We assume a phenotypic value of 1 for the genotype that is identical to the consensus sequence. For each mismatch between a genotype and the consensus sequence, we deduct a small positive value e, such that the phenotypic value of a genotype with m mismatches is $A_m = 1 - m \cdot e$. Due to the permutation invariance, the genotype-phenotype landscape is highly degenerate, such that the number of genotypes N_m in a mismatch class m is distributed according to the asymmetric binomial distribution:

$$N_m = (a-1)^m \frac{L!}{(L-m)!m!}, \tag{4}$$

where a is the number of alleles per locus ($a = 4$ for transcription factor binding sites, because they are DNA sequences). Fig. 6A shows this distribution for transcription factor binding sites of length $L = 8$. Note that N_m is maximized at $m = 6$.

Fig. 6B shows the incongruence between genotype-phenotype landscapes and fitness landscapes of transcription factor-DNA interactions constructed using the mismatch model, under selection for an optimal mismatch class m_{opt}, reported in terms of the absolute change in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape. Based on our analysis of Mt. Fuji genotype-phenotype landscapes, we anticipated asymmetric incongruence about the mismatch class $m_{\text{opt}} = 6$, because the distribution of the number of genotypes per mismatch class is asymmetric with a maximum at mismatch class $m = 6$ (Fig. 6A) and we expected all genotypes in the optimal mismatch class m_{opt} to be peaks. However,

we observe symmetric incongruence about $m_{\text{opt}} = 4$ (Fig. 6B). This occurs because $a > 2$, which renders some genotypes in the same mismatch class mutational neighbours. Consequently, the peaks can be broad and include more than one genotype, thus resembling plateaus. Specifically, each genotype has $(a-1)^m - 1$ mutational neighbours that are in the same mismatch class and are therefore part of the same peak. This leaves $\binom{L}{m}$ clusters of genotypes to be peaks in each mismatch class m, an expression that is maximized with $m = 4$ when $L = 8$, thus forcing symmetry in the absolute change in the number of peaks about $m_{\text{opt}} = 4$. However, the width of the peaks increases as $(a-1)^m$, leading to a tripling of peak width for each increment in m (Fig. 6B). Thus, even in this idealised genotype-phenotype landscape, features of empirical landscapes of TF-DNA interactions begin to emerge, such as broad peaks.

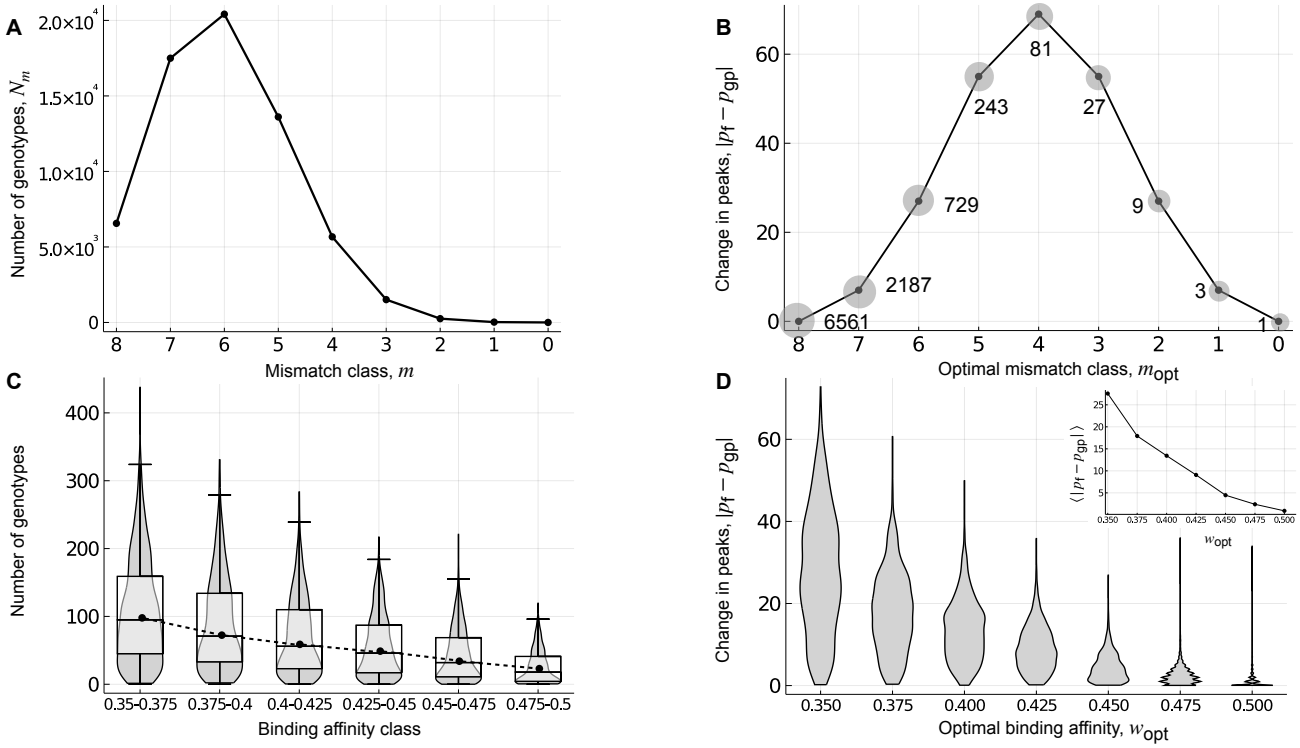


Figure 6. Global incongruence : genotype-phenotype landscapes of transcription factor-DNA interactions. Landscapes constructed using (A,B) the mismatch model and (C,D) experimental measurements from protein binding microarrays for 1,137 eukaryotic transcription factors. (A) The number of genotypes, shown in relation to mismatch class. (B) The absolute change in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape, shown in relation to the optimal mismatch class m_{opt}. Labels indicate the number of genotypes per peak in the fitness landscape. Note the symmetry in the absolute change in the number of peaks around mismatch class $m_{\text{opt}} = 4$, as well as the tripling of the number of genotypes per peak for each increment in m_{opt}. The grey shaded circles are a schematic representation of the growing width of the peaks. (C) The number of genotypes per binding affinity class, where protein binding microarray E-scores are used as a proxy for relative binding affinity. Violin plots show the distribution, and box-and-whisker plots the 25-75% quartiles, across genotype-phenotype landscapes for the 1,137 transcription factors. Closed symbols and the dashed line denote the median of each distribution. (D) Violin plots of the distribution of the absolute change in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape, shown in relation to the optimal binding affinity w_{opt} for $\sigma = 0.15$. The inset shows the mean absolute change in the number of peaks, in relation to w_{opt}. The x-axes in (A,B) are arranged such that binding affinity increases when read from left to right, in qualitative agreement with the x-axes in (C,D). The results in panels A-C are independent of σ.

Empirical landscapes

We now study genotype-phenotype landscapes of transcription factor-DNA interactions generated using experimental data from protein binding microarrays[76] (Methods). For all possible DNA sequences of length $L = 8$, these data include an enrichment score (E-score) ranging from -0.5 to 0.5 that serves as a proxy for relative binding affinity, such that higher E-scores correspond to higher binding affinities[75,76]. We have previously used these data to construct genotype-phenotype landscapes for 1,137 eukaryotic transcription factors, in which the surface of each landscape was defined by the E-score[21]. Due to limitations in the reproducibility of E-scores across microarray designs for genotypes that are bound non-specifically or with very low affinity[75,76], each genotype-phenotype landscape only includes DNA sequences with an E-score exceeding a threshold of 0.35, which corresponds to a false discovery rate of 0.001[75]. As shown in our previous work[21], these landscapes tend to exhibit little, if any, reciprocal sign epistasis and therefore comprise few peaks. As such, they bear resemblance to the genotype-phenotype landscapes constructed using the mismatch model. An important difference, however, is that genotypes in the lower half of the E-score range are omitted from each landscape due to the reproducibility issues mentioned above. Fig. 6C shows the distributions of the number of genotypes across all 1,137 genotype-phenotype landscapes, grouped into six binding affinity classes. Whereas the lowest binding affinity class contains the most genotypes, we cannot determine if this is the true maximum, because we do not know what these distributions look like for lower binding affinity classes. However, assuming the energetic contribution of each binding site to be additive[66], we expect lower binding affinity classes to have fewer genotypes and the maximum to occur at an intermediate binding affinity class, as can be seen in the mismatch model and other models in literature[43].

Fig. 6D shows the incongruence between these 1,137 empirical genotype-phenotype landscapes and their corresponding fitness landscapes, under selection for an optimal binding affinity w_{opt}, reported in terms of the absolute change in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape for $\sigma = 0.15$. Interestingly, the effect of changing σ is small (Supplementary Fig. 5B) because increasing σ not only decreases the range of variation of fitness values, but also decreases the uncertainty in these values (Methods), leaving the number of peaks largely unchanged. The mean and variance in the absolute change in the number of peaks decreases as w_{opt} increases to 0.5, in line with the intuition that selection for high w_{opt} generates fitness landscapes that are topographically similar to the underlying genotype-phenotype landscape. More precisely, the percentage of landscapes that show a change in the number of peaks decreases from around 99% to 40%, as we go from $w_{\text{opt}} = 0.35$ to $w_{\text{opt}} = 0.5$. As anticipated from our results with additive genotype-phenotype landscapes (Mt. Fuji and the mismatch model), selection for low or intermediate phenotypic values is more likely to increase than to decrease the number of peaks, although the relative fraction of increase depends upon w_{opt} – while around 99% of the landscapes show an increase in peaks for $w_{\text{opt}} = 0.35$, this value decreases to around 40% for $w_{\text{opt}} = 0.5$. Further, when we separately analyzed the single-peaked ($\approx 66.75\%$) and multi-peaked ($\approx 33.25\%$) genotype-phenotype landscapes, we found the single-peaked landscapes to be more incongruent (Supplementary Fig. 5A), in line with our results from the previous section. In sum, these results show that selection for low or intermediate phenotypic values tends to increase the ruggedness of a fitness landscape,

relative to the underlying genotype-phenotype landscapes, rendering genotype-phenotype landscapes a poor proxy for fitness landscapes under such selection.

Mismatch model and empirical landscapes show different kinds of global incongruence

In contrast to genotype-phenotype landscapes constructed using the mismatch model, the incongruence of genotype-phenotype landscapes constructed using protein binding microarray data is highest for the lowest binding affinity class, rather than an intermediate class. There are three non-mutually exclusive explanations for this. First, these empirical landscapes are not purely additive[21], unlike the landscapes constructed with the mismatch model, so we do not expect perfect symmetry about an intermediate w_{opt}. Second, as previously mentioned, protein binding microarray data do not capture the full range of binding affinity, so the lowest binding affinity class in our data (E-score = 0.35), which contains the most genotypes (Fig. 6C), is unlikely to be the lowest binding affinity class, but rather an intermediate binding affinity class. Third, while the binding affinity of a sequence is highly correlated with the binding affinities of its mutational neighbors[77], this correlation is not perfect, so neighboring genotypes that are in the same mismatch class may not have sufficiently similar binding affinities to be considered part of the same peak, unlike in the mismatch model.

There are two additional differences between the incongruence of the empirical genotype-phenotype landscapes and those constructed using the mismatch model that are worth highlighting. First, in the empirical landscapes, the average height of the peaks is maximised when selecting for intermediate binding affinities and is lowest when selecting for the highest affinity (Supplementary Fig. 6A), whereas peak height is independent of m_{opt} in the mismatch model. Second, in the empirical landscapes, peak width is maximized when selecting for the highest binding affinity ($w_{\text{opt}} = 0.5$) (Supplementary Fig. 6B), whereas it is maximized when selecting for the lowest binding affinity ($m_{\text{opt}} = 8$) in the mismatch model (Fig. 6B). These two differences in incongruence are important for understanding evolutionary simulations on these landscapes, which are the focus of the next section.

Evolutionary consequences

A key finding of our analyses so far is that selection for low or intermediate phenotypic values is more likely to increase than to decrease the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape. Since the ruggedness of a fitness landscape has important implications for a diversity of evolutionary phenomena[59–62], we now study the evolutionary consequences of this finding. We do so using two metrics: (1) the length $\langle l \rangle$ of a greedy adaptive walk, averaged over all possible genotypes as starting points, and (2) the mean population fitness at equilibrium $\langle f \rangle$ under deterministic mutation-selection dynamics (i.e., assuming an infinite population size). These metrics provide complementary information to measures of landscape ruggedness, such as the number of peaks. To see why, consider our results with the mismatch model, in which the fitness landscape had one or more global peaks of equal height (Fig. 6A,B). We observed single-peaked fitness landscapes when the optimal mismatch class was $m_{\text{opt}} = 0$ or $m_{\text{opt}} = 8$, and the most rugged fitness landscapes when the optimal mismatch class was $m_{\text{opt}} = 4$. We also observed a tripling of the number of genotypes per peak as m_{opt} increased from 0 to 8, such that

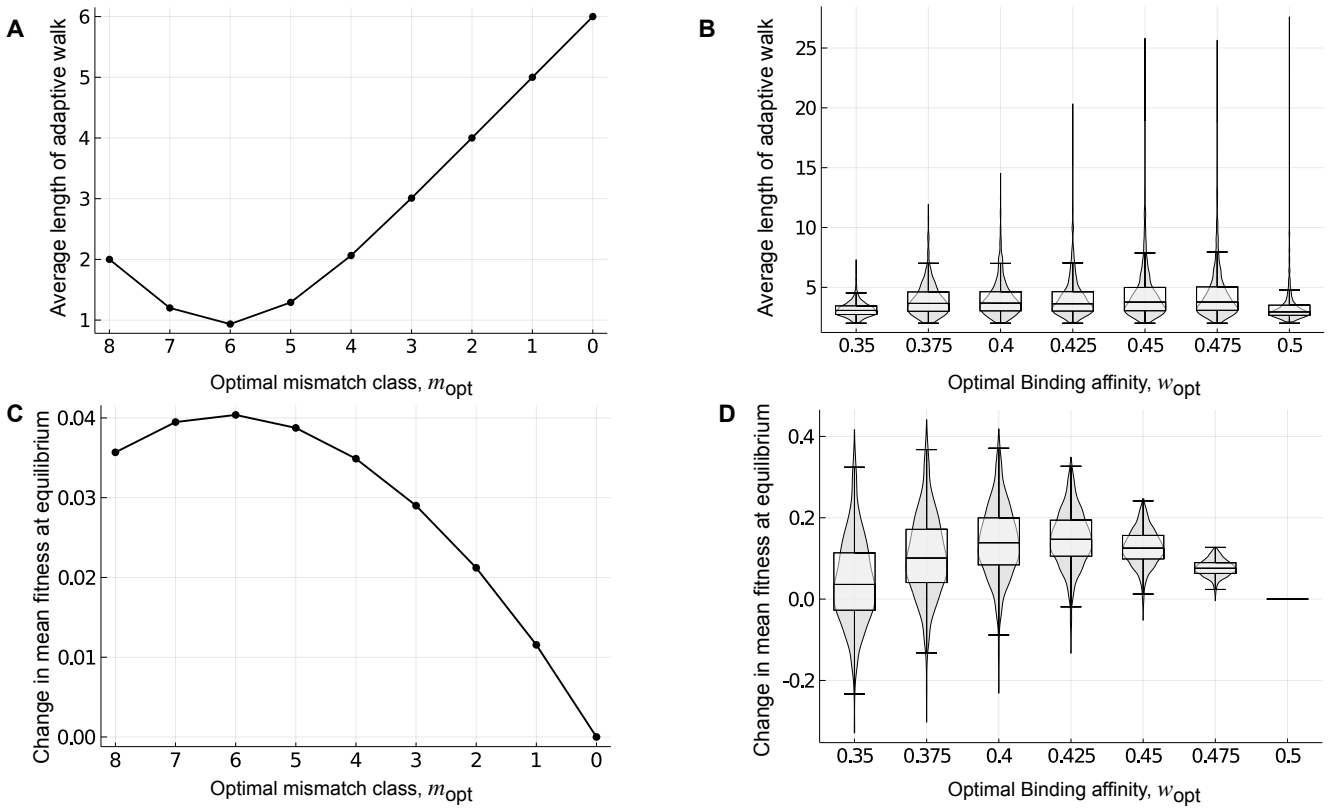


Figure 7. Rugged fitness landscapes need not impede adaptation. The average length of the shortest adaptive walk to a fitness landscape's global peak $\langle l \rangle$ and the change in mean population fitness at equilibrium $\langle f \rangle$ is shown for landscapes of transcription factor-DNA interactions generated using (A,C) the mismatch model for $e = 0.05$ and (B,D) protein binding microarray data. In (B,D), violin plots show the distribution, and box-and-whisker plots the 25-75% quartiles, across the 1,137 empirical landscapes for each optimal binding affinity w_{opt}. Closed symbols and dashed line represent the mean. In (C), $\mu = 0.1$, $\sigma = 1$ while in (D) $\mu = 0.1$, $\sigma = 0.15$.

the landscape with $m_{opt} = 0$ comprised a single peak with one genotype, the landscape with $m_{opt} = 4$ comprised 70 peaks with 81 genotypes per peak, and the landscape with $m_{opt} = 8$ comprised 1 peak with 6,561 genotypes. Which landscape topography is most conducive to adaptive evolutionary change?

NK landscapes

On NK landscapes, the length of the greedy adaptive walk $\langle l \rangle$ is minimised when $w_{opt} = 0.5$ for $K < L - 1$, whereas for $K = L - 1$ (House-of-Cards landscapes), $\langle l \rangle$ is independent of w_{opt}. Moreover, the change in mean fitness at equilibrium is always positive, increases with L and K, and tends to be maximized at $w_{opt} = 0.5$ (Supplementary Material, Note 2). In sum, these results show that on biallelic landscapes, selection for an intermediate phenotypic value decreases the length of a greedy adaptive walk and increases mean fitness at equilibrium, despite increasing the overall ruggedness of the fitness landscape, relative to the genotype-phenotype landscape. Below, we show that similar results hold for the mismatch model and the empirical landscapes of TF-DNA interactions.

Mismatch model

In the fitness landscapes generated using the mismatch model, the length of the greedy adaptive walk averaged over all starting genotypes is given by

$$\langle l \rangle = \frac{\Sigma_{m=0}^{L}(a-1)^m \cdot \frac{L!}{(L-m)!m!} \cdot |m_{\text{opt}} - m|}{a^L}. \tag{5}$$

Fig. 7A shows this expression for $L = 8$ and $a = 4$. Due to the additive nature and degeneracy of this genotype-phenotype landscape, the fitness landscapes have monotonically decreasing fitness values as the mutational distance from any peak increases. Therefore, the length of any individual greedy adaptive walk is simply the absolute difference of the mismatch class of the starting genotype and that of the optimal mismatch class (i.e., $|m_{\text{opt}} - m|$). As such, the average length of the greedy adaptive walk is minimized when selecting for $m_{\text{opt}} = 6$ (Fig. 7A), because this maximizes the number of genotypes in adaptive peaks (Fig. 6B; 28 global peaks $\times$ 729 genotypes per peak). If instead, we chose to start the greedy walk from only non-peak genotypes, the walk would still be minimised for $m_{\text{opt}} = 6$, because this class has the maximum number of genotypes that are Hamming distance one away from the peak genotypes. Recall that in each fitness landscape generated upon selection for m_{opt}, all the peaks are of the same height and therefore, the greedy walk always terminates on a global peak. This is in contrast with the NK landscapes, and as we will see below, the empirical landscapes.

For each m_{opt}, we next calculated the change in mean fitness at equilibrium under deterministic mutation-selection dynamics, relative to the fitness landscape generated for $m_{\text{opt}} = 0$. To do so, we exploit the permutation-invariance of these landscapes to group genotypes into a lower-dimensional state space defined by mismatch class (Methods). Specifically, we construct a transition matrix that defines the probability that a genotype from one mismatch class mutates into another, based on the frequency and fitness of genotypes in each mismatch class. We iterate this matrix until we reach steady state, which is guaranteed by the Frobenius-Perron theorem to be independent of initial conditions, because the matrix is irreducible[78]. Fig. 7C shows the change in mean fitness at equilibrium in relation to the optimal mismatch class $m_{\text{opt}} = 0$. Our first observation is that the change in mean fitness is always positive for $m_{\text{opt}} > 0$ when $\mu = 0.1$ and $\sigma = 1$, even though selection for such mismatch classes always causes an increase in the number of peaks in the fitness landscape, relative to the underlying genotype-phenotype landscape (Fig. 6B). Our second observation is that the change in mean fitness relative to m_{opt} is always unimodal. For example, selection for $m_{\text{opt}} = 6$ maximizes the change in mean fitness when $\mu = 0.1$ and $\sigma = 1$ (Fig. 7C). More generally, the m_{opt} that maximizes the change in mean fitness depends on the interplay between the mutation rate μ and the strength of selection σ. See Supplementary Fig. 10 for the phase diagram. When selection is strong or mutation is weak, mismatch class $m_{\text{opt}} = 8$ maximizes the change in mean fitness. As the strength of selection decreases or the mutation rate increases, the mismatch class that maximizes the change in mean fitness decreases from $m_{\text{opt}} = 8$ to $m_{\text{opt}} = 7$ and then to $m_{\text{opt}} = 6$, where it remains for weak selection or high mutation rates. This is because when σ is small (i.e., selection is strong), it is costly to step down from a peak and therefore, selecting for the class with the broadest peak ($m_{\text{opt}} = 8$) leads to the highest equilibrium mean population

fitness. As σ increases and selection becomes weaker, it is no longer as costly to step down from a peak and therefore, selecting for the class with the maximum number of genotypes in peaks ($m_{\text{opt}} = 6$) maximizes mean fitness at equilibrium. While this effect is reminiscent of the "survival of the flattest"[79] phenomenon, the crucial difference is that in the mismatch model, the peaks always have the same height and thus, there is no trade-off between peak height and width. Another way of altering the strength of selection is by changing e, the energetic cost of a mismatch. Larger e corresponds to stronger selection and the phase diagram changes accordingly (Supplementary Fig. 11).

Empirical landscapes

The empirical genotype-phenotype landscapes are topologically more complex[77,80] than the genotype-phenotype landscapes generated with the mismatch model, which are regular graphs (i.e., every genotype has $(a-1)\cdot L$ mutational neighbors). We therefore used simulations to calculate the average length of a greedy adaptive walk $\langle l \rangle$ in these empirical landscapes[81], initiating the walks from all non-peak genotypes in the fitness landscape. Moreover, each binding affinity measurement (E-score) in the empirical landscapes is associated with a noise threshold that is used to determine whether two genotypes truly differ in phenotype (Methods). This noise threshold can cause landscapes to have large non-peak plateaus, in which many genotypes have indistinguishable fitness. We therefore modified the greedy adaptive walk such that when a non-peak plateau was encountered, we chose a random mutational neighbor of indistinguishable fitness for the next step in the walk, disallowing reversion mutations. We repeated this process until the plateau was traversed and a sequence with higher fitness was reached. Finally, we terminated the walk when a peak sequence was reached. Therefore, the walk was primarily a deterministic greedy walk, with some stochasticity due to the non-peak plateaus.

Fig. 7B shows $\langle l \rangle$, averaged over 100 simulations of the adaptive walk from each initial condition, in relation to w_{opt}. In contrast to the mismatch model, $\langle l \rangle$ is shortest for $w_{\text{opt}} = 0.35$ and $w_{\text{opt}} = 0.5$ and slightly higher for intermediate w_{opt}. This is because, as in the mismatch model, $\langle l \rangle$ is correlated with the total number of genotypes in peaks, which depends upon both the number of peaks and their widths. Whereas selecting for $w_{\text{opt}} = 0.35$ leads to the largest number of peaks (Fig. 6D), selecting for $w_{\text{opt}} = 0.5$ leads to the broadest peaks (Supplementary Fig. 6B), thus explaining the minimisation of $\langle l \rangle$ when selecting for these extreme phenotypes. When selecting for low or intermediate phenotypic values, most walks terminate at a local, rather than the global fitness peak. However, these local peaks tend to be nearly as high as the global peak, especially when selecting for intermediate phenotypic values (Supplementary Fig. 9).

Next, we simulated deterministic mutation selection dynamics (Methods). For each w_{opt}, Fig. 7D shows the change in mean fitness at equilibrium for $\mu = 0.1, \sigma = 0.15$, relative to the fitness landscape generated after selecting for $w_{\text{opt}} = 0.5$. As in the mismatch model, the change in mean fitness tends to be positive, despite the increase in the number of peaks caused by selection for low or intermediate phenotypic values (Fig. 6D). Moreover, the mean change in fitness at equilibrium is maximized when selection favors an intermediate phenotypic value ($w_{\text{opt}} = 0.425$ for $\mu = 0.1, \sigma = 0.15$). This can be explained by the average peak heights and widths that occur when selecting for intermediate phenotypic values (Supplementary Fig. 6A and B for $\sigma = 0.15$). Exactly which w_{opt} maximizes the change in mean fitness depends on the interplay between the mutation rate μ and

the strength of selection σ (Supplementary Fig. 12), converging on 0.425 for large μ and σ, similar to the convergence seen at $m_{\text{opt}} = 6$ for the mismatch model.

Discussion

Non-linear relationships between phenotype and fitness have been observed in a diversity of biological systems[28,31,33,35,82,83], often reflecting a trade-off between the costs and benefits of a particular phenotype, such as antibiotic resistance[84,85]. It is well established that these non-linearities, either in the genotype-phenotype or phenotype-fitness map are a cause of epistasis[36–38,72]. Moreover, when mutations have epistatic interactions in their contribution to phenotype, a non-linear phenotype-fitness map can change the form of these interactions from negative to positive, or vice versa[16], as well as introduce or remove sign epistasis[82]. Our work complements these empirical observations and expands upon previous theoretical work[72,86], by systematically quantifying how and how often selection for a low or intermediate phenotypic value introduces or removes epistasis in the fitness landscape. Specifically, we show that the probability of changing the type of magnitude epistasis (e.g., positive to negative) is highest when selecting for low phenotypic values and the probability of introducing or removing sign epistasis is highest when selecting for intermediate phenotypic values. Further, we show that the simple sign epistasis motif cannot be converted into reciprocal sign epistasis, implying that genotype-phenotype landscapes with mainly simple sign epistasis motifs will remain globally congruent with their corresponding fitness landscapes.

Another key finding of our analysis is that selection for low or intermediate phenotypic values is more likely to increase than to decrease the number of peaks, with the probability of the two types of change being equal only in House-of-Cards genotype-phenotype landscapes. This means that additive genotype-phenotype landscapes will tend to be incongruent with their fitness landscapes, whereas rugged genotype-phenotype landscapes will not. While increased landscape ruggedness is typically thought to frustrate the evolutionary process, because it limits the amount of adaptive phenotypic variation mutation can bring forth[4], our evolutionary simulations show this need not be the case. Specifically, we find that the rugged fitness landscapes caused by selection for low or intermediate phenotypic values comprise local adaptive peaks that are nearly as tall as the global adaptive peak. Moreover, these local peaks tend to be accessible from throughout the landscape via a small number of sequential mutations that monotonically increase fitness. As a result, the mean population fitness at equilibrium is almost always higher when selecting for low or intermediate phenotypic values than when selecting for a high phenotypic value.

Finally, while there have been several attempts at investigating genotype-phenotype-fitness landscapes in the past[72,87,88] – some models have been very specific to the system of interest and others are agnostic to any mechanistic details[89]. We tried to bridge this gap, by applying a Fisher's Geometric model-like phenotype-fitness function to biophysically motivated and empirically determined genotype-phenotype landscapes. Further, our results on the mismatch model and the 1,137 landscapes of TF-DNA interactions may help to explain the prevalence of low- and intermediate-affinity binding sites in the control of gene expression. Prior work has suggested an entropic argument[90]: As with certain RNA secondary structures[91] or regulatory circuit motifs[92], low- and intermediate-affinity binding sites appear more frequently simply because they are more "findable". That is,

because a transcription factor binds more distinct DNA sequences with low or intermediate affinity than with high affinity, low- and intermediate-affinity binding sites are more likely to evolve to control gene expression. Our work complements this "arrival of the frequent" argument[93] by showing that low- and intermediate-affinity binding sites are not only more likely to arise *de novo* due to their increased frequency, selection for such sites also generates fitness landscapes that are more conducive to adaptation – in terms of increased mean fitness at equilibrium and decreased average length of adaptive walks, than fitness landscapes that were generated by selection for high affinity binding sites.

We made several modeling assumptions to simplify our analyses of transcription-factor DNA interactions, the relaxation of which may open new avenues for future research. First, we assumed a single nonlinearity in the relationship between genotype, phenotype, and fitness. As noted by Domingo et al.[37], "from transcription to RNA processing, translation, and protein folding and all the way up to protein activity and cellular fitness, there are many layers of biological organization where the effects of a mutation can be transformed." To date, our knowledge of how such nonlinearities combine to modify genotype-phenotype landscapes is based on a small number of experimental studies (e.g., ref[16]). A systematic theoretical analysis could build off the work presented here, for example by incorporating the sigmoidal relationship between binding site occupancy and binding affinity in modeling transcription factor-DNA interactions[94], such that fitness depends nonlinearly on occupancy, rather than affinity. Other nonlinearities, such as those caused by transcription factor cooperativity[94], could also be included.

Second, we assumed that selection acts directly on a single phenotype — binding affinity. While this assumption is common in models of the evolution of transcription factor binding sites[21,43,46] and is supported by empirical data[43,44], the relationship between binding affinity and fitness is not so direct, because it is modulated by gene expression. Gene expression depends on a variety of factors, including the presence, arrangement, and affinities of binding sites for other competing or cooperating transcription factors in promoters and enhancers[95], as well as local sequence context[96], chromatin context[97], DNA methylation[98], and local transcription factor concentrations[99]. Existing modeling frameworks that relate the architecture of entire regulatory regions to gene expression patterns may provide a path forward[100], facilitating the study of landscape incongruence when fitness depends upon the multitude of molecular phenotypes characteristic of eukaryotic gene regulation. Alternatively, our modeling framework could be extended to include multiple phenotypes by defining fitness in terms of the differences between a vector of phenotypes and a vector of optimal phenotypes, rather than the scalars considered here. Incongruence could then be quantified between the fitness landscape generated by selecting for the highest phenotypic value of each phenotype and that generated by selecting for a combination of low and intermediate values of the phenotypes. Our results correspond to a special case of this extended model, wherein all phenotypes except one are exactly attuned to their optimal values.

Third, we assumed a Gaussian phenotype-fitness map, which is commonly employed in a diversity of modeling frameworks[29,68,70,101], including those for transcription factor-DNA interactions[46]. Alternative symmetric phenotype-fitness maps (e.g., ref.[102]) will only affect our results quantitatively, because many of our findings, such as the changes in sign epistasis motifs and number of peaks, only depend on the rank ordering of fitness values. However, we expect asymmetric phenotype-

fitness maps, such as those uncovered in experimental studies of biological systems such as viruses[30] and yeast[103], to affect our results qualitatively. For example, in our analyses of NK landscapes, we often observed symmetries in incongruence around an intermediate phenotypic value. These symmetries will almost certainly be lost. Understanding how asymmetric phenotype-fitness maps affect the incongruence of genotype-phenotype landscape is therefore an outstanding challenge.

Finally, our study may open new lines of research on dynamic genotype-phenotype and fitness landscapes[104–110]. For example, whereas we studied selection for a fixed phenotypic optimum, this optimum may in fact change in space or in time. Our results imply that even gradual changes in the phenotypic optimum may lead to abrupt changes in fitness landscape topography, which may have implications for an evolving population's ability to track this optimum and thus for population persistence and extinction. Moreover, because our measures of incongruence can be applied to any pair of landscapes so long as they are defined over the same set of genotypes, they may shed light on the kinds of topographical alterations induced by fluctuating environmental factors, such as DNA methylation[98] or the presence of protein partners[111] on transcription factor-DNA interactions. As our ability to experimentally interrogate such complexities in the relationship between genotype, phenotype, and fitness continues to improve, we anticipate a sharpened focus on landscape dynamics and their implications for the evolutionary process.

Methods

Epistasis

In a genotype-phenotype landscape, we classify the type of magnitude epistasis between a pair of loci using the following linear combination of phenotypic values:

$$\varepsilon_{\mathrm{gp}} = w_{00} + w_{11} - w_{01} - w_{10}, \tag{6}$$

where w_i represents the phenotype of genotype $i \in \{0,1\}^2$. When $\varepsilon_{\mathrm{gp}} = 0$, there is no magnitude epistasis, because the phenotypic effects of the two alleles combine additively; when $\varepsilon_{\mathrm{gp}} > 0$, there is positive epistasis, because the phenotypic effects of the two alleles are greater than expected based on their individual phenotypic effects; when $\varepsilon_{\mathrm{gp}} < 0$, there is negative epistasis, because the phenotypic effects of the two alleles are less than expected based on their individual phenotypic effects.

Analogously, in the fitness landscape, we classify the type of magnitude epistasis between a pair of loci using the following linear combination of fitness values:

$$\varepsilon_{\mathrm{f}} = F_{00} + F_{11} - F_{01} - F_{10}, \tag{7}$$

where F_i represents the fitness of the corresponding genotype i. As in the genotype-phenotype landscape, $\varepsilon_{\mathrm{f}} = 0$, $\varepsilon_{\mathrm{f}} > 0$, and $\varepsilon_{\mathrm{f}} < 0$ indicate additive, positive, and negative epistatic interactions among loci, respectively.

We use $\varepsilon_{\mathrm{gp}}$ and ε_{f} to calculate the fraction of pairs of loci that have the same type of epistasis in the genotype-phenotype landscape and the fitness landscape, which we determine as the product of $\varepsilon_{\mathrm{gp}}$ and ε_{f}. This is because the type of epistasis is the same in the two landscapes when $\varepsilon_{\mathrm{gp}} \cdot \varepsilon_{\mathrm{f}} > 0$. While it is theoretically possible for both $\varepsilon_{\mathrm{gp}}$ and ε_{f} to be 0, in which case the type of epistasis would be the same in the two landscapes yet the condition $\varepsilon_{\mathrm{gp}} \cdot \varepsilon_{\mathrm{f}} > 0$ would not be satisfied, this never happens in practice.

NK Landscapes

We constructed the NK landscapes using the adjacency neighbourhood scheme, wherein each locus i of a genotype of length L, interacts with K adjacent loci to the right of locus i and $0 \leq K \leq L-1$. We used periodic boundary conditions, such that the L-th locus interacts with the first K loci, and so on.

The phenotype $w(\tau)$ of genotype τ is computed as the sum of the individual contributions of all loci, each of which depends on K other interacting loci:

$$w(\tau) = \Sigma_{i=1}^{L} f(\tau_i; \tau_i^1, \tau_i^2 \ldots \tau_i^K), \tag{8}$$

where $f(\tau_i; \tau_i^1, \tau_i^2 \ldots \tau_i^K)$ represents the contribution of the i-th locus, which depends on K other loci $\tau_i^1, \tau_i^2 \ldots \tau_i^K$. We drew the contributions of each of the 2^{K+1} possible configurations from a uniform distribution between 0 and 1. Finally, we re-scaled the phenotypic values by subtracting the minimum and dividing by the maximum, such that they fell between 0 and 1.

Empirical genotype-phenotype landscapes of transcription factor-DNA interactions

We studied the incongruence of 1,137 genotype-phenotype landscapes of transcription factor-DNA interactions. The procedure for constructing these landscapes has been described elsewhere[21]. In brief, each landscape corresponds to a single transcription factor, the genotypes it contains represent DNA sequences of length $L = 8$ that specifically bind the transcription factor, and the phenotype of each sequence is a quantitative proxy for relative binding affinity, which defines the surface of the landscape. These phenotypes are reported as enrichment scores (E-scores) derived from protein binding microarrays. In each landscape, two genotypes are considered mutational neighbors if they differ by a single small mutation, specifically a point mutation or a small indel that shifts an entire contiguous binding site by a single base[77]. We performed all analyses on the dominant genotype network (i.e., the largest connected component of the network), which comprises the vast majority of genotypes in each landscape[21].

We used the Genonets Python package (version 0.31) to characterize the topographical properties of the empirical genotype-phenotype landscapes[112]. Specifically, we used this package to compute the number of peaks per landscape and the number of genotypes in the peaks of each landscape. These calculations rely on a noise threshold δ, which is used to determine whether two genotypes actually differ in phenotype. For each transcription factor, we used the value of δ reported in ref.[21], which was derived from a comparison of binding affinity measurements across two protein binding microarray designs.

While characterizing the topographies of the resulting fitness landscapes, we also had to transform δ following the rules of error propagation. Accordingly, the noise in fitness values, dF, depends on the noise in the phenotypic values $dw = \delta$ as follows:

$$dF = -2F \cdot \frac{w - w_{\text{opt}}}{\sigma^2} \cdot \delta. \tag{9}$$

To compute the number of peaks and their widths in the fitness landscapes, we adapted Genonets to specify different noise threshold values (dF) for each genotype.

Mutation-selection dynamics for the mismatch model

We grouped genotypes according to their mismatch class m and iterated a series of selection and mutation steps until the population reached an equilibrium distribution. In each selection step, the frequency of each mismatch class X_m was scaled by its fitness F_m and then normalized:

$$X_m^S(t+1) = \frac{F_m \cdot X_m(t)}{\Sigma_i F_i \cdot X_i}, \tag{10}$$

where $X_m^S(t+1)$ is the frequency of the mismatch class m after the selection step.

The selection step was followed by a mutation step, in which genotypes could either mutate within their mismatch class or mutate to an adjacent mismatch class. The total mutation probability is μ, so with probability $1-\mu$, the genotype does not mutate. The frequency of each mismatch class was thus updated as

$$X_m^M(t+1) = \left[\mu \cdot \frac{L-(m-1)}{L}\right] \cdot X_{m-1}^S(t+1) + \left[(1-\mu) + \mu \cdot \frac{(a-2)\cdot m}{(a-1)\cdot L}\right] \cdot X_m^S(t+1) + \left[\mu \cdot \frac{m+1}{(a-1)\cdot L}\right] \cdot X_{m+1}^S(t+1), \tag{11}$$

where $X_m^M(t+1)$ is the frequency of the mismatch class m after the mutation step and μ is the mutation probability. The mutation step implicitly accounts for the different number of genotypes per mismatch class. Finally, $X_m(t+1)$ was set to $X_m^M(t+1)$ and t was incremented. This process was repeated until equilibrium.

Mutation-selection dynamics for the empirical landscapes

The empirical landscapes comprise far fewer genotypes than the mismatch landscape. We therefore defined the state space of the empirical landscapes in terms of the individual genotypes in each landscape, merging each genotype with its reverse complement[77]. The mutational neighbors of each genotype were those that differed by a single small mutation, namely a point mutation or an indel that shifted an entire contiguous binding site by a single base[77]. In this analysis, we did not account for uncertainties in the fitness values, because these do not influence our results in this evolutionary regime (see below).

The recursion relation for mutation-selection dynamics in discrete time is

$$x_i(t+1) = \sum_j \mu^{d(\tau_i,\tau_j)} \cdot (1-\mu)^{L-d(\tau_i,\tau_j)} \frac{f_j}{\bar{f}(\vec{x},t)} x_j(t), \tag{12}$$

where $\vec{x}$ is the vector of genotype frequencies, x_i is the frequency of the ith genotype, μ is the probability of a single point mutation or a small indel mutation, L is the length of the genotypes, $d(\tau_i,\tau_j)$ is the minimum of the mutational distance between genotypes τ_i and τ_j and between genotypes τ_i and the reverse compliment of τ_j, f_i is the fitness of the ith genotype, and $\bar{f}(\vec{x},t)$ is the mean population fitness at time t.

We linearized the dynamics by substituting $z_i(t) = \frac{x_i(t)}{\prod_{\tau=1}^{t-1} \bar{f}(\vec{x},\tau)}$, yielding

$$z_i(t+1) = \sum_j \mu^{d(\tau_i,\tau_j)} (1-\mu)^{L-d(\tau_i,\tau_j)} f_j z_j(t), \tag{13}$$

from which we retrieved the normalized genotype frequencies with $\vec{x} = \vec{z}/(\sum_i z_i)$. In matrix form, the dynamics are

$$\vec{z}(t+1) = M \cdot S \cdot \vec{z}(t), \tag{14}$$

where the mutation matrix M has elements $M_{ij} = \mu^{d(\tau_i,\tau_j)} \cdot (1-\mu)^{L-d(\tau_i,\tau_j)}$ and the selection matrix S is a diagonal matrix with $S_{ii} = f_i$, where f_i is the fitness of sequence τ_i. We calculated the eigenvector corresponding to the largest eigenvalue of $M \cdot S$ to determine the equilibrium state of the dynamics, which is guaranteed by the Frobenius-Perron theorem to be unique and stable.

Because the fitness values have some uncertainty around them, we additionally performed a sensitivity analysis by adding Gaussian noise to the fitness values, with standard deviation equal to $1/3 \cdot dF$, such that 99.7% of the sampled fitness values fell within $F \pm \delta$. Our results are robust to these perturbations, with the exception of two parameter combinations – $\sigma = 0.1, \mu = 0.001$ and $\sigma = 0.1, \mu = 0.0025$. For small σ, dF is large and therefore such sensitivity is expected. However, the quantitative changes in our results were small, leading us to conclude that fitness uncertainties do not significantly influence our results in this evolutionary regime.

Acknowledgements

This work was supported by the Swiss National Science Foundation (Grant numbers PP00P3_170604 and 310030_192541). We thank members of the Computational Biology Group at ETH, Michael Manhart, Joachim Krug, and David M. McCandlish for discussions.

Data Availability

All data used in this study are publicly available. See Aguilar-Rodriguez et al. (2017) "A thousand empirical adaptive landscapes and their navigability." Nature Ecology & Evolution, 1, 0045. Code used in this study is available here.

References

1. Alberch, P. From genes to phenotype: dynamical systems and evolvability. *Genetica* **84**, 5–11 (1991).
2. Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nat. Rev. Genet.* **14**, 168–178 (2013).
3. Julien, P., Miñana, B., Baeza-Centurion, P., Valcárcel, J. & Lehner, B. The complete local genotype–phenotype landscape for the alternative splicing of a human exon. *Nat. Commun.* **7** (2016).
4. Payne, J. L. & Wagner, A. The causes of evolvability and their evolution. *Nat. Rev. Genet.* **20**, 24–38 (2019).
5. Poelwijk, F., Tănase-Nicola, S., Kiviet, D. & Tans, S. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* **272**, 141–4 (2011).
6. Østman, B., Hintze, A. & Adami, C. Impact of epistasis and pleiotropy on evolutionary adaptation. *Proc. Royal Soc. B: Biol. Sci.* **279**, 247–256 (2012).
7. Weinreich, D., Watson, R. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).
8. Kinney, J. B. & McCandlish, D. M. Massively parallel assays and quantitative sequence–function relationships. *Annu. Rev. Genomics Hum. Genet.* **20**, 99–127 (2019).
9. Qiu, C. *et al.* High-resolution phenotypic landscape of the RNA polymerase II trigger loop. *PLoS Genet.* **12** (2016).
10. Wu, N., Dai, C. L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**, e16965 (2016).
11. Tack, D. *et al.* The genotype-phenotype landscape of an allosteric protein. *Mol. Syst. Biol.* **17** (2021).
12. Sarkisyan, K. S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397–401 (2016).
13. de Boer, C. G. *et al.* Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat. Biotechnol.* **38**, 56–65 (2019).
14. Lagator, M., Paixao, T., Barton, N. H., Bollback, J. P. & Guet, C. C. On the mechanistic nature of epistasis in a canonical cis-regulatory element. *eLife* **6**, e25192 (2017).
15. Schaerli, Y. *et al.* Synthetic circuits reveal how mechanisms of gene regulatory networks constrain evolution. *Mol. Syst. Biol.* **14** (2018).
16. Li, X., Lalic, J., Baeza-Centurion, P., Dhar, R. & Lehner, B. Changes in gene expression predictably shift and switch genetic interactions. *Nat. Commun.* **10** (2019).
17. Bassalo, M. C. *et al.* Deep scanning lysine metabolism in escherichia coli. *Mol. Syst. Biol.* **14** (2018).

656 **18.** Schenk, M. F., Szendro, M. L., Salverda, I. G., Krug, J. & de Visser, J. A. G. Patterns of epistasis between beneficial
657 mutations in an antibiotic resistance gene. *Mol. Biol. Evol.* **30**, 1779–87 (2013).

658 **19.** Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein.
659 *eLife* **2**, e00631 (2013).

660 **20.** Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. & Chen, I. A. Comprehensive fitness landscape for
661 RNA. *Proc. Natl. Acad. Sci.* **110**, 376–379 (2013).

662 **21.** Aguilar-Rodríguez, J., Payne, J. & Wagner, A. A thousand empirical adaptive landscapes and their navigability. *Nat. Ecol.*
663 *Evol.* **1** (2017).

664 **22.** Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire
665 protein domain. *Curr. Biol.* **24**, 2643–2651 (2014).

666 **23.** Hartman, E. C. *et al.* Experimental evaluation of coevolution in a self-assembling particle. *Biochemistry* **58**, 1527–1538
667 (2019).

668 **24.** Pitt, J. & Ferré-D'Amaré, A. Rapid construction of empirical RNA fitness landscapes. *Science* **330**, 376–379 (2010).

669 **25.** Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proc. 6th Int. Congr. Genet.*
670 356–366 (1932).

671 **26.** Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**
672 (2009).

673 **27.** Kinsler, G., Geiler-Samerotte, K. & Petrov, D. Fitness variation across subtle environmental perturbations reveals local
674 modularity and global pleiotropy of adaptation. *eLife* **9**, e61271 (2020).

675 **28.** Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–592
676 (2005).

677 **29.** Lynch, M. & Hagner, K. Evolutionary meandering of intermolecular interactions along the drift barrier. *Proc. Natl. Acad.*
678 *Sci.* **112** (2015).

679 **30.** Rokyta, D. R. *et al.* Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. *PLoS*
680 *Genet.* **7** (2011).

681 **31.** Duveau, F., Toubiana, W. & Wittkopp, P. J. Fitness effects of cis-regulatory variants in the saccharomyces cerevisiae
682 TDH3 promoter. *Mol. Biol. Evol.* **34**, 2908–2912 (2017).

683 **32.** Rest, J. S. *et al.* Nonlinear Fitness Consequences of Variation in Expression Level of a Eukaryotic Gene. *Mol. Biol. Evol.*
684 **30**, 448–456 (2012).

685 **33.** Keren, L. *et al.* Massively parallel interrogation of the effects of gene expression levels on fitness. *Cell* **166**, 1–13 (2016).

686 **34.** Labourel, F. & Rajon, E. Resource uptake and the evolution of moderately efficient enzymes. *Mol. Biol. Evol.* **38**,
687 3938–3952 (2021).

688 **35.** Perfeito, L., Ghozzi, S., Berg, J., Schnetz, K. & Lässig, M. Nonlinear fitness landscape of a molecular pathway. *PLoS*
689 *Genet.* **7** (2011).

690 **36.** Otwinowski, J., McCandlish, D. & Plotkin, J. Inferring the shape of global epistasis. *Proc. Natl. Acad. Sci.* **115**,
691 E7550–E7558 (2018).

692 **37.** Domingo Espinós, J., Baeza-Centurion, P. & Lehner, B. The causes and consequences of genetic interactions (epistasis).
693 *Annu. Rev. Genomics Hum. Genet.* **20**, 17.1–17.28 (2019).

694 **38.** Kryazhimskiy, S. Emergence and propagation of epistasis in metabolic networks. *eLife* **10**, e60200 (2021).

695 **39.** Crocker, J., Preger-Ben Noon, E. & Stern, D. The soft touch: Low-affinity transcription factor binding sites in development
696 and evolution. *Curr. Top. Dev. Biol.* **117**, 455–469 (2015).

697 **40.** Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell*
698 *Biol.* **19**, 621–637 (2018).

699 **41.** Shultzaberger, R., Malashock, D., Kirsch, J. & Eisen, M. The fitness landscapes of cis-acting binding sites in different
700 promoter and environmental contexts. *PLoS Genet.* **6** (2010).

701 **42.** Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically
702 designed promoters. *Nat. Biotechnol.* **30**, 521–530 (2012).

703 **43.** Mustonen, V., Kinney, C., J.and Callan & Lässig, M. Energy-dependent fitness: a quantitative model for the evolution of
704 yeast transcription factor binding sites. *Proc. Natl. Acad. Sci.* **105**, 12376–12381 (2008).

705 **44.** Haldane, A., Manhart, M. & Morozov, A. Biophysical fitness landscapes for transcription factor binding sites. *PLoS*
706 *Comput. Biol.* **10**, 1–16 (2014).

707 **45.** Gerland, U. & Hwa, T. On the selection and evolution of regulatory DNA motifs. *J. Mol. Evol.* **55**, 386–400 (2002).

708 **46.** Berg, J., Willmann, S. & Lässig, M. Adaptive evolution of transcription factor binding sites. *BMC Evol. Biol.* **4** (2004).

709 **47.** Tuğrul, M., Paixão, T., Barton, N. & Tkačik, G. Dynamics of transcription factor binding site evolution. *PLoS Genet.* **11**,
710 273–283 (2015).

711 **48.** Weirauch, M. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**,
712 1431–1443 (2014).

713 **49.** Crocker, J. *et al.* Low affinity binding site clusters confer HOX specificity and regulatory robustness. *Cell* **160**, 191–203
714 (2015).

715 **50.** Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-affinity binding sites and the transcription factor
716 specificity paradox in eukaryotes. *Annu. Rev. Cell Dev. Biol.* **35**, 357–379 (2019).

51. Grönlund, A., Lötstedt, P. & Elf, J. Transcription factor binding kinetics constrain noise suppression via negative feedback. *Nat. Commun.* **4** (2013).

52. Farley, E., Olson, K., Zhang, W. & et al. Suboptimization of developmental enhancers. *Science* **350**, 325–28 (2015).

53. Farley, E., Olson, K., Zhang, W., Rokhsar, D. & Levine, M. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc. Natl. Acad. Sci.* **113**, 6508–6513 (2016).

54. Tanay, A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res.* 962–972 (2006).

55. Ramos, A. & Barolo, S. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philos. Transactions Royal Soc. B* **368** (2013).

56. Tsai, A. *et al.* Nuclear microenvironments modulate transcription from low-affinity enhancers. *eLife* **6**, e28975 (2017).

57. Cary, G. A., Jarvela, A. M. C., Francolini, R. D. & Hinman, V. F. Genome-wide use of high- and low-affinity Tbrain transcription factor binding sites during echinoderm development. *Proc. Natl. Acad. Sci.* **114**, 5854–5861 (2017).

58. Wang, J., Malecka, A., Trøen, G. & Delabie, J. Comprehensive genome-wide transcription factor analysis reveals that a combination of high affinity and low affinity DNA binding is needed for human gene regulation. *BMC genomics* **16** (2015).

59. Clarke, B. The evolution of genetic diversity. *Proc. Royal Soc. London. Ser. B, Biol. Sci.* **205**, 453–474 (1979).

60. Gavrilets, S. Fitness landscapes and the origin of species. *Princet. Univ. Press.* (2004).

61. Misevic, D., Kouyos, R. D. & Bonhoeffer, S. Predicting the evolution of sex on complex fitness landscapes. *PLoS Comput. Biol.* **5** (2009).

62. de Visser, J. A. G. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15**, 480–490 (2014).

63. Aita, Y., Takuyo anf Husimib. Fitness spectrum among random mutants on mt. fuji type fitness landscape. *J. Theor. Biol.* **182**, 469–485 (1996).

64. Kingman, J. On the properties of bilinear models for the balance between genetic mutation and selection. *Math. Proc. Camb. Philos. Soc.* **81**, 443–453 (1977).

65. Kauffman, S. A. & Weinberger, E. D. The NK model of rugged fitness landscapes and its application to the maturation of the immune response. *J. Theor. Biol.* **141**, 211–245 (1989).

66. Berg, O. G. & von Hippel, P. H. Selection of DNA binding sites by regulatory proteins. *J. Mol. Biol.* **193**, 723–750 (1987).

67. Poelwijk, F., Kiviet, D., Weinreich, D. & Tans, S. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–386 (2007).

68. Fisher, R. A. The genetical theory of natural selection. *Clarendon Press. Oxford.* (1930).

69. Tenaillon, O. The utility of fisher's geometric model in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* **45**, 179–201 (2014).

70. Dieckmann, U. & Doebeli, M. On the origin of species by sympatric speciation. *Nature* **400**, 354–357 (1999).

71. Gao, H., Granka, J. M. & Feldman, M. W. On the classification of epistatic interactions. *Genetics* **184**, 827–837 (2010).

72. Hwang, S., Park, S.-C. & Krug, J. Genotypic complexity of fisher's geometric model. *Genetics* **206**, 1049–1079 (2017).

73. Hordijk, W. & Stadler, P. Amplitude spectra of fitness landscapes. *Adv. Complex Syst.* **1**, 39–66 (1998).

74. Macken, C. A. & Perelson, A. S. Protein evolution on rugged landscapes. *Proc. Natl. Acad. Sci.* **86**, 6191–6195, (1989).

75. Badis-Breard, G. *et al.* Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009).

76. Berger, M. *et al.* Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* **24**, 1429–35 (2006).

77. Payne, J. & Wagner, A. The robustness and evolvability of transcription factor binding sites. *Science* **343**, 875–877 (2014).

78. Caswell, H. *Matrix population models* (Sinauer, Sunderland, Mass., 2001).

79. Wilke, C., Wang, J., Ofria, C., Lenski, R. & Adami, C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**, 331–333 (2001).

80. Aguilar-Rodriguez, J., Peel, L., Stella, M., Wagner, A. & Payne, J. The architecture of an empirical genotype-phenotype map. *Evolution* **72**, 1242–1260 (2018).

81. Franke, J., Klozer, A., de Visser, J. & Krug, J. Evolutionary accessibility of mutational pathways. *PLoS Comput. Biol.* **7** (2011).

82. Chou, H.-H., Delaney, N., Draghi, J. & Marx, C. Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLoS genetics* **10** (2014).

83. Kemble, H. *et al.* Flux, toxicity, and expression costs generate complex genetic interactions in a metabolic pathway. *Sci. Adv.* **6** (2020).

84. Das, S. G., Direito, S. O., Waclaw, B., J Allen, R. & Krug, J. Predictable properties of fitness landscapes induced by adaptational tradeoffs. *eLife* e55155 (2020).

85. Pinheiro, F., Warsi, O., Andersson, D. I. & Lässig, M. Metabolic fitness landscapes predict the evolution of antibiotic resistance. *Nat. Ecol. Evol.* **5**, 677–687 (2021).

86. Blanquart, F., Achaz, G., Bataillon, T. & Tenaillon, O. Properties of selected mutations and genotypic landscapes under fisher's geometric model. *Evolution* **68**, 3537–3554 (2014).

87. Friedlander, T., Prizak, R., Barton, N. H. & Tkačik, G. Evolution of new regulatory functions on biophysically realistic fitness landscapes. *Nat. Commun.* **8** (2017).

88. Bershtein, S., Serohijos, A. W. & Shakhnovich, E. I. Bridging the physical scales in evolutionary biology: from protein sequence space to fitness of organisms and populations. *Curr. Opin. Struct. Biol.* **42**, 31–40 (2017).

89. Fragata, I., Blanckaert, A., Louro, M. A. D., Liberles, D. A. & Bank, C. Evolution in the light of fitness landscape theory. *Trends Ecol. & Evol.* **34**, 69–82 (2019).

90. He, X., Duque, T. & Sinha, S. Evolutionary origins of transcription factor binding site clusters. *Mol. Biol. Evol.* **29**, 1059–1070 (2011).

91. Dingle, K., Ghaddar, F., Sulc, P. & Louis, A. Phenotype bias determines how natural RNA structures occupy the morphospace of all possible shapes. *Mol. Biol. Evol.* **39** (2022).

92. Xiong, K., Gerstein, M. & Masel, J. Differences in evolutionary accessibility determine which equally effective regulatory motif evolves to generate pulses. *Genetics* **219** (2021).

93. Schaper, S. & Louis, A. A. The arrival of the frequent: How bias in genotype-phenotype maps can steer populations to local optima. *PLoS One* **9**, e86635 (2014).

94. Bintu, L. *et al.* Transcriptional regulation by the numbers: Models. *Curr. Opin. Genet. Dev.* **15**, 116–124 (2005).

95. Siggers, T. & Gordân, R. Protein–DNA binding: Complexities and multi-protein codes. *Nucleic Acids Res.* **42**, 1–13 (2013).

96. Dror, I., Golan, T., Levy, C., Rohs, R. & Mandel-gutfreund, Y. A widespread role of the motif environment on transcription factor binding across diverse protein families. *Genome Res.* **25**, 1268–1280 (2015).

97. Xin, B. & Rohs, R. Relationship between histone modifications and transcription factor binding is protein family specific. *Genome Res.* **28**, 321–333 (2018).

98. Yin, Y. *et al.* Impact of cytosine methylation on dna binding specificities of human transcription factors. *Science* **356** (2017).

99. Tsai, A., Galupa, R. & Crocker, J. Robust and efficient gene regulation through localized nuclear microenvironments. *Development* **147** (2020).

100. Duque, T. *et al.* Simulations of enhancer evolution provide mechanistic insights into gene regulation. *Mol. Biol. Evol.* **31**, 184–200 (2013).

101. Martin, G., Elena, S. & Lenormand, T. Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat. genetics* **39**, 555–60 (2007).

102. Roughgarden, J. Species packing and the competition function with illustrations from coral reef fish. *Theor. population biology* **5**, 163–86 (1974).

103. Duveau, F. *et al.* Fitness effects of altering gene expression noise in saccharomyces cerevisiae. *eLife* **7**, e37272 (2018).

104. Vos, M. G. D., Dawid, A., Sunderlikova, V. & Tans, S. J. Breaking evolutionary constraint with a tradeoff ratchet. *Proc. Natl. Acad. Sci.* **112**, 14906–14911 (2015).

105. Steinberg, B. & Ostermeier, M. Environmental changes bridge evolutionary valleys. *Sci. Adv.* **2** (2016).

106. Hayden, E. J. & Wagner, A. Environmental change exposes beneficial epistatic interactions in a catalytic RNA. *Proc. Royal Soc. B: Biol. Sci.* **279**, 3418–3425 (2012).

107. Li, C. & Zhang, J. Multi-environment fitness landscapes of a tRNA gene. *Nat. Ecol. Evol.* **2**, 1025–1032 (2018).

108. Gorter, F. A., Aarts, M. G., Zwaan, B. J. & Visser, J. A. G. D. Local fitness landscapes predict yeast evolutionary dynamics in directionally changing environments. *Genetics* **208**, 307–322 (2018).

109. Bank, C., Matuszewski, S., Hietpas, R. T. & Jensen, J. D. On the (un)predictability of a large intragenic fitness landscape. *Proc. Natl. Acad. Sci.* **113**, 14085–14090 (2016).

110. Bajić, D., Vila, J., Blount, Z. & Sanchez, A. On the deformability of an empirical fitness landscape by microbial evolution. *Proc. Natl. Acad. Sci.* **115**, 201808485 (2018).

111. Barrera, L. *et al.* Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science* **351**, 1450–1454 (2016).

112. Khalid, F., Aguilar-Rodriguez, J., Wagner, A. & Payne, J. Genonets server - A web server for the construction, analysis, and visualization of genotype networks. *Nucleic Acids Res.* **44**, W70–W76 (2016).

113. Baldi, P. & Rinott, Y. Asymptotic normality of some graph-related statistics. *J. Appl. Probab.* **26**, 171–175 (1989).

Supplementary material

Derivations

Derivation 1: Derivation of equation 2

The epistasis in the fitness landscape is given by equation 7. We assume σ to be large, such that $(w_i - w_{\text{opt}})^2/\sigma^2 \ll 1 \; \forall i \in \{0,1\}^2$, an approximation that holds when selection is weak. Taylor expanding F_i and neglecting higher order terms, we find

$$F_i = \exp\left[-\left(\frac{w_i - w_{\text{opt}}}{\sigma}\right)^2\right] \approx 1 - (w_i - w_{\text{opt}})^2/\sigma^2, \tag{15}$$

where $i \in \{0,1\}^2$. Let $\Delta_i = w_i - w_{\text{opt}}$, then

$$\varepsilon_{\text{f}} \approx \frac{-1}{\sigma^2}\left(\Delta_{00}^2 + \Delta_{11}^2 - \Delta_{01}^2 - \Delta_{10}^2\right) = \frac{-1}{\sigma^2}\left[(\Delta_{00}+\Delta_{11})^2 - 2\Delta_{00}\Delta_{11} - (\Delta_{01}+\Delta_{10})^2 + 2\Delta_{10}\Delta_{01}\right] \tag{16}$$

$$= \frac{-1}{\sigma^2}\left[(\Delta_{00}+\Delta_{11}+\Delta_{01}+\Delta_{10})(\Delta_{00}+\Delta_{11}-\Delta_{01}-\Delta_{10}) - 2(\Delta_{00}\Delta_{11} - \Delta_{10}\Delta_{01})\right]$$

$$= \frac{-1}{\sigma^2}\left[(\Sigma_i \Delta_i)\varepsilon_{\text{gp}} - 2\varepsilon_{m,\Delta}\right]$$

where $\varepsilon_{m,\Delta} = \Delta_{00}\Delta_{11} - \Delta_{10}\Delta_{01} = \varepsilon_{m,\Delta} = (w_{00} - w_{\text{opt}})(w_{11} - w_{\text{opt}}) - (w_{01} - w_{\text{opt}})(w_{10} - w_{\text{opt}}) \implies \varepsilon_{m,\Delta} = \varepsilon_m - w_{\text{opt}} \cdot \varepsilon_{\text{gp}}$, and $\varepsilon_m = w_{00}w_{11} - w_{01}w_{10}$ is the multiplicative epistasis. Substituting for $\varepsilon_{m,\Delta}$, expanding the $\Sigma_i \Delta_i$ term and rearranging gives:

$$\varepsilon_{\text{f}} = \frac{1}{\sigma^2}\left[2\varepsilon_m + (2w_{\text{opt}} - \Sigma_i w_i)\cdot \varepsilon_{\text{gp}})\right] \implies \varepsilon_{\text{f}} \cdot \varepsilon_{\text{gp}} = \frac{1}{\sigma^2}\left[2\varepsilon_m \cdot \varepsilon_{\text{gp}} + (2w_{\text{opt}} - \Sigma_i w_i)\cdot \varepsilon_{\text{gp}}^2\right]. \tag{17}$$

Derivation 2: Derivation of equation 3

The absolute difference between the number of peaks in the genotype-phenotype landscape and the fitness landscape is

$$|p_{\text{f}} - p_{\text{gp}}| = \begin{cases} p_{\text{f}} - p_{\text{gp}}, & \text{if } p_{\text{f}} > p_{\text{gp}} \\ p_{\text{gp}} - p_{\text{f}}, & \text{if } p_{\text{f}} < p_{\text{gp}}. \end{cases} \tag{18}$$

We can rule out the case where $p_{\text{f}} = p_{\text{gp}}$, because the probability of change in the number of peaks tends towards 1 as L increases, so long as w_{opt} is sufficiently far from 1 (Fig. 4D). This implies

$$\langle |p_{\text{f}} - p_{\text{gp}}| \rangle = \langle p_{\text{f}} - p_{\text{gp}} \rangle_{p_{\text{f}} > p_{\text{gp}}} P(p_{\text{f}} > p_{\text{gp}}) + \langle p_{\text{gp}} - p_{\text{f}} \rangle_{p_{\text{gp}} > p_{\text{f}}} P(p_{\text{gp}} > p_{\text{f}}), \tag{19}$$

where $\langle p_{\text{f}} - p_{\text{gp}} \rangle_{p_{\text{f}} > p_{\text{gp}}}$ is the conditional mean of $p_{\text{f}} - p_{\text{gp}}$, given that $p_{\text{f}} > p_{\text{gp}}$.

As shown below, the probability of increase in the number of peaks is equal to the probability of decrease in the number of

peaks (Proof 3); i.e., $P(p_{\mathrm{f}} > p_{\mathrm{gp}}) = P(p_{\mathrm{gp}} > p_{\mathrm{f}}) = 0.5$. This implies

$$\langle |p_{\mathrm{f}} - p_{\mathrm{gp}}| \rangle = 0.5(\langle p_{\mathrm{f}} - p_{\mathrm{gp}} \rangle_{p_{\mathrm{f}} > p_{\mathrm{gp}}} + \langle p_{\mathrm{gp}} - p_{\mathrm{f}} \rangle_{p_{\mathrm{gp}} > p_{\mathrm{f}}}). \tag{20}$$

In genotype-phenotype landscapes generated by the House-of-Cards model, the phenotypic values of mutational neighbors are completely uncorrelated, so selection for a low or intermediate phenotypic value generates a fitness landscape in which the fitness values of mutational neighbors are also completely uncorrelated. Thus, the distribution of the number of peaks is the same for the genotype-phenotype landscape and the fitness landscape. Moreover, p_{f} is independent of p_{gp} because p_{f} depends upon the number of genotypes with phenotypes close to the optimal phenotypic value, and this number is independent of the number of genotypes with phenotypes close to 1. Therefore, p_{gp} and p_{f} are independent and identically distributed (i.i.d) random variables. Now as $L \to \infty$, the distribution of the number of peaks in the House-of-Cards model tends to a normal distribution[113] with mean $\frac{a^L}{(a-1)\cdot L+1}$ and variance $\tau^2 = \frac{a^L \cdot ((a-1)\cdot L-(a-1)}{2((a-1)\cdot L+1)^2}$[74], where a is the number of alleles at each site and here $a = 2$. This implies the distribution of the random variable $p_{\mathrm{f}} - p_{\mathrm{gp}}$ is also normally distributed with mean zero and variance $2\tau^2$; i.e.,

$$p_{\mathrm{f}} - p_{\mathrm{gp}} \coloneqq x \sim \mathcal{N}(0, 2\tau^2), \tag{21}$$

which is symmetric about zero, implying $\langle p_{\mathrm{f}} - p_{\mathrm{gp}} \rangle_{p_{\mathrm{f}} > p_{\mathrm{gp}}} = \langle p_{\mathrm{gp}} - p_{\mathrm{f}} \rangle_{p_{\mathrm{gp}} > p_{\mathrm{f}}}$. Thus,

$$\langle |p_{\mathrm{f}} - p_{\mathrm{gp}}| \rangle = \langle p_{\mathrm{f}} - p_{\mathrm{gp}} \rangle_{p_{\mathrm{f}} > p_{\mathrm{gp}}} = \frac{1}{\sqrt{2\pi}\tau'} \int_0^\infty x \cdot \exp\left[-\frac{x^2}{2\tau'^2}\right] dx = \frac{\tau}{\sqrt{\pi}} = \frac{1}{\sqrt{\pi}} \sqrt{\frac{a^L((a-1)\cdot L-(a-1))}{2((a-1)\cdot L+1)^2}}, \tag{22}$$

where $\tau'^2 = 2\tau^2$.

So, for $a = 2$, as in our biallelic landscapes,

$$\lim_{L\to\infty} \langle |p_{\mathrm{f}} - p_{\mathrm{gp}}| \rangle = \sqrt{\frac{2^L \cdot (L-1)}{2\pi(L+1)^2}}. \tag{23}$$

Proofs

Proof 1: Selection for w_{opt} cannot transform a simple sign epistasis motif into a reciprocal sign epistasis motif

In order to have a reciprocal sign epistasis motif:

- No two adjacent arrows can point in the same direction.
- No two parallel arrows can point in the same direction.

Because sign epistasis motifs cannot be cyclic, the two conditions above imply one another. The simple sign epistasis motif determines the complete rank ordering of the phenotypic values, unlike the other two motifs, which only give a partial ordering. Therefore, upon selecting for any w_{opt}, at least two arrows in the genotype-phenotype landscape remain the same as in the fitness landscape or reverse in direction together. This results in a pattern of arrows in the fitness landscape that always contradicts both the above mentioned conditions for reciprocal sign epistasis (see Supplementary Fig. 1). ■

Proof 2: The phenotype-fitness map changes a reciprocal sign epistasis motif into a no sign epistasis motif or a simple sign epistasis motif with equal probability

The reciprocal sign epistasis motif does not determine the complete rank ordering of the four phenotypic values. In particular, it does not determine the ordering of the two peaks and the two valleys. Let the phenotypic value of the higher of the two peaks be w_{p1} and that of the lower peak be w_{p2}. Similarly, let the labels w_{v1} and w_{v2} refer to the higher and lower valleys, respectively. By design, these values lie on the real number line between 0 and 1. We define the "neighbourhood" of each of the phenotypic values as the region on the number line around the phenotypic value that is closer to it than to any other phenotypic value (Supplementary Fig. 2). For transformation of the reciprocal sign epistasis motif into one of the other two motifs, w_{opt} must lie in the neighbourhood of either w_{p2} or w_{v1} and it must be second closest to the other (i.e., $F_{\text{p2}} > F_{\text{v1}} > \text{rest}$ or $F_{\text{v1}} > F_{\text{p2}} > \text{rest}$, where F_i represents the fitness corresponding to the phenotype w_i). If not, the reciprocal sign epistasis motif will be retained. So, for any w_{opt}, we only consider those configurations that satisfy this condition. Amongst these configurations, both phenotypic values are equally likely to be the closest to w_{opt}. Let's call the probability of either event happening $P_0 = P(F_{\text{v1}} > F_{\text{p2}} > \text{rest}) = P(F_{\text{p2}} > F_{\text{v1}} > \text{rest})$. We denote the transition probabilities amongst motifs as $P(x|rs)$, where x represents either the no sign epistasis motif, ns, or the simple sign epistasis motif, ss, and rs represents the reciprocal sign epistasis motif. Now, for the no sign epistasis motif to emerge, the fitness values of the peaks or valleys need to be "separated", while for the simple sign epistasis motif to emerge, they need to be "interspersed" (see Supplementary Fig. 2), giving us the following:

$$P(ns|rs) = P(F_{\text{v1}} > F_{\text{p2}} > \text{rest}) \cdot P(F_{\text{p1}} > F_{\text{v2}}) + P(F_{\text{p2}} > F_{\text{v1}} > \text{rest}) \cdot P(F_{\text{p1}} < F_{\text{v2}}), \tag{24}$$

and

$$P(ss|rs) = P(F_{\text{v1}} > F_{\text{p2}} > \text{rest}) \cdot P(F_{\text{p1}} < F_{\text{v2}}) + P(F_{\text{p2}} > F_{\text{v1}} > \text{rest}) \cdot P(F_{\text{p1}} > F_{\text{v2}}). \tag{25}$$

Finally, using $P_0 = P(F_{\text{v1}} > F_{\text{p2}} > \text{rest}) = P(F_{\text{p2}} > F_{\text{v1}} > \text{rest})$ gives:

$$P(ns|rs) = P_0 \cdot P(F_{\text{p1}} > F_{\text{v2}}) + P_0 \cdot P(F_{\text{p1}} < F_{\text{v2}}) \text{ and } P(ss|rs) = P_0 \cdot P(F_{\text{p1}} < F_{\text{v2}}) + P_0 \cdot P(F_{\text{p1}} > F_{\text{v2}}) \implies P(ns|rs) = P(ss|rs). \blacksquare \tag{26}$$

Proof 3: For House-of-Cards genotype-phenotype landscapes, selection for low or intermediate phenotypic values is equally likely to increase or decrease the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape

For any number of loci L, the probability that selection for a low or intermediate phenotypic value causes an increase in the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape, is

$$P_{\text{inc}} = \Sigma_{i>j} P(p_{\text{f}} = i | p_{\text{gp}} = j) \cdot P(p_{\text{gp}} = j), \tag{27}$$

whereas the probability of a decrease in the number of peaks is

$$P_{\text{dec}} = \Sigma_{i>j} P(p_{\text{f}} = j | p_{\text{gp}} = i) \cdot P(p_{\text{gp}} = i), \tag{28}$$

where $i, j \in \{1, 2, ...2^{L-1}\}$, $P(p_{\text{f}} = j | p_{\text{gp}} = i)$ is the probability of having j peaks in the fitness landscape after selection for w_{opt} given the genotype-phenotype landscape has i peaks, and $P(p_{\text{gp}} = i)$ is the probability of i peaks in the genotype-phenotype landscape. The number of peaks in the fitness landscape, p_{f}, is independent of the number of peaks in the genotype-phenotype landscape, p_{gp}, because p_{f} depends upon the number of genotypes with phenotypes close to the optimal phenotypic value, and this number is independent of the number of genotypes with phenotypes close to 1. This implies:

$$P(p_{\text{f}} = i | p_{\text{gp}} = j) \cdot P(p_{\text{gp}} = j) \; = \; P(p_{\text{f}} = i) \cdot P(p_{\text{gp}} = j), \tag{29}$$

and

$$P(p_{\text{f}} = j | p_{\text{gp}} = i) \cdot P(p_{\text{gp}} = i) \; = \; P(p_{\text{f}} = j) \cdot P(p_{\text{gp}} = i). \tag{30}$$

Moreover, p_{gp} and p_{f} are identically distributed; i.e., $P(p_{\text{gp}} = i) = P(p_{\text{f}} = i)$

$$\implies P_{\text{inc}} = P_{\text{dec}} \blacksquare \tag{31}$$

Notes

Note 1: Number of maxima in the plot of $\langle |p_{\mathbf{f}} - p_{\mathbf{gp}}| \rangle$ versus $w_{\mathbf{opt}}$ for additive biallelic landscapes

Any biallelic, additive landscape can be relabelled such that the genotype with the lowest phenotypic value is represented by 000.....0 and the genotype with the highest phenotypic value by 111.....1. Moreover, the phenotypic values can be re-scaled such that genotype 000.....0 has phenotypic value zero and genotype 111.....1 has phenotypic value one. This yields a monotonically increasing genotype-phenotype landscape, wherein the phenotypic effects of every single mutant is larger than zero and they add up to 1; i.e.,

$$w_{100...0} + w_{010...0} + ... + w_{000...1} = 1 \tag{32}$$

where $w_{100...0}$ is the phenotypic value of a single mutant with the mutation at the 1st site, and so on. Equation 32 implies that the vector of single mutant phenotypes belongs to a standard $L-1$-simplex[‡].

Two-locus case

A two-locus biallelic fitness landscape can have a maximum of two peaks. Because the additive genotype-phenotype landscape has only one peak, selection for a low or intermediate phenotypic value can increase the number of peaks in the fitness landscape, relative to the genotype-phenotype landscape, by at most one. Let the fitness of one of the single mutants be x, forcing the fitness of the other single-mutant to be $1-x$. Due to the relabelling of the genotype-phenotype landscape, a two-peaked fitness landscape is only possible when the two single-mutants are peaks. This will occur when the two single-mutants have phenotypes that are closer to w_{opt} than the phenotypes of both genotypes 00 and 11. This leads to the following conditions on x:

$$1 - 2 \cdot w_{\text{opt}} < x < 2 \cdot w_{\text{opt}}, \forall w_{\text{opt}} < 0.5 \tag{33}$$

and

$$2 \cdot w_{\text{opt}} - 1 < x < 2 - 2 \cdot w_{\text{opt}}, \forall w_{\text{opt}} > 0.5. \tag{34}$$

From Eqs. 33 and 34, we can infer the shape of the $L = 2$ curves in Figures 4A,B. Specifically, for $w_{\text{opt}} < 0.25$ and $w_{\text{opt}} > 0.75$, we get the absurd condition $0.5 < x < 0.5$, which implies there is no possible value of x that will lead to two peaks. Further, for $w_{\text{opt}} = 0.5$, $0 < x < 1$, which implies there will always be two peaks. For $0.25 \leq w_{\text{opt}} < 0.5$ or $0.5 < w_{\text{opt}} \leq 0.75$, the fitness landscape can have either one or two peaks.

Three-locus case

[‡] An n dimensional simplex is the convex hull of $n+1$ affinely independent points. E.g., a 0-simplex is a point, a 1-simplex is a line, a 2-simplex is a triangle, and a 3-simplex is a tetrahedron.

Here, the single-mutant phenotypic values belong to the 2-simplex formed by the points (0,0,1), (0,1,0) and (1,0,0). Let (w_1, w_2, w_3) be an arbitrary point in the simplex, where w_i is the phenotypic value of the single mutant with a mutation at the ith site and $i \in \{1,2,3\}$. Although, a three-locus, biallelic landscape can have at most four peaks, the transformed additive genotype-phenotype landscapes can have at most three, due to the correlations between the phenotypic values. So for instance, if the three single mutants become peaks, then the triple mutant cannot be a peak because the double mutants will be assigned a higher fitness than the triple mutant. This is caused by correlations in the additive landscape, which makes the double-mutants closer in phenotypic value to the single-mutants than the triple mutant. Thus, it is easy to see that the only three peak configuration possible is one in which either all the single mutants become peaks or all the double mutants become peaks. For the former to be the case, each single mutant phenotypic value w_i must satisfy the following conditions:

$$(w_i - w_{\text{opt}})^2 < (0 - w_{\text{opt}})^2 \tag{35}$$

and

$$(w_i - w_{\text{opt}})^2 < (1 - w_j - w_{\text{opt}})^2 \ \forall j \neq i. \tag{36}$$

While Eq. 35 necessitates the single mutant fitness values to be higher than the wild type fitness, Eq. 36 requires them to be higher than the neighbouring double mutant fitness values. On average, $w_i = 1/3 \ \forall i$, so on average, the simplex area enclosed by the above conditions will be maximised when $w_{\text{opt}} = 1/3$. A similar analysis can be done for the case when the three double mutants become peaks and in this case, the area is maximised when $w_{\text{opt}} = 2/3$. This explains the presence of the two maxima in Fig. 4A at $w_{\text{opt}} = 1/3$ and $w_{\text{opt}} = 2/3$. Further, the dip at $w_{\text{opt}} = 0.5$ can be explained by the following argument: Assuming the single mutant phenotypic values are of the type $(x, y, 1-x-y)$, the double mutant phenotypic value will be of the type $(x+y, 1-x, 1-y)$. At $w_{\text{opt}} = 0.5$, the wild type and triple mutant immediately become minima and leave an opportunity for the intermediate mutants to become peaks. However, if one of the single mutants (say with fitness x) is closest in fitness to 0.5, the other two cannot be peaks because the double mutant formed by combining these two mutations will have the same fitness i.e., $|x - 0.5| = |(1-x) - 0.5|$. So at $w_{\text{opt}} = 0.5$, there are two peaks after selection for a low phenotypic value and therefore the net change is 1, which is smaller than what we see at $w_{\text{opt}} = 1/3, 2/3$.

Four and more locus cases

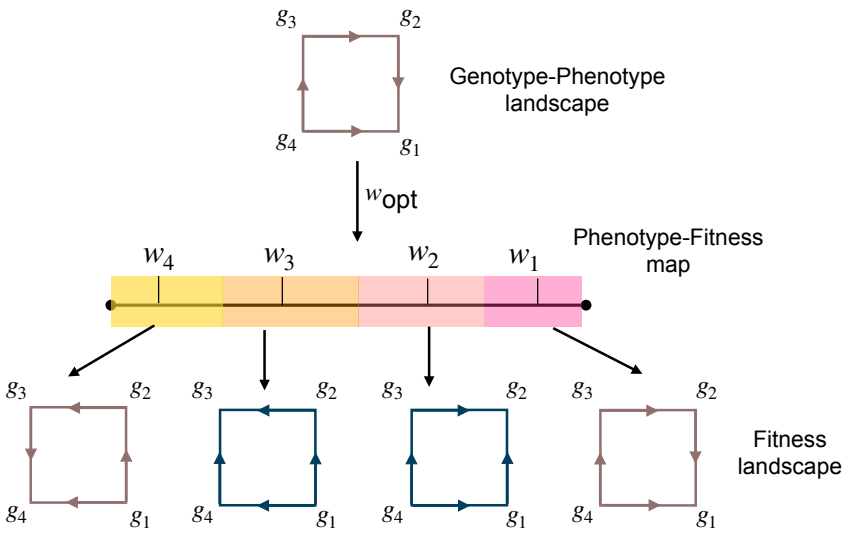
We could explicitly derive the shape of the curves for the $L = 2$ and $L = 3$ cases because it was easy to visualise the simplices. For $L \geq 4$, we can form an intuition about the expected result, by considering the average additive fitness landscape, which guides the behavior of the average change in the number of peaks. Because the only randomness in these landscapes arises from the effect of single mutations, which as described above, belong to the L-dimensional simplex, all we need to look at is the average L-dimensional vector of single mutation effects $\langle \vec{w}_1^L \rangle$ and the deviations around it. These random vectors belong to an $L-$dimensional Dirichlet distribution with all the concentration parameters equal to one. It can be shown by symmetry

that $\langle \vec{w}_1^L \rangle = (1/L, 1/L, , , 1/L)$. Therefore, on average, the fitness of a mutant with n mutations is n/L. As w_{opt} increases from zero, the number of peaks will gradually start to increase as w_{opt} gets closer the value of the fitness of single mutants. Then as w_{opt} approaches $1/L$, we encounter the first maxima because at this w_{opt}, all single mutants get to be peaks. Next, there is a slight decline due to competition between the single and double mutants as peaks, followed by the emergence of the second maxima at $2/L$ when all the double mutants become peaks, and so on. Naturally, because the number of genotypes with n mutations increases as n reaches the value $L/2$, the height of the maxima increases accordingly. Thereafter, we observe a fall in $\langle |p_{\text{f}} - p_{\text{gp}}| \rangle$. Therefore, for a genotype of length L, there are $L-1$ maxima in the plot of $\langle |p_{\text{f}} - p_{\text{gp}}| \rangle$ vs. w_{opt}. However, the maxima are close together and therefore difficult to visually discern.

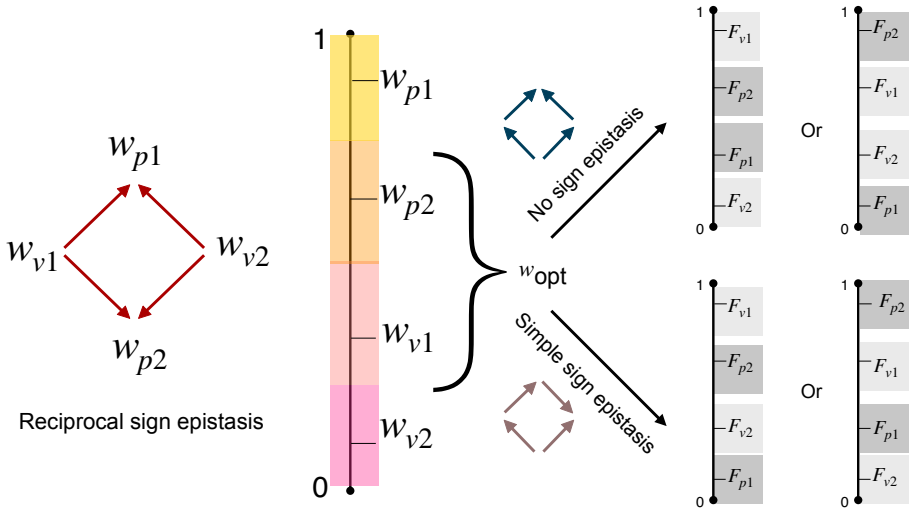
Note 2: Evolution on NK landscapes

Supplementary Fig. 3A,C shows the average length of a greedy adaptive walk $\langle l \rangle$ in NK landscapes with $L = 5$ and $L = 8$. For all $K < L-1$, $\langle l \rangle$ is minimised when $w_{\text{opt}} = 0.5$, whereas for $K = L-1$ (House-of-Cards landscapes), $\langle l \rangle$ is independent of w_{opt}. This can be understood by first considering additive genotype-phenotype landscapes ($K = 0$), wherein genotypes with a phenotype $w \approx 0.5$ are by definition, approximately located half-way between the wild-type and its antipodal sequence. Selecting for $w_{\text{opt}} = 0.5$ thus minimizes the average length of a greedy adaptive walk to these peak genotypes. As K increases, correlations among the phenotypes of mutationally-neighboring genotypes are gradually broken and the landscape becomes more rugged. Yet, genotypes with a phenotype $w \approx 0.5$ tend to remain near the center of the landscape. When K is increased to $L-1$, the phenotypes of mutationally-neighboring genotypes become completely uncorrelated and monotonic increases in phenotype along mutational paths to peak genotypes become exceedingly rare, rendering $\langle l \rangle$ independent of w_{opt}. Thus the "U"-shape of the curve of $\langle l \rangle$ versus w_{opt} flattens as K increases from 0 to $L-1$. Interestingly, so long as $K > 0$, the average height of the peak reached by a greedy adaptive walk is also maximized at $w_{\text{opt}} = 0.5$, and these local peaks are >97% the height of the global peak for $\sigma = 1$ (Supplementary Fig. 7), an observation we explain below.

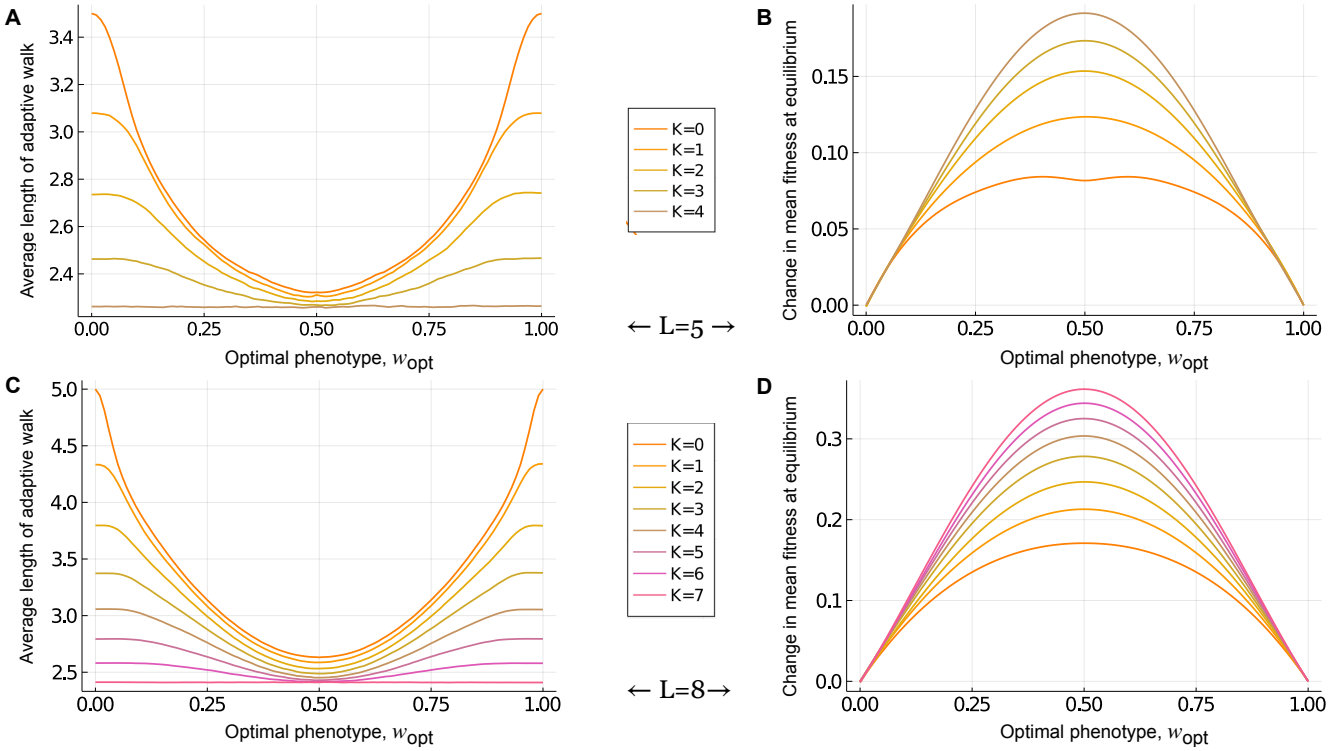
We next simulated deterministic mutation-selection dynamics. Supplementary Fig. 3B,D show the change in mean fitness at equilibrium when selecting for w_{opt}, relative to fitness at equilibrium when selecting for $w_{\text{opt}} = 1$, for $\sigma = 0.5$ and $\mu = 0.1$ (see Supplementary Fig. 8 for other values of σ and μ). This change is always positive, increases with L and K, and tends to be maximized at $w_{\text{opt}} = 0.5$ when $K > 0$. The latter occurs because the optimal phenotype $w_{\text{opt}} = 0.5$ is, on average, closer to all other phenotypic values in the landscape. For instance for $K = L-1$, the sum $\Sigma_{i=1}^{2^L}(w_i - w_{\text{opt}})^2$ is minimized when $w_{\text{opt}} = \Sigma_{i=1}^{2^L} w_i / 2^L = \langle w \rangle = 0.5$. Thus, selecting for $w_{\text{opt}} = 0.5$ creates a fitness landscape with higher average fitness than fitness landscapes under selection for any other w_{opt}. This also explains why the average height of local adaptive peaks is maximized at $w_{\text{opt}} = 0.5$ (Supplementary Fig. 7). In sum, these results show that selection for an intermediate phenotypic value decreases the length of a greedy adaptive walk and increases mean fitness at equilibrium, despite increasing the overall ruggedness of the fitness landscape, relative to the genotype-phenotype landscape.



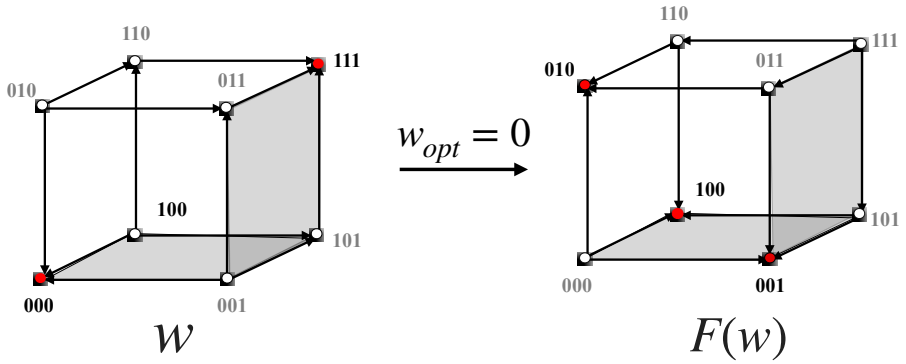
Supplementary Figure 1. A simple sign epistasis motif (shown in brown) cannot be changed into a reciprocal sign epistasis motif, for any w_{opt}. The no sign epistasis motif is shown in blue. The remaining colours represent the neighbourhood of the phenotypic values (w_i) corresponding to each genotype (g_i), where $i \in \{1,2,3,4\}$ and the genotypes are labelled in descending order, such that the genotype with the highest phenotypic value is labelled 1, and so on. The four bottom arrows point to the transformed motifs when w_{opt} belongs to the neighbourhood from which the arrow emanates.



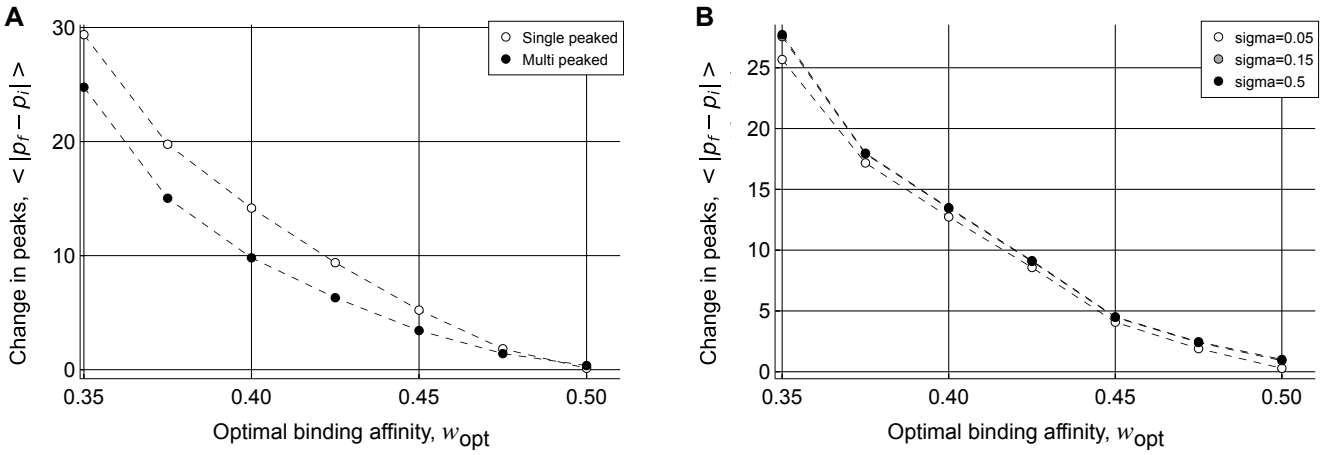
Supplementary Figure 2. Selection for w_{opt} transforms the reciprocal sign epistasis motif (shown in red) into the no sign epistasis motif (shown in blue) and the simple sign epistasis motif (shown in brown) with equal probability. The neighbourhood of each phenotypic value is shown in a different colour. For the no sign epistasis motif to emerge (top), the fitness values need to be "separated", while for the simple sign epistasis motif to emerge (bottom), the fitness values need to be "interspersed".



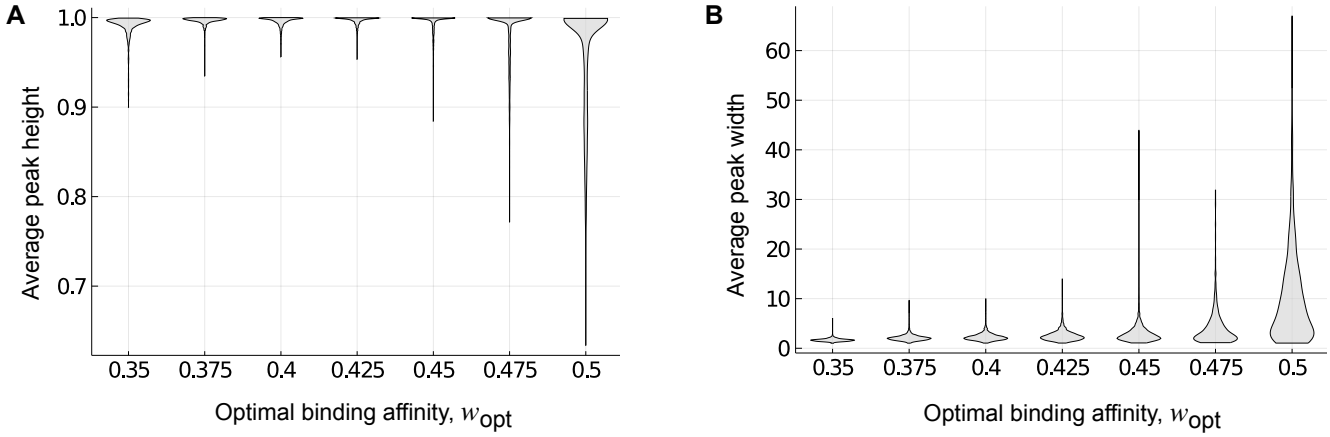
Supplementary Figure 3. Evolutionary consequences of landscape ruggedness in NK landscapes. (A,C) The average length of a greedy adaptive walk and (B,D) the change in mean fitness at equilibrium, relative to fitness at equilibrium when selecting for $w_{\text{opt}} = 1$, are shown for (A,B) $L = 5$ and (C,D) $L = 8$. In (B,D), $\mu = 0.1$ and $\sigma = 0.5$.



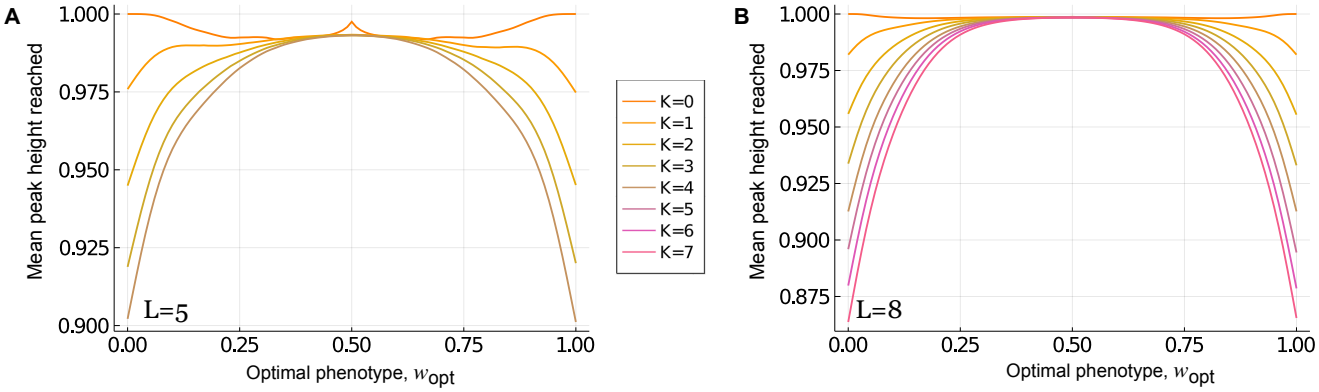
Supplementary Figure 4. Selection for $w_{\text{opt}} = 0$ does not change the type of epistasis motif for any "square" in the fitness landscape, relative to the genotype-phenotype landscape, yet it can change the number and location of peaks. To understand how, consider that any two adjacent faces of the hypercube (e.g., gray faces above) are sufficient to determine whether the genotypes on their common edge are peaks or not. After selecting for $w_{\text{opt}} = 0$, the type of epistasis motif does not change in the adjacent faces, yet the number and location of the peaks on their common edge does change. Peak genotypes are shown in red. Arrows point from lower to higher phenotypic or fitness values.



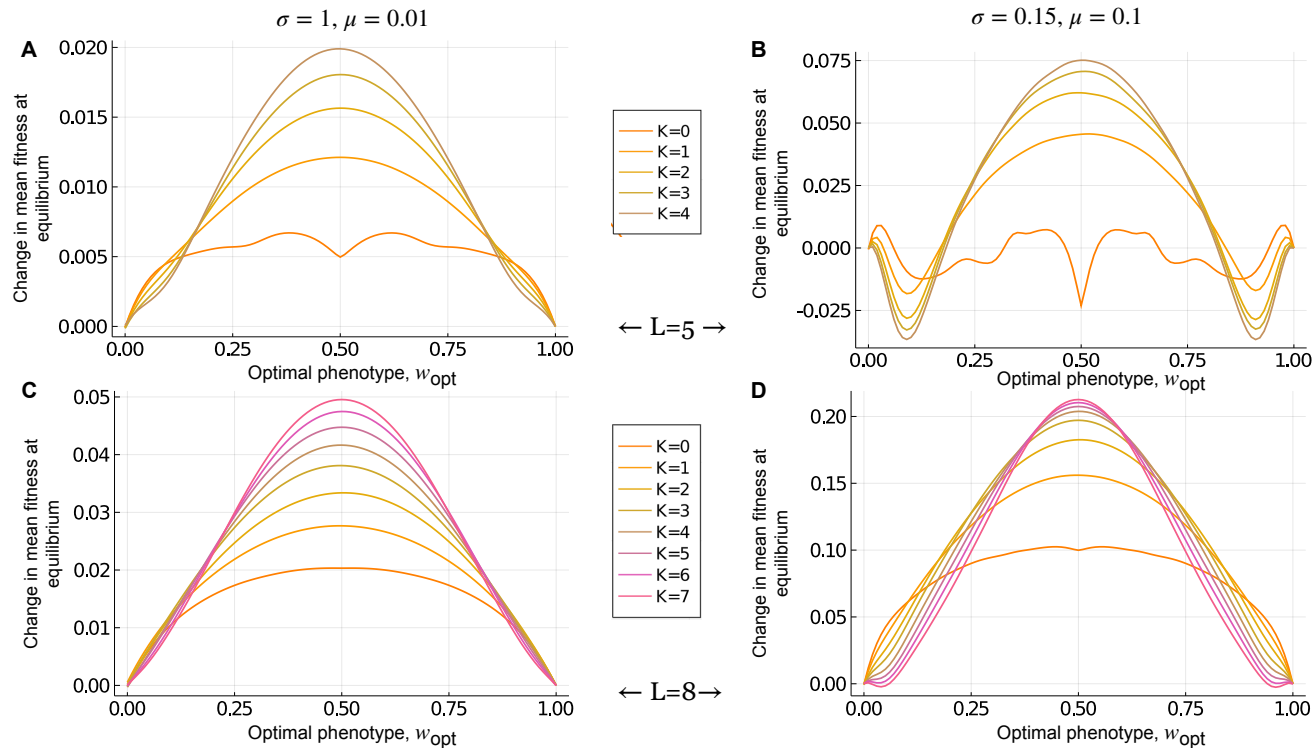
Supplementary Figure 5. Mean absolute change in the number of peaks in the fitness landscape relative to the genotype-phenotype landscape, shown in relation to w_{opt}, for (A) single-peaked vs. multi-peaked genotype-phenotype landscapes and (B) three values of σ.



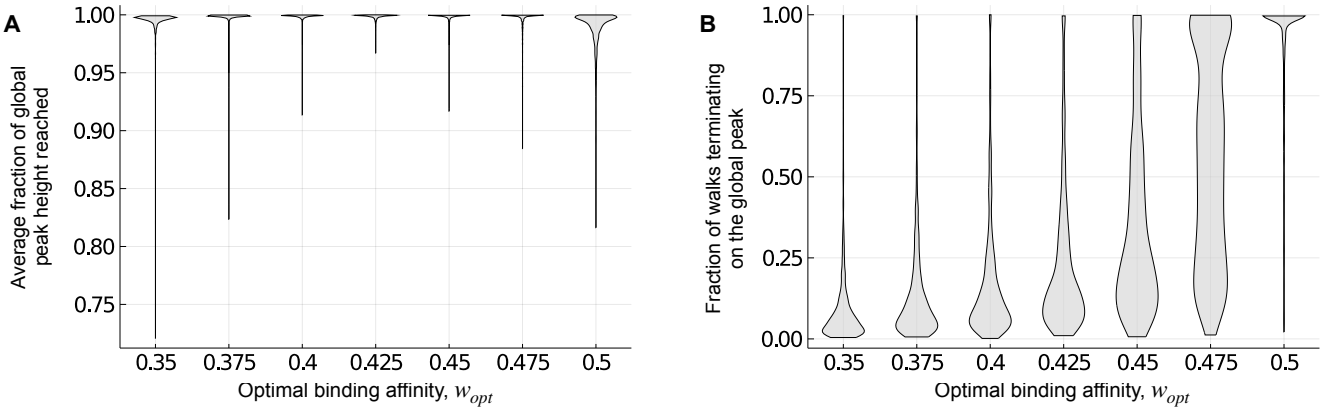
Supplementary Figure 6. Average peak (A) height and (B) width for 1,137 empirical landscapes, shown in relation to the optimal binding affinity w_{opt}. Violin plots show the distribution across the landscapes for each w_{opt}. Data include both local and global peaks.



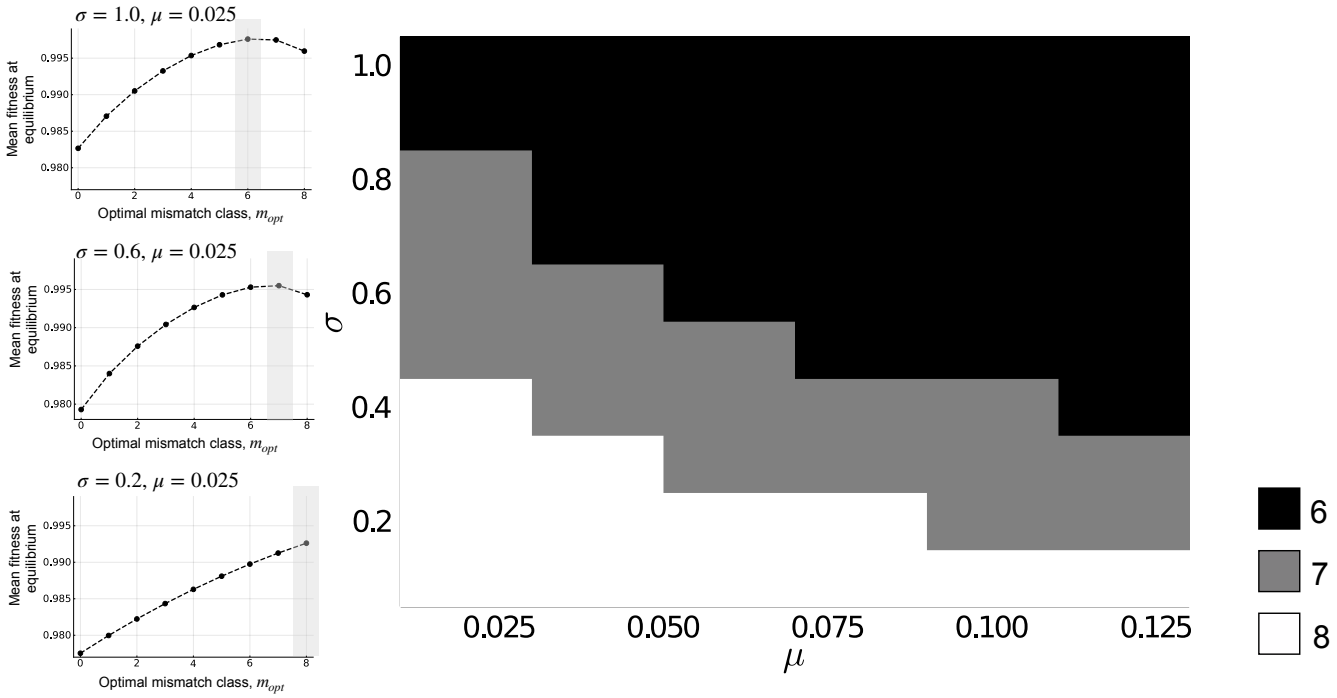
Supplementary Figure 7. The average height of peaks reached by greedy adaptive walks in NK landscapes with $\sigma = 0.5$, shown in relation to w_{opt}, for (A) $L = 5$ and (B) $L = 8$, with $K = 0 \ldots L-1$.



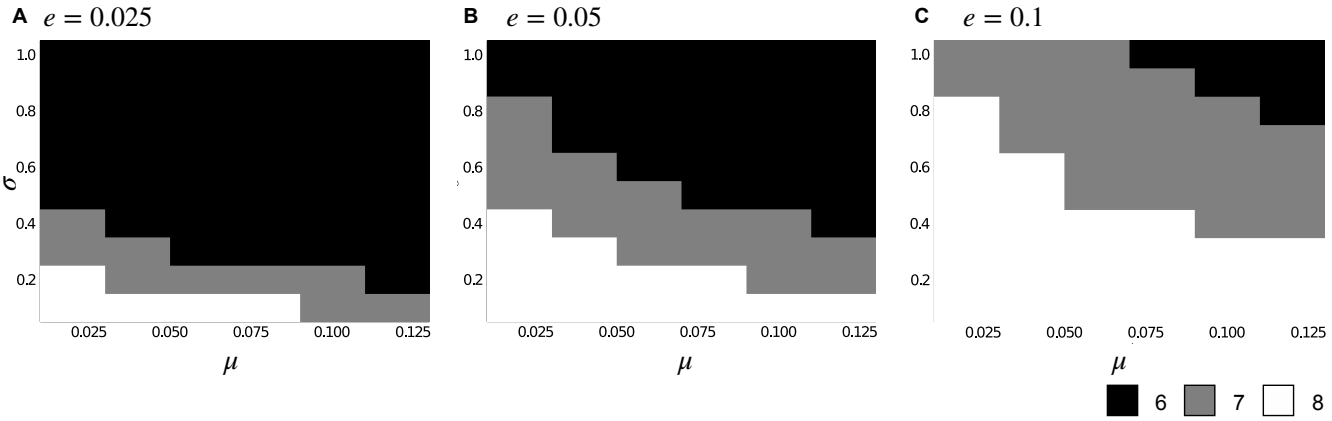
Supplementary Figure 8. Change in mean fitness at equilibrium for NK landscapes with (A,B) L=5 and (C,D) L=8, shown in relation to w_{opt}, for different values of σ and μ.



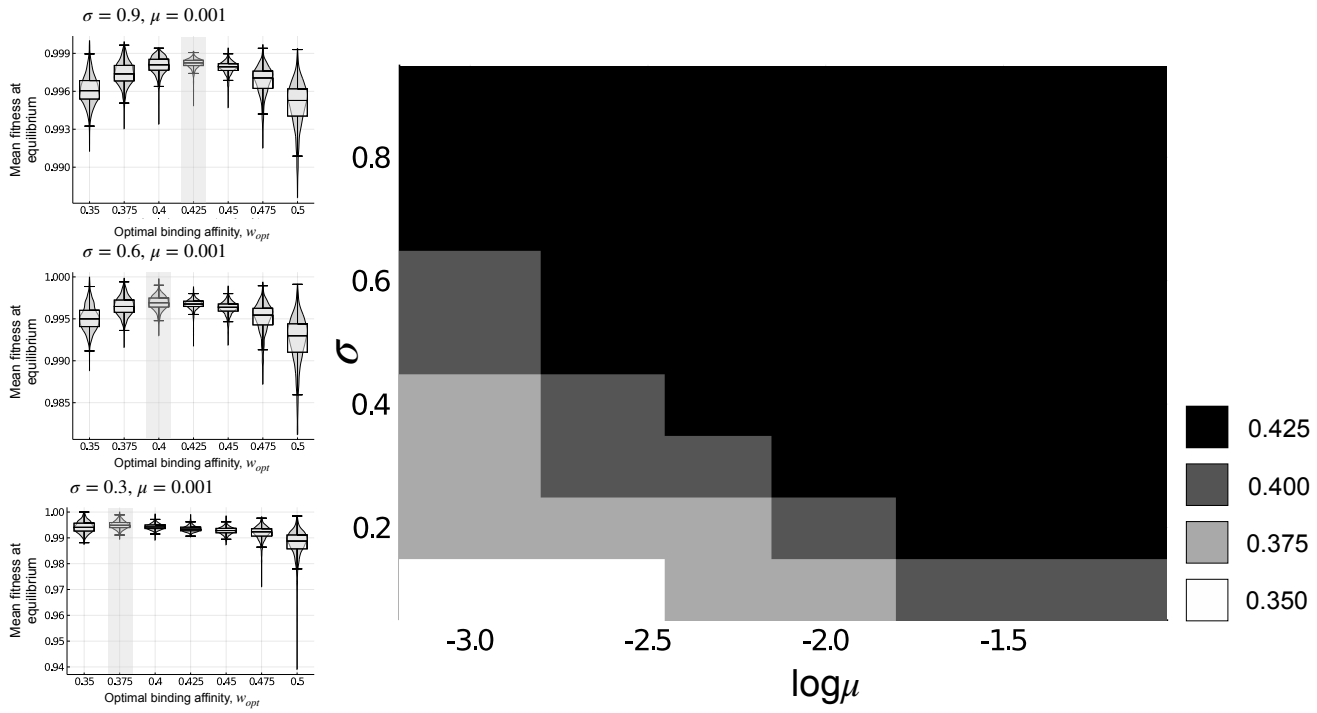
Supplementary Figure 9. (A) Local peaks on which the adaptive walks terminate tend to be nearly as tall as the global peaks in the 1,137 empirical landscapes. Violin plots show the distribution of the fractional height of local peaks reached by greedy adaptive walks, relative to the height of the global peak, for each optimal binding affinity w_{opt}. (B) Violin plots show the distribution of the fraction of walks terminating on the global peak, for each optimal binding affinity w_{opt}.



Supplementary Figure 10. The mistmatch class m_{opt} that maximizes mean fitness at equilibrium is shown in relation to the strength of selection σ and the mutation rate μ. The three smaller panels show mean fitness at equilibrium in relation to m_{opt} for three combinations of σ and μ. The value of m_{opt} that maximizes mean fitness is indicated with a gray rectangle.



Supplementary Figure 11. The mistmatch class m_{opt} that maximizes mean fitness at equilibrium is shown in relation to the strength of selection σ and the mutation rate μ for three different values of the mismatch penalty e: (A) $e = 0.025$, (B) $e = 0.05$ and (C) $e = 0.1$.



Supplementary Figure 12. The binding affinity w_{opt} that maximizes mean fitness at equilibrium is shown in relation to the strength of selection σ and the logarithm of the mutation rate ($\log \mu$) for the 1,137 empirical landscapes. The three smaller panels show the distributions of mean fitness at equilibrium as violin plots, in relation to w_{opt}, for three combinations of σ and μ. Box-and-whisker plots show the 25-75% quartiles. The value of w_{opt} that maximizes mean fitness is indicated. These results are robust to perturbations of the fitness values (Methods), with the exception of two parameter combinations – $\sigma = 0.1, \mu = 0.001$ and $\sigma = 0.1, \mu = 0.0025$. For these parameter combinations, the binding affinity w_{opt} that maximizes mean fitness at equilibrium changes from 0.35 to 0.385 (on average) and 0.375 respectively.