Genotypic complexity of Fisher's geometric model

Sungmin Hwang[*]**, Su-Chan Park**[†,1] **and Joachim Krug**[*]

[*]Institut für Theoretische Physik, Universität zu Köln, 50937 Köln, Germany, [†]Department of Physics, The Catholic University of Korea, Bucheon 14662, Republic of Korea

ABSTRACT Fisher's geometric model was originally introduced to argue that complex adaptations must occur in small steps because of pleiotropic constraints. When supplemented with the assumption of additivity of mutational effects on phenotypic traits, it provides a simple mechanism for the emergence of genotypic epistasis from the nonlinear mapping of phenotypes to fitness. Of particular interest is the occurrence of reciprocal sign epistasis, which is a necessary condition for multipeaked genotypic fitness landscapes. Here we compute the probability that a pair of randomly chosen mutations interacts sign-epistatically, which is found to decrease with increasing phenotypic dimension n, and varies non-monotonically with the distance from the phenotypic optimum. We then derive expressions for the mean number of fitness maxima in genotypic landscapes composed of all combinations of L random mutations. This number increases exponentially with L, and the corresponding growth rate is used as a measure of the complexity of the landscape. The dependence of the complexity on the model parameters is found to be surprisingly rich, and three distinct phases characterized by different landscape structures are identified. Our analysis shows that the phenotypic dimension, which is often referred to as phenotypic complexity, does not generally correlate with the complexity of fitness landscapes and that even organisms with a single phenotypic trait can have complex landscapes. Our results further inform the interpretation of experiments where the parameters of Fisher's model have been inferred from data, and help to elucidate which features of empirical fitness landscapes can be described by this model.

KEYWORDS fitness landscape; genotype-phenotype map; epistasis; adaptation; fitness peaks

A fundamental question in the theory of evolutionary adaptation concerns the distribution of mutational effect sizes and the relative roles of mutations of small versus large effects in the adaptive process (Orr 2005). In his seminal 1930 monograph, Ronald Fisher devised a simple geometric model of adaptation in which an organism is described by n phenotypic traits and mutations are random displacements in the trait space (Fisher 1930). Each trait has a unique optimal value and the combination of these values defines a single phenotypic fitness optimum that constitutes the target of adaptation. Because random mutations act pleiotropically on multiple traits, the probability that a given mutation brings the phenotype closer to the target decreases with increasing n. Fisher's analysis showed that, for large n, the mutational step size in units of the distance to the optimum must be smaller than $1/\sqrt{n}$ in order for the mutation to be beneficial with an appreciable probability. He thus concluded that the evolution of complex adaptations involving a large number of traits must rely on mutations of small effect. This conclusion was subsequently qualified by the realization that small effect mutations are likely to be lost by genetic drift, and therefore mutations of intermediate size contribute most effectively to adaptation (Kimura 1983; Orr 1998, 2000).

During the past decade Fisher's geometric model (FGM) has become a standard reference point for theoretical and experimental work on fundamental aspects of evolutionary adaptation (Tenaillon 2014). In particular, it has been found that FGM provides a versatile and conceptually simple mechanism for the emergence of epistatic interactions between genetic mutations in their effect on fitness (Martin *et al.* 2007; Gros *et al.* 2009; Blanquart *et al.* 2014). For this purpose two extensions of Fisher's original formulation of the model have been suggested. First, phenotypes are assigned an explicit fitness value, which is usually taken to be a smooth function on the trait

space with a single maximum at the optimal phenotype. Second, and more importantly, mutational effects on the phenotypes are assumed to be additive. As a consequence, any deviations from additivity that arise on the level of fitness are solely due to the nonlinear mapping from phenotype to fitness, or, in mathematical terms, due to the curvature of the fitness function. Because the curvature is largest around the phenotypic optimum, epistasis generally increases upon approaching the optimal phenotype and is weak far away from the optimum. Several recent studies have made use of the framework of FGM to interpret experimental results on pairwise epistatic interactions and to estimate the parameters of the model from data (Martin *et al.* 2007; Velenich and Gore 2013; Weinreich and Knies 2013; Perfeito *et al.* 2014; Schoustra *et al.* 2016).

A particularly important form of epistatic interaction is sign epistasis, where a given mutation is beneficial or deleterious depending on the genetic background (Weinreich *et al.* 2005). Two types of sign epistasis are distinguished depending on whether one of the mutations affects the effect sign of the other but the reverse is not true (*simple sign epistasis*), or whether the interaction is reciprocal (*reciprocal sign epistasis*); for a pictorial representation of the two kinds of sign epistasis, see, for example, Poelwijk *et al.* (2007). Sign epistasis can arise in FGM either between large effect beneficial mutations that in combination overshoot the fitness optimum, or between mutations of small fitness effect that display antagonistic pleiotropy (Blanquart *et al.* 2014). The presence of sign epistasis is a defining feature of genotypic fitness landscapes that are complex, in the sense that not all mutational pathways are accessible through simple hill-climbing and multiple genotypic fitness peaks may exist (Weinreich *et al.* 2005; Franke *et al.* 2011; de Visser and Krug 2014). Specifically, reciprocal sign epistasis is a necessary condition for the existence of multiple fitness peaks (Poelwijk *et al.* 2011; Crona *et al.* 2013).

Following a common practice, here a genotypic fitness landscape is understood to consist in the assignment of fitness values to all combinations of L haploid, biallelic loci that together constitute the L-dimensional genotype space. A peak in such a landscape is a genotype that has higher fitness than all its L neighbors that can be reached by a single point mutation (Kauffman and Levin 1987). Note that, in contrast to the continuous phenotypic space on which FGM is defined, the space of genotypes is discrete.

Blanquart *et al.* (2014) showed that an ensemble of L-dimensional genotypic landscapes can be constructed from FGM by combining subsets of L randomly chosen mutational displacements. Each sample of L mutations defines another realization of the landscape ensemble, and the exploratory simulations reported by Blanquart *et al.* (2014) indicate a large variability among the realized landscapes. Nevertheless some general trends in the properties of the genotypic landscapes were identified. In particular, as expected on the basis of the considerations outlined above, the genotypic landscapes are essentially additive when the focal phenotype representing the unmutated wild type is far away from the optimum and become increasingly rugged as the optimal phenotype is approached.

In this article we present a detailed and largely analytic study of the properties of genotypic landscapes generated under FGM. The focus is on two types of measures of landscape complexity, that is, the fraction of sign-epistatic pairs of random mutations and the number of fitness maxima in the genotypic landscape. A central motivation for our investigation is to assess the potential of FGM and related phenotypic models to explain the properties of empirical genotypic fitness landscapes of the kind that have been recently reported in the literature (Szendro *et al.* 2013; Weinreich *et al.* 2013; de Visser and Krug 2014). The ability of nonlinear phenotype-fitness maps to explain epistatic interactions among multiple loci has been demonstrated for a virus (Rokyta *et al.* 2011) and for an antibiotic resistance enzyme (Schenk *et al.* 2013), but a comparative study of several different data sets using Approximate Bayesian Computation has questioned the broader applicability of phenotype-based models (Blanquart and Bataillon 2016). It is thus important to develop a better understanding of the structure of genotypic landscapes generated by phenotypic models such as FGM.

In the next section we describe the mathematical setting and introduce the relevant model parameters: the phenotypic and genotypic dimensionalities n and L, the distance of the focal phenotype to the optimum, and the standard deviation of mutational displacements. As in previous studies of FGM, specific scaling relations among these parameters have to be imposed in order to arrive at meaningful results for large n and L. We then present analytic results for the probability of sign epistasis and the behavior of the number of fitness maxima for large L, both in the case of fixed phenotypic dimension n and for a situation where the joint limit $n, L \to \infty$ is taken at constant ratio $\alpha = n/L$.

Similar to other probabilistic models of genotypic fitness landscapes (Kauffman and Levin 1987; Weinberger 1991; Evans and Steinsaltz 2002; Durrett and Limic 2003; Limic and Pemantle 2004; Neidhart *et al.* 2014), the number of maxima generally increases exponentially with L, and we use the exponential growth rate as a measure of genotypic complexity. We find that this quantity displays several phase transitions as a function of the parameters of FGM which separate parameter regimes characterized by qualitatively different landscape structures. Depending on the regime, the genotypic landscapes induced by FGM become more or less rugged with increasing phenotypic dimension. This indicates that the role of the number of phenotypic traits in shaping the fitness landscapes of FGM is much more subtle than has been previously appreciated, and that the sweeping designation of n as (phenotypic) "complexity" can be misleading. Further implications of our study for the theory of adaptation and the interpretation of empirical data will be elaborated in **Discussion**.

Model

Basic properties of FGM

In FGM, the phenotype of an organism is modeled as a set of n real-valued traits and represented by a vector $\vec{y} = (y_1, y_2, \ldots, y_n)$ in the n dimensional Cartesian space, $\vec{y} \in \mathbb{R}^n$. The fitness $W(\vec{y})$ is assumed to be a smooth, single-peaked function of the phenotype $\vec{y}$. By choosing an appropriate coordinate system, the optimum phenotype, i.e., the combination of phenotypic traits with the highest fitness value, can be placed at the origin in $\mathbb{R}^n$. We also assume that the fitness $W(\vec{y})$ depends on the distance to the optimum $|\vec{y}|$ but not on the direction of $\vec{y}$, which can be justified by arguments based on random matrix theory (Martin 2014). The uniqueness of the phenotypic optimum at the origin implies that $W(\vec{y})$ is a decreasing function of $|\vec{y}|$. The form of the fitness function will be specified

below when needed. Most of the results presented in this paper are however independent of the explicit shape of $W(\vec{y})$, as they rely solely on the relative ordering of different genotypes with respect to their fitness.

When a mutation arises the phenotype of the mutant becomes $\vec{y} + \vec{\xi}$, where $\vec{y}$ is the parental phenotype and the mutational vector $\vec{\xi}$ corresponds to the change of traits due to the mutation. The key result derived by Fisher (1930) concerns the fraction P_b of beneficial mutations arising from a wild type phenotype located at distance d from the optimum. Assuming that mutational displacements have a fixed length $|\vec{\xi}| = r$ and random directions, he showed that for $n \gg 1$

$$P_b = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt = \frac{1}{2}\mathrm{erfc}(x/\sqrt{2}), \qquad (1)$$

where erfc denotes the complementary error function and $x = r\sqrt{n}/(2d)$. Thus, for large n the mutational step size has to be much smaller than the distance to the optimum, $r \sim d/\sqrt{n} \ll d$, for the mutation to have a chance of increasing fitness.

As has become customary in the field, we here assume that the mutational displacements are independent and identically distributed (i.i.d.) random variables drawn from a n-dimensional Gaussian distribution with zero mean. The covariance matrix can be taken to be of diagonal form $\sigma^2 I$, where I is the n-dimensional identity matrix and σ^2 is the variance of a single trait (Blanquart *et al.* 2014). In the limit $n \to \infty$ the form of the distribution of the mutational displacements becomes irrelevant owing to the central limit theorem, and therefore Fisher's result Equation 1 holds also in the present setting of Gaussian mutational displacements of mean size $r = \sigma\sqrt{n}$ (Waxman and Welch 2005; Ram and Hadany 2015); an explicit derivation will be provided below. Because lengths in the phenotype space can be naturally measured in units of σ, the parameters d and σ should always appear as the ratio d/σ as can be seen in Equation 1. Thus, without loss of generality, we can set $\sigma = 1$. In the following we denote the (scaled) wild type phenotype by $\vec{Q}$, its distance to the optimum by

$$Q = |\vec{Q}| = \frac{d}{\sigma}, \qquad (2)$$

and draw the displacement vectors $\vec{\xi}$ from the n-dimensional Gaussian $p(\vec{\xi})$ with unit covariance matrix.

By normalizing phenotypic distances to the standard deviation σ of the mutational effect on a single trait, we are adopting a particular pleiotropic scaling that has been referred to as the "Euclidean superposition model" (Wagner *et al.* 2008; Hermisson and McGregor 2008). An alternative choice which is closer to Fisher's original formulation but appears to have less empirical support is the "total effect model", wherein the total length r of the mutational displacements is taken to be independent of n. Since $r = \sigma\sqrt{n}$, this implies that the single trait effect size decreases with n as $\sigma \sim 1/\sqrt{n}$. As a consequence the parameter Q defined by Equation 2 becomes n-dependent and increases as $\sqrt{n}$, provided d does not depend on n (Orr 2000). The results presented below will always be given in terms of ratios of the basic parameters of FGM, such that their translation to the total effect model is in principle straightforward. We will nevertheless explicitly point out instances where the two settings give rise to qualitatively different behaviors.
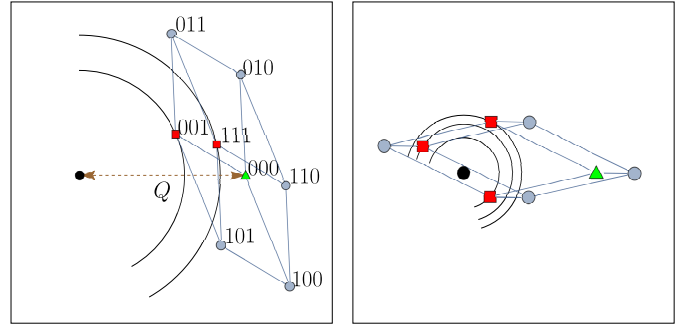

Figure 1 Examples of three-dimensional genotypic fitness landscapes induced by FGM with two phenotypic dimensions ($L = 3$ and $n = 2$). The panels show the projection of the discrete genotype space onto the phenotype plane, where the phenotypic optimum is represented by a black dot. In the left panel the binary sequence notation for genotypes is indicated. The wild type genotype 000, marked by a green triangle, is located at distance Q from the phenotypic optimum. The nodes represented by red squares are local fitness maxima of the genotypic landscapes, as can be seen from the contour lines of constant fitness. In the right panel the mutant phenotypes overshoot the optimum, whereas in the left panel they do not.

The genotypic fitness landscape induced by FGM

In order to study epistasis within FGM, Fisher's original definition has to be supplemented with a rule for how the effects of multiple mutations are combined. Based on earlier work (Lande 1980) in quantitative genetics, Martin *et al.* (2007) introduced the assumption that mutations act additively on the level of the phenotype. Thus the phenotype arising from two mutations $\vec{\xi}_1$, $\vec{\xi}_2$ applied to the wild type $\vec{Q}$ is simply given by $\vec{Q} + \vec{\xi}_1 + \vec{\xi}_2$. This definition suffices to associate an L-dimensional genotypic fitness landscape to any set of L mutational displacements $\vec{\xi}_1, \vec{\xi}_2, \ldots, \vec{\xi}_L$ (Blanquart *et al.* 2014). For this purpose the haploid genotype τ is represented by a binary sequence with length L, $\tau = (\tau_1, \tau_2, \ldots, \tau_L)$ with $\tau_i = 1$ ($\tau_i = 0$) in the presence (absence) of the i'th mutation. For the wild type $\tau_i = 0$ for all i, and in general the phenotype vector associated with the genotype τ reads

$$\vec{z}(\tau) = \vec{Q} + \sum_{i=1}^L \tau_i \vec{\xi}_i. \qquad (3)$$

Two examples illustrating this genotype-phenotype map and the resulting genotypic fitness landscapes with $L = 3$ and $n = 2$ are shown in Figure 1.

Since fitness decreases monotonically with the distance to the optimum phenotype, a natural proxy for fitness is the negative squared magnitude of the phenotype vector

$$-|\vec{z}(\tau)|^2 = -|\vec{Q}|^2 - 2\sum_{i=1}^L (\vec{Q} \cdot \vec{\xi}_i)\tau_i - \sum_{i,j=1}^L (\vec{\xi}_i \cdot \vec{\xi}_j)\tau_i\tau_j, \qquad (4)$$

where $\vec{x} \cdot \vec{y}$ denotes the scalar product between two vectors $\vec{x}$ and $\vec{y}$. This quantity is thus seen to consist of a part that is additive across loci with coefficients given by the scalar products $\vec{Q} \cdot \vec{\xi}_i$ and a pairwise epistatic part with coefficients $\vec{\xi}_i \cdot \vec{\xi}_j$.

It is instructive to decompose Equation 4 into contributions from the mutational displacements parallel and perpendicular

to $\vec{Q}$. Writing $\vec{\xi}_i = \xi_i^{\parallel} Q^{-1}\vec{Q} + \vec{\xi}_i^{\perp}$ with $\vec{Q} \cdot \vec{\xi}_i^{\perp} = 0$, Equation 4 can be recast into the form

$$-|\vec{z}(\tau)|^2 = -\left(Q + \sum_{i=1}^{L} \xi_i^{\parallel} \tau_i\right)^2 - \sum_{i,j=1}^{L} (\vec{\xi}_i^{\perp} \cdot \vec{\xi}_j^{\perp})\tau_i\tau_j. \quad (5)$$

The first term on the right hand side contains both additive and epistatic contributions associated with displacements along the $\vec{Q}$-direction. The second term is dominated by the diagonal contributions with $i = j$ and is of order $L(n-1)$ because $|\vec{\xi}_i^{\perp}|^2 = n - 1$ on average.

We now show how the first term on the right hand side of Equation 5 can be made to vanish for a range of Q. For this purpose consider the subset of phenotypic displacement vectors for which the component $\xi_i^{\parallel}$ in the direction of $\vec{Q}$ is negative. There are on average $L/2$ such mutations, and the expected value of each component is

$$2\int_{-\infty}^{0} dy \, \frac{y}{\sqrt{2\pi}} e^{-y^2/2} = -\sqrt{\frac{2}{\pi}} \equiv -2q_0, \quad (6)$$

where the factor 2 in front of the integral arises from conditioning on $\xi_i^{\parallel} < 0$. Setting $\tau_i = 1$ for s out of these $L/2$ vectors and $\tau_i = 0$ for all other mutations, the sum inside the brackets in Equation 5 becomes approximately equal to $-2q_0 s$ which cancels the Q-term for $s = Q/2q_0$. Since s can be at most $L/2$ in a typical realization, such genotypes can be constructed with a probability approaching unity provided $Q < q_0 L$.

We will see below that the structure of the genotypic fitness landscapes induced by FGM depends crucially on whether or not the phenotypes of multiple mutants are able to closely approach the phenotypic optimum. Assuming that the contributions from the perpendicular displacements in Equation 5 can be neglected, which will be justified shortly, the simple argument given above shows that a close approach to the optimum is facile when $Q < q_0 L$ but becomes unlikely when $Q \gg q_0 L$. This observation hints at a possible transition between different types of landscape topographies at some value of Q which is proportional to L. The existence and nature of this transition is a central theme of this paper.

Scaling limits

Since we are interested in describing complex organisms with large phenotypic and genotypic dimensions, appropriate scaling relations have to be imposed to arrive at meaningful asymptotic results. Three distinct scaling limits will be considered.

(i) Fisher's classic result (Equation 1) shows that the distance of the wild type from the phenotypic optimum has to be increased with increasing n in order to maintain a nonzero fraction of beneficial mutations for $n \to \infty$. In our notation Fisher's parameter is

$$x = \frac{n}{2Q} \quad (7)$$

and hence *Fisher scaling* implies taking $n, Q \to \infty$ at fixed ratio n/Q. We will extend Fisher's analysis by computing the probability of sign epistasis between pairs of mutations for fixed x and large n, which amounts to characterizing the shape of genotypic fitness landscapes of size $L = 2$.

(ii) We have argued above that the distance towards the phenotypic optimum that can be covered by typical multiple mutations is of order L, and hence the limit $L \to \infty$ is naturally

accompanied by a limit $Q \to \infty$ at fixed ratio

$$q = \frac{Q}{L}. \quad (8)$$

From a biological point of view one expects that $L \gg n \gg 1$, which motivates to consider the limit $L, Q \to \infty$ at constant phenotypic dimension n. Under this scaling the first term on the right hand side of Equation 5 is of order L^2, whereas the contribution from the perpendicular displacements is only $(n-1)L$. Thus in this regime the topography of the fitness landscape is determined mainly by the one-dimensional mutational displacements in the $\vec{Q}$-direction, which is reflected by the fact that the genotypic complexity is independent of n to leading order and coincides with its value for the case $n = 1$, in which the perpendicular contribution in Equation 5 does not exist (see **Results**).

(iii) By contrast, the perpendicular displacements play an important role when both the phenotypic and genotypic dimensions are taken to infinity at fixed ratio

$$\alpha = \frac{n}{L}. \quad (9)$$

Combining this with the limit $Q \to \infty$ at fixed $q = Q/L$, both terms on the right hand side of Equation 5 are of the same order $\sim L^2$. Fisher's parameter (Equation 7) is then also a constant given by $x = \alpha/(2q)$.

Preliminary considerations about genotypic fitness maxima

To set the stage for the detailed investigation of the number of genotypic fitness maxima in **Results**, it is useful to develop some intuition for the behavior of this quantity based on the elementary properties of FGM that have been described so far. For this purpose we consider the probability P_{wt} for the wild type to be a local fitness maximum, which is equal to the probability that all the L mutations are deleterious. Since mutations are statistically independent, we have

$$P_{\text{wt}} = [1 - P_b]^L = 2^{-L}\left[1 + \text{erf}(x/\sqrt{2})\right]^L, \quad (10)$$

where $\text{erf} = 1 - \text{erfc}$ is the error function. Under the (highly questionable) assumption that this estimate can be applied to all 2^L genotypes in the landscape, we arrive at the expression

$$\mathcal{N}_{\text{wt}} = 2^L P_{\text{wt}} = \left[1 + \text{erf}(x/\sqrt{2})\right]^L \quad (11)$$

for the expected number of genotypic fitness maxima.

Consider first the scaling limit (ii), where $x = n/(2Q) = n/(2qL) \to 0$. Expanding the error function for small arguments as $\text{erf}(y) \approx 2y/\sqrt{\pi}$ we obtain

$$\mathcal{N}_{\text{wt}} \approx \left[1 + \frac{2x}{\sqrt{2\pi}}\right]^L \to \exp\left(\frac{q_0 n}{q}\right) \quad (12)$$

for $L \to \infty$, where $q_0 = 1/\sqrt{2\pi}$ was defined in Equation 6. We will show below that this expression correctly captures the asymptotic behavior for very large q but generally grossly underestimates the number of maxima. The reason is that for moderate values of q (in particular for $q < q_0$) the relevant mutant phenotypes are much closer to the origin than the wild type, which entails a mechanism for generating a large number of fitness maxima that grows exponentially with L.

Such an exponential dependence on L is expected from Equation 11 in the scaling limit (iii), where $x = \alpha/(2q)$ is a nonzero

Table 1 List of Mathematical Symbols

Symbols	Description
n	Number of phenotypic traits, also referred to as phenotypic dimension.
L	Length of the binary genetic sequence, also referred to as genotypic dimension.
α	Ratio of phenotypic to genotypic dimension ($\alpha = n/L$).
$\vec{Q}, Q$	n-dimensional vector $\vec{Q}$ representing the wild type phenotype and its magnitude $Q = |\vec{Q}|$.
$\vec{q}, q$	Wild type phenotype vector in units of L ($\vec{q} = \vec{Q}/L$) and its magnitude ($q = |\vec{q}|$).
τ	Genotype represented by a binary sequence of length L.
τ_i	Binary number indicating absence (0) or presence (1) of a mutation at site i ($i = 1, \ldots, L$) of the genotype τ.
$\vec{z}(\tau), z(\tau)$	Phenotype vector corresponding to the genotype τ (Equation 3) and its magnitude $z = |\vec{z}|$.
$\vec{\xi}, \xi$	Random phenotypic displacement vector representing a mutation and its magnitude $\xi = |\vec{\xi}|$.
$p(\vec{\xi})$	Probability density of the random vector $\vec{\xi}$. In this paper, the density is Gaussian with unit covariance matrix .
x	Fisher's scaling parameter; in our notation $x = \frac{n}{2Q}$.
$\mathcal{N}$	Total number of fitness maxima in a genotypic fitness landscape averaged over all realizations of sets of $\vec{\xi}$'s.
Σ^*	Genotypic complexity defined as the ratio of $\ln \mathcal{N}$ to L for $L \to \infty$ (Equation 13).
q_c	Transition point of q that separates regimes I and II.
q_0	Half of the average mutational displacement ξ_i of a single trait conditioned on being positive ($q_0 = 1/\sqrt{2\pi}$).
ρ	Fraction of mutations that are present in a genotype τ, $\rho = L^{-1} \sum_{i=1}^{L} \tau_i$.
ρ^*	Mean value of ρ of a local maximum, also referred to as mean genotypic distance from the wild type.
z^*	Mean value of $z(\tau)$ of a local maximum, also referred to as mean phenotypic distance from the optimum.

constant and the expression in the square brackets is larger than 1. Although this general prediction is confirmed by the detailed analysis for this case, the behavior of the number of maxima predicted by Equation 11 will again turn out to be valid only when q is very large. In particular, whereas Equation 11 is an increasing function of α for any q, we will see below that the expected number of maxima actually decreases with increasing phenotypic dimension (hence increasing α) in a substantial range of q. In qualitative terms, this can be attributed to the effect of the perpendicular displacements in Equation 5, which grows with α and makes it increasingly more difficult for the mutant phenotypes to closely approach the origin.

The observation that the number of genotypic fitness maxima grows exponentially with L in most cases motivates to make use of the corresponding growth rate as a measure of the ruggedness of the landscape. We therefore define the *genotypic complexity* Σ^* through the limiting relation

$$\Sigma^* = \lim_{L \to \infty} \frac{\ln \mathcal{N}}{L}, \tag{13}$$

where $\mathcal{N}$ is the *average* number of genotypic fitness maxima and L is the sequence length. Since the total number of binary genotypes is 2^L, the complexity is bounded from above by $\ln 2$. If any genotype had the same probability $P_{\max}$ of being a fitness maximum (which is in fact not the case for FGM), we could write $\mathcal{N} = 2^L P_{\max}$ and hence $P_{\max} \sim \exp[-(\ln 2 - \Sigma^*)L]$.

Data availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within

the article. All numerical calculations including simulations[1] described in this work were implemented in Mathematica and C++. All relevant source codes are available upon request.

Results

Preliminary note

In the following sections our results on the structure of genotypic fitness landscapes induced by FGM are stated in precise mathematical terms and the key steps of their derivation are outlined, with some technical details relegated to the Appendices. In order to facilitate the navigation through the inevitable mathematical formalism, we display the definitions of the most commonly used mathematical symbols in Table 1. Moreover, we provide numbered summaries at the end of each subsection which state the main results without resorting to mathematical expressions.

Sign epistasis

We first study the local topography of the fitness landscape around the wild type, focusing on the epistasis between two random mutations with phenotypic displacements $\vec{\xi}$ and $\vec{\eta}$. Since fitness is determined by the magnitude of a phenotypic vector, i.e., the distance of the phenotype from the origin, the epistatic effect of the two mutations can be understood by analyzing how the magnitudes of the four vectors $\vec{Q}, \vec{Q} + \vec{\xi}, \vec{Q} + \vec{\eta}$

---

[1] When counting the number of local genotypic maxima, we checked all genotypes and counted the number exactly for a randomly realized landscape, then took an average.
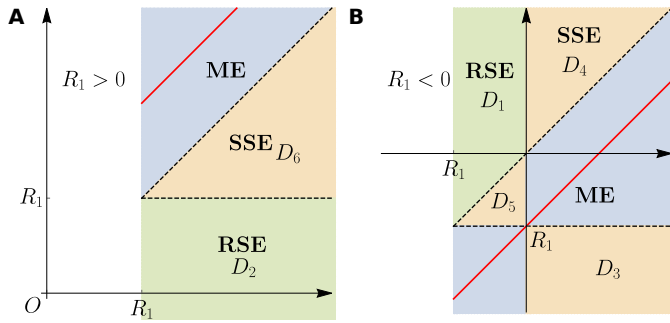
Figure 2 Domains in the (R_2, R)-plane contributing to different types of epistasis: magnitude epistasis (ME), simple sign epistasis (SSE) and reciprocal sign epistasis (RSE). The two panels illustrate the two cases: (**A**) $R_1 > 0$ and (**B**) $R_1 < 0$. The red solid lines indicate $R = R_1 + R_2$. The labeling of the domains $D_1, \ldots, D_6$ is used in the derivation in Appendix B.

and $\vec{Q} + \vec{\xi} + \vec{\eta}$ are ordered. To this end, we introduce the quantities

$$R_1 \equiv \frac{1}{n}\left(|\vec{\xi} + \vec{Q}|^2 - Q^2\right), \quad R_2 \equiv \frac{1}{n}\left(|\vec{\eta} + \vec{Q}|^2 - Q^2\right),$$
$$R \equiv \frac{1}{n}\left(|\vec{\xi} + \vec{\eta} + \vec{Q}|^2 - Q^2\right), \tag{14}$$

where division by n guarantees the existence of a finite limit for $n \to \infty$. The sign of these quantities determines whether a mutation is beneficial or deleterious. For example, if $R_1 < 0$, the mutation $\vec{\xi}$ is beneficial; if $R > 0$ the two mutations combined together confer a deleterious effect; and so on. We will see later that $R_{1,2}$ and R are actually closely related to the selection coefficients of the respective mutations.

We proceed to express the different types of pairwise epistasis defined by Weinreich *et al.* (2005) and Poelwijk *et al.* (2007) in terms of conditions on the quantities defined in Equation 14. Without loss of generality we assume $R_1 < R_2$ and consider first the case where both mutations are beneficial, $R_1 < R_2 < 0$. Then magnitude epistasis (ME), the absence of sign epistasis, applies when the fitness of the double mutant is higher than that of each of the single mutants, i.e., $R < R_1 < R_2 < 0$. Similarly for two deleterious mutations the condition for ME reads $R > R_2 > R_1 > 0$. When one mutant is deleterious and the other beneficial, in the case of ME the double mutant fitness has to be intermediate between the two single mutants, which implies that $R_1 < R < R_2$ when $R_2 > 0 > R_1$.

The condition for reciprocal sign epistasis (RSE) reads $R > R_2 > R_1$ when both single mutants are beneficial and $R < R_1 < R_2$ when both are deleterious, and the remaining possibility $R_1 < R < R_2$ corresponds to simple sign epistasis (SSE) between two mutations of the same sign. If the two single mutant effects are of different signs, RSE is impossible and SSE applies when $R < R_1 < 0 < R_2$ or $R > R_2 > 0 > R_1$. Figure 2 depicts the different categories of epistasis as regions in the (R_2, R)-plane. Note that the corresponding picture for $R_1 > R_2$ is obtained by exchanging $R_1 \leftrightarrow R_2$.

To find the probability of each epistasis, we require the joint probability density $\mathcal{P}(R_1, R_2, R)$. In Appendix A it is shown

that

$$\mathcal{P}(R_1, R_2, R) = \frac{x^2 n^{1/2}}{4\sqrt{2}\pi^{3/2}} e^{-\frac{1}{8}n(-R+R_1+R_2)^2 - \frac{x^2}{2}((R_1-1)^2+(R_2-1)^2)}$$
$$\times \left(1 + O\left(\frac{1}{n}\right)\right), \tag{15}$$

which can be obtained rather easily by resorting to the central limit theorem (CLT). The applicability of the CLT follows from the fact that $R_{1,2}$ and R are sums of a large number of independent terms for $n \to \infty$ (Waxman and Welch 2005; Ram and Hadany 2015). According to the CLT, it is sufficient to determine the first and second cumulants of these quantities. Denoting averages by angular brackets, we find the mean $\langle R_i \rangle = 1$, the variance $\langle R_i^2 \rangle - \langle R_i \rangle^2 = 1/x^2$, and the covariance $\langle R_1 R_2 \rangle - \langle R_1 \rangle \langle R_2 \rangle = 0$ ($i = 1, 2$). Similarly, the corresponding quantitites evaluated for $R - R_1 - R_2$ are $\langle R - R_1 - R_2 \rangle = 0$, $\langle (R - R_1 - R_2)^2 \rangle - \langle R - R_1 - R_2 \rangle^2 = 4/n$, and $\langle (R - R_1 - R_2)R_i \rangle - \langle R - R_1 - R_2 \rangle \langle R_i \rangle = 0$ ($i = 1, 2$). With an appropriate normalization constant, this leads directly to Equation 15.

As a first application, we rederive Fisher's Equation 1 by integrating $\mathcal{P}(R_1, R_2, R)$ over the region $R_1 < 0$ for all R_2 and R, which indeed yields

$$P_b = \int_{-\infty}^{0} dR_1 \int_{-\infty}^{\infty} dR_2 \int_{-\infty}^{\infty} dR \, \mathcal{P}(R_1, R_2, R) = \frac{1}{2}\text{erfc}\left(\frac{x}{\sqrt{2}}\right).$$

An immediate conclusion from the form of $\mathcal{P}(R_1, R_2, R)$ is that it is unlikely to observe sign epistasis for large n, because $\mathcal{P}(R_1, R_2, R)$ becomes concentrated along the line $R = R_1 + R_2$ as n increases. As can be seen in Figure 2, this line touches the region of SSE in one point for $R_1 < 0$, whereas it maintains a finite distance to the region of RSE everywhere. This indicates that the probability of reciprocal sign epistasis decays more rapidly with increasing n than the probability of simple sign epistasis. Moreover, one expects the latter probability to be proportional to the width of the region around the line $R = R_1 + R_2$ where the joint probability in Equation 15 has appreciable weight, which is of order $1/\sqrt{n}$.

To be more quantitative, we need to integrate $\mathcal{P}(R_1, R_2, R)$ over the domains in Figure 2 corresponding to the different categories of epistasis. In Appendix B, we obtain the asymptotic expressions

$$P_{\text{RSE}} = \frac{2x^2}{\pi n} e^{-x^2} + O(n^{-3/2}), \tag{16}$$

$$P_{\text{SSE}} = \frac{4x}{\pi\sqrt{n}} e^{-x^2/2} + O\left(n^{-1}\right), \tag{17}$$

for the probabilities of reciprocal (P_{RSE}) and simple (P_{SSE}) sign epistasis. Due to the non-linearity of the phenotype-fitness map, FGM does not allow for strictly non-epistatic combination of fitness effects. The probability of magnitude epistasis, therefore, is given by $P_{\text{ME}} = 1 - P_{\text{RSE}} - P_{\text{SSE}}$. Interestingly, the probability of sign epistasis varies non-monotonically with x. To confirm our analytic results, we compare our results with simulations in Figure 3, which shows an excellent agreement.

Similarly, we can calculate the probabilities of sign epistasis conditioned on both mutations being beneficial, which in our setting means $R_2 < 0$. The conditioning requires normalization by the unconditional probability of two random mutations being beneficial, which is given by the square of P_b in Equation 1.
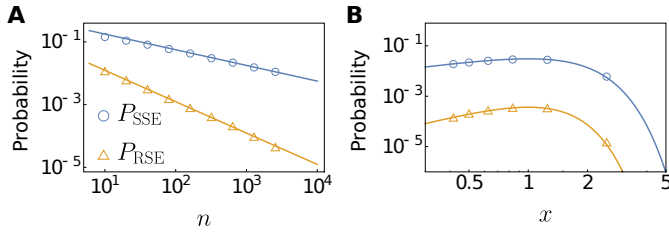
Figure 3 Comparison of analytic results for the probability of epistasis with simulations. Depicted are probabilities of simple (P_{SSE}) and reciprocal (P_{RSE}) sign epistasis between two randomly chosen mutations among nearest neighbor genotypes of the wild type (**A**) as functions of n for fixed Fisher parameter $x = 0.5$ and (**B**) as function of x for fixed phenotypic dimension $n = 640$. For each parameter set, 10^4 randomly generated landscapes were analyzed. The asymptotic expressions provide accurate approximations even for moderate $n > 10$. The non-monotonic behavior with respect to x means that the probabilities are non-monotonic functions of Q for fixed n and vice versa.

Hence

$$P_{RSE}^b = \frac{2\Pr(D_1)}{P_b^2} \approx \frac{4x^2}{\pi n \, \mathrm{erfc}\left(x/\sqrt{2}\right)^2} e^{-x^2} \qquad (18)$$

and

$$P_{SSE}^b = \frac{2\Pr(D_5)}{P_b^2} \approx \frac{4x}{\pi \sqrt{n} \, \mathrm{erfc}\left(x/\sqrt{2}\right)} e^{-x^2/2}, \qquad (19)$$

where $\Pr(D_i)$ denotes the integral of the joint probability density over the domain D_i in Figure 2 (see Appendix B).

As anticipated from the form of Equation 15, the fraction of sign epistatic pairs of mutations decreases with increasing phenotypic dimension n, and this decay is faster for reciprocal sign epistasis ($\sim 1/n$) than for simple sign epistasis ($\sim 1/\sqrt{n}$). At first glance this might seem to suggest that FGM has little potential for generating rugged genotypic fitness landscapes. However, as we will see below, the results obtained in this section apply only to the immediate neighborhood of the wild type phenotype. They are modified qualitatively in the presence of a large number of mutations that are able to substantially displace the phenotype and allow it to approach the phenotypic optimum.

As a slight variation to the previous setting, one may consider the fraction of sign epistasis conditioned on the two single mutations to have the same selection strength, as recently investigated by Schoustra *et al.* (2016). In our notation this implies that $R_1 = R_2 \equiv \tilde{R}$, and it is easy to see that sign epistasis is always reciprocal in this case. If the two mutations are beneficial $\tilde{R} < 0$, and the condition for (reciprocal) sign epistasis is $R > \tilde{R}$. The corresponding probability is

$$\tilde{P}_{RSE}(\tilde{R}) = \frac{\int_{\tilde{R}}^{\infty} \mathcal{P}(\tilde{R}, \tilde{R}, R)\, dR}{\int_{-\infty}^{\infty} \mathcal{P}(\tilde{R}, \tilde{R}, R)\, dR} = \frac{1}{2}\mathrm{erfc}\left(-\frac{\sqrt{n}\tilde{R}}{2\sqrt{2}}\right). \qquad (20)$$

Following the same procedure for deleterious mutations ($\tilde{R} > 0$) one finds that the probability is actually symmetric around $\tilde{R} = 0$ and hence depends only on $|\tilde{R}|$.

In order to express $\tilde{P}_{RSE}$ in terms of the selection coefficient of the single mutations we introduce a Gaussian phenotypic fitness function of the form

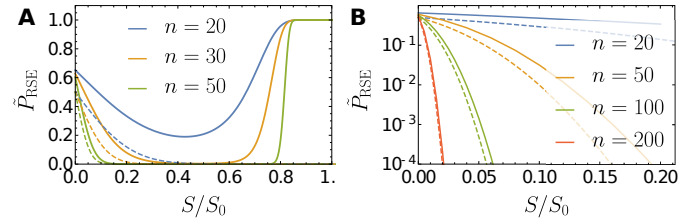$$W(\vec{y}) = W_0 \exp(-\lambda |\vec{y}|^2), \qquad (21)$$

Figure 4 Probability of reciprocal sign epistasis $\tilde{P}_{RSE}$ conditioned on the selection coefficients S of the two single mutations to be equal and positive. Here, the fitness of a phenotype $\vec{y}$ is assumed to be $W(\vec{y}) = W_0 \exp(-\lambda |\vec{y}|^2)$, where the parameter λ is related to the maximal beneficial selection coefficient S_0 through the relation $S_0 = \lambda Q^2$. Dashed lines depict the asymptotic expression Equation 23, and solid lines were obtained numerically using the Gaussian approximation for the distribution of epistasis developed by Schoustra *et al.* (2016).

where $\lambda > 0$ is a measure for the strength of selection. The selection coefficient of a mutation with phenotypic effect $\vec{\xi}$ is then given by

$$S = \ln\left[\frac{W(\vec{Q} + \vec{\xi})}{W(\vec{Q})}\right] = -\lambda\left(|\vec{Q} + \vec{\xi}|^2 - |\vec{Q}|^2\right) = -\lambda n \tilde{R}. \qquad (22)$$

To fix the value of λ we note that the largest possible selection coefficient, which is achieved for mutations that reach the phenotypic optimum, is $S_0 = \lambda Q^2$, and hence $\tilde{R}$ is related to the selection coefficient through $\tilde{R} = -\frac{Q^2}{n}\frac{S}{S_0}$. With this substitution the result in Equation 20 becomes

$$\tilde{P}_{RSE}(S) = \frac{1}{2}\mathrm{erfc}\left(\frac{n^{3/2}}{8\sqrt{2}x^2}\frac{|S|}{S_0}\right). \qquad (23)$$

The probability of sign epistasis conditioned on selection strength takes on its maximal value $\tilde{P}_{RSE} = \frac{1}{2}$ in the neutral limit $S \to 0$ and decreases monotonically with $|S|$. Similar to the results of Equation 16, Equation 17 and Equation 18 for unconstrained mutations, it also decreases with increasing phenotypic dimension n when S and x are kept fixed.

In a previous numerical study carried out at finite Q and n it was found that $\tilde{P}_{RSE}$ varies non-monotonically with S for the case of beneficial mutations, and displays a second peak at the maximum selection coefficient $S = S_0$ (Schoustra *et al.* 2016). The two peaks were argued to reflect the two distinct mechanisms giving rise to sign epistasis within FGM (Blanquart *et al.* 2014). Mutations of small effect correspond to phenotypic displacements that proceed almost perpendicular to the direction of the phenotypic optimum, and sign epistasis is generated through antagonistic pleiotropy. On the other hand, for mutations of large effect the dominant mechanism for sign epistasis is through overshooting of the phenotypic optimum. Because of the Fisher scaling implemented in this section with $Q, n \to \infty$ at fixed $x = n/(2Q)$, the second class of mutations cannot be captured by our approach and only the peak at small S remains. Figure 4 (A) shows the full two-peak structure for a few representative values of n, and Figure 4 (B) illustrates the convergence to the asymptotic expression Equation 23 for the left peak. Using the results of Schoustra *et al.* (2016), it can be shown that the right peak becomes a step function for $n \to \infty$, displaying a discontinuous jump from $\tilde{P}_{RSE} = 0$ to $\tilde{P}_{RSE} = 1$ at $S/S_0 = 8/9 = 0.888\ldots$.
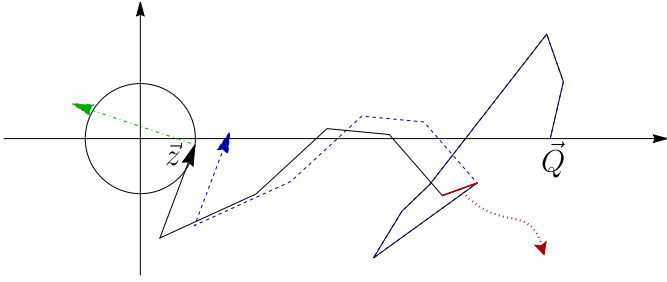
Figure 5 Illustration of the condition for a genotype to be a local fitness maximum. The black circle encloses phenotypes that have higher fitness than the focal phenotype $\vec{z}(\tau)$. For τ to be a genotypic fitness maximum, both a phenotype with a further mutation (dash-dotted green arrow) and a phenotype without one of the mutations in τ (red segment and blue dotted arrows) should lie outside the circle.

Summary 1. When the phenotypic dimension n is large and the Fisher parameter x is moderate, the probability of reciprocal sign epistasis decays as $1/n$, while that of simple sign epistasis decays as $1/\sqrt{n}$. Although these probabilities decrease monotonically with n at fixed x, they have a non-monotonic behavior as a function of x: For small x they increase with x and for large x they decrease with x (see Figure 3). Under the pleiotropic scaling adopted in this work this implies that the probabilities are non-monotonic function of the wild type distance Q at fixed n and vice versa. In contrast, under the total effect model, where both the wild type distance Q and x scale as $\sqrt{n}$, the probabilities decrease monotonically and exponentially with n.

Genotypic complexity at fixed phenotypic dimension

In this section, we are interested in the number of local maxima in the genotypic fitness landscape. We focus on the expected number of maxima, which we denote by $\mathcal{N}$, and analyze how this quantity behaves in the limit of large genotypic dimension, $L \to \infty$, when the phenotypic dimension n is fixed. For the sake of clarity, the (unique) maximum of the phenotypic fitness landscape will be referred to as the phenotypic *optimum* throughout.

The number of local fitness maxima. Since fitness decreases monotonically with the distance to the phenotypic optimum, a genotype τ is a local fitness maximum if the corresponding phenotype defined by Equation 3 satisfies

$$|\vec{z}(\tau)| < |\vec{z}(\tau) + (1 - 2\tau_i)\vec{\xi}_i|, \tag{24}$$

for all $1 \leq i \leq L$. The phenotype vector appearing on the right hand side of this inequality arises from $\vec{z}(\tau)$ either by removing a mutation vector that is already part of the sum in Equation 3 ($\tau_i = 1$) or by adding a mutation vector that was not previously present ($\tau_i = 0$). The condition Equation 24 is obviously always fulfilled if $\vec{z}(\tau) = 0$, that is, if the phenotype is optimal, and we will see that in general the probability for this condition to be satisfied is larger, the more closely the phenotype approaches the origin. A graphical illustration of the condition Equation 24 is shown in Figure 5.

The ability of a phenotype $\vec{z}(\tau)$ to approach the origin clearly depends on the number $s = \sum_{i=1}^{L} \tau_i$ of mutant vectors it is composed of, and all phenotypes with the same number of mutations are statistically equivalent. The expected number of fit-

ness maxima can therefore be decomposed as

$$\mathcal{N} = \sum_{s=0}^{L} \binom{L}{s} \mathcal{R}_s(L), \tag{25}$$

where $\binom{L}{s}$ is the number of possible combinations of s out of L mutation vectors and $\mathcal{R}_s(L)$ is the probability that a genotype with s mutations is a fitness maximum. The latter can be written as

$$\mathcal{R}_s(L) = \int_n d\vec{z} \left[\prod_{i=s+1}^{L} \int_{\mathcal{D}(-\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right] \times$$
$$\left[\prod_{i=1}^{s} \int_{\mathcal{D}(\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right] \delta\left(\vec{z} - \vec{Q} - \sum_{i=1}^{s} \vec{\xi}_i \right), \tag{26}$$

with

$$\mathcal{D}(\vec{y}) \equiv \left\{ \vec{\xi} \in \mathbb{R}^n \middle| |\vec{\xi} - \vec{y}| > |\vec{y}| \right\}. \tag{27}$$

Here and below, $\int_n$ stands for the integral over $\mathbb{R}^n$.

Equation 26 can be understood as follows. First, the delta function $\delta\left(\vec{z} - \vec{Q} - \sum_{i=1}^{s} \vec{\xi}_i \right)$ constrains $\vec{z}$ to be the phenotype of τ as defined in Equation 3. Next, the integration domains of the $\vec{\xi}_i$'s reflect the condition in Equation 24. Assuming without loss of generality that the L genetic loci are ordered such that $\tau_i = 1$ for $i \leq s$ and $\tau_i = 0$ for $i > s$, the maximum condition for $i \leq s$ requires $|\vec{z}| < |\vec{z} - \vec{\xi}_i|$, so the integration domain should be $\mathcal{D}(\vec{z})$, whereas for $i > s$ the condition is $|\vec{z}| < |\vec{z} + \vec{\xi}_i|$, corresponding to the integration domain $\mathcal{D}(-\vec{z})$. Using the integral representation of the delta function

$$\delta(\vec{y}) = \frac{1}{(2\pi)^n} \int_n d\vec{k} \exp\left(i\vec{k} \cdot \vec{y} \right), \tag{28}$$

we can write

$$\mathcal{R}_s(L) = \int_n \int_n \frac{d\vec{z} d\vec{k}}{(2\pi)^n} \exp\left[i\vec{k} \cdot \left(\vec{z} - \vec{Q} \right) \right] F(\vec{k}, \vec{z})^s F(0, -\vec{z})^{L-s}, \tag{29}$$

where

$$F(\vec{k}, \vec{z}) \equiv \int_{\mathcal{D}(\vec{z})} d\vec{\xi} p(\vec{\xi}) \exp\left(-i\vec{k} \cdot \vec{\xi} \right). \tag{30}$$

It was argued on qualitative grounds in **Model** that phenotypes that approach arbitrarily close to the origin are easily generated when the scaled wild type distance q is small, but they become rare for large q. As a consequence, it turns out that the main contribution to the integral over $\vec{z}$ in Equation 29 comes from the region around the origin $\vec{z} = 0$ for small q but shifts to a distance $z \sim L$ along the $\vec{Q}$-direction for large q. To account for this possibility, it is necessary to divide the integral domain into two parts $|\vec{z}| < z_0$ and $|\vec{z}| > z_0$, where z_0 is an arbitrary non-zero number with $z_0/L \to 0$ as $L \to \infty$. Thus, we write $\mathcal{R}_s(L)$ as

$$\mathcal{R}_s(L) = \mathcal{R}_s^<(L) + \mathcal{R}_s^>(L), \tag{31}$$

where

$$\mathcal{R}_s^<(L) = \int_{|\vec{z}|<z_0} d\vec{z} \int_n \frac{d\vec{k}}{(2\pi)^n} e^{i\vec{k}\cdot(\vec{z}-\vec{Q})} F(\vec{k}, \vec{z})^s F(0, -\vec{z})^{L-s},$$

$$\mathcal{R}_s^>(L) = \int_{|\vec{z}|>z_0} d\vec{z} \int_n \frac{d\vec{k}}{(2\pi)^n} e^{i\vec{k}\cdot(\vec{z}-\vec{Q})} F(\vec{k}, \vec{z})^s F(0, -\vec{z})^{L-s}, \tag{32}$$
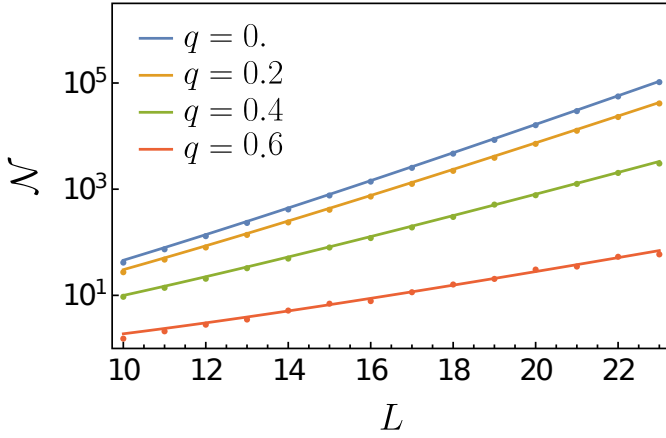
Figure 6 Plots of mean number of local maxima $\mathcal{N}$ as a function of the genotypic dimension L for $q = 0, 0.2, 0.4$, and 0.6 with $n = 1$ on a semi-logarithmic scale. Data from numerical simulations are represented as dots, and the analytical prediction Equation 42 is shown as solid lines. Each dot represents the average over 10^5 realizations of landscapes. In this parameter regime, $\mathcal{N}$ grows exponentially with L and the growth rate (i.e., the slopes of the lines) decreases with increasing q.

and correspondingly define $\mathcal{N}_<$ and $\mathcal{N}_>$ as

$$\mathcal{N}_< = \sum_s \binom{L}{s} \mathcal{R}_s^<(L), \quad \mathcal{N}_> = \sum_s \binom{L}{s} \mathcal{R}_s^>(L). \quad (33)$$

The total number of local maxima is then $\mathcal{N} = \mathcal{N}_< + \mathcal{N}_>$.

Regime I. We first consider $\mathcal{R}_s^<(L)$. Expanding $F(\vec{k}, \vec{z})$ around the origin $\vec{z} = 0$, we show in Appendix C that

$$\mathcal{R}_s^<(L) \approx \frac{s^{-n/2} \exp\left[-Q^2/(2s)\right]}{s \exp[-Q^2/(2s^2)] + L - s}. \quad (34)$$

For an interpretation of Equation 34 it is helpful to refer to Figure 5. Note first that the probability that $\vec{z} = \vec{Q} + \sum_{i=1}^s \vec{\xi}_i$ lies in the ball $|\vec{z}| < \zeta$ with radius $\zeta \ll 1$ is

$$\text{Prob}(|\vec{z}| < \zeta) \approx \frac{V_n}{(2\pi)^{n/2}} s^{-n/2} \exp\left[-Q^2/(2s)\right], \quad (35)$$

where $V_n(\zeta) \sim \zeta^n$ is the volume of the ball. We need to estimate how small ζ has to be for τ to be a local fitness maximum with an appreciable probability. Since the s random vectors contributing to $\vec{z}$ are statistically equivalent, it is plausible to assume that their average component parallel to $\vec{Q}$ is $\xi_i^{\parallel} \approx -Q/s$. We further assume that the conditional probability density $\tilde{p}_s(\vec{\xi})$ of these vectors, conditioned on their sum $\vec{z}$ reaching the ball around the origin, can be approximated by a Gaussian, which consequently has the form

$$\tilde{p}_s(\vec{\xi}) \approx \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\left|\vec{\xi} + \frac{\vec{Q}}{s}\right|^2\right). \quad (36)$$

For $\vec{z}$ to be a phenotype vector of a local maximum, all these random vectors should lie in the region $\mathcal{D}(\vec{z})$ and the remaining (unconstrained) $L - s$ vectors should lie in $\mathcal{D}(-\vec{z})$. This event
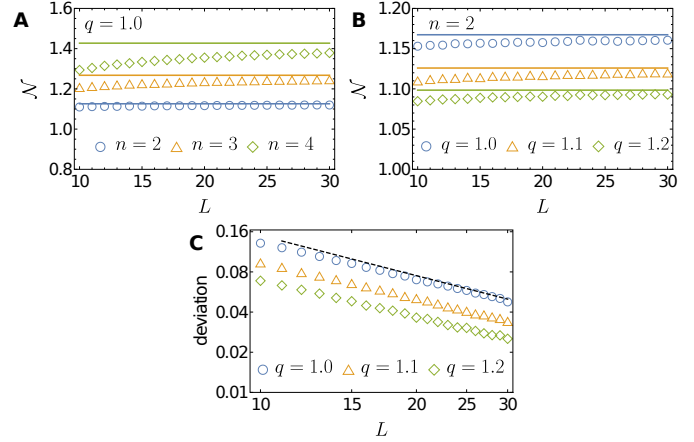
Figure 7 Comparison of simulation results (symbols) of the mean number of local maxima $\mathcal{N}$ with analytic approximations (lines) for $q > q_c$. Each symbol is the result of averaging over 2×10^6 realizations. (**A**) $\mathcal{N}$ is shown to increase with n for fixed q. (**B**) $\mathcal{N}$ is shown to decrease with q for fixed n. (**C**) Deviation of the analytic expression from the simulation results, defined as $1 - \frac{\mathcal{N}_{\text{data}}}{\mathcal{N}_{\text{theory}}}$, is depicted as a function of L on a double logarithmic scale. The phenotypic dimension for this panel is $n = 4$, where the largest deviations are observed in panel **A**. The deviation decreases inversely with L as indicated by the black dashed line with slope -1.

happens with probability

$$\left(\int_{\mathcal{D}(\vec{z})} d\vec{\xi} \tilde{p}_s(\vec{\xi})\right)^s \left(\int_{\mathcal{D}(-\vec{z})} d\vec{\xi} p(\vec{\xi})\right)^{L-s}$$

$$\approx \left\{1 - \frac{V_n}{(2\pi)^{n/2}} \exp[-Q^2/(2s^2)]\right\}^s \left(1 - \frac{V_n}{(2\pi)^{n/2}}\right)^{L-s}$$

$$\approx \exp\left[-\frac{V_n}{(2\pi)^{n/2}}\left\{s \exp[-Q^2/(2s^2)] + L - s\right\}\right]. \quad (37)$$

Thus, we can estimate the typical value of ζ as the solution of

$$\frac{V_n(\zeta)}{(2\pi)^{n/2}} \approx \left[s \exp[-Q^2/(2s^2)] + L - s\right]^{-1}, \quad (38)$$

which, combined with Equation 35, indeed gives Equation 34.

To find the asymptotic behavior of $\mathcal{N}_<$ for large L, we use Stirling's formula in Equation 33 and approximate the summation over s by an integral over $\rho \equiv s/L$. This yields

$$\mathcal{N}_< \approx \int_0^1 d\rho \frac{1}{L^{n/2} \rho^{n/2}} \frac{e^{L\Sigma(\rho)}}{\sqrt{2\pi L \rho(1-\rho)}} \frac{1}{1 - \rho + \rho e^{-\frac{q^2}{2\rho^2}}}, \quad (39)$$

where the exponent $\Sigma(\rho)$ is given by

$$\Sigma(\rho) \equiv -\rho \ln \rho - (1 - \rho) \ln(1 - \rho) - \frac{q^2}{2\rho}. \quad (40)$$

Under the condition $L \gg 1$, the remaining integral with respect to ρ can be performed by expanding $\Sigma(\rho)$ to second order around the saddle point ρ^* determined by the condition

$$0 = \left.\frac{\partial}{\partial \rho}\Sigma(\rho)\right|_{\rho=\rho^*} = \frac{q^2}{2(\rho^*)^2} - \ln\frac{\rho^*}{1 - \rho^*}. \quad (41)$$

Fisher's geometric model   9

Performing the resulting Gaussian integral with respect to ρ one finally obtains

$$\mathcal{N}_< \approx \frac{1}{L^{1+n/2}} \sqrt{\frac{1}{1+(1-\rho^*)(q/\rho^*)^2}} \frac{(\rho^*)^{-n/2}e^{L\Sigma(\rho^*)}}{1-\rho^*+\rho^*e^{-\frac{q^2}{2(\rho^*)^2}}}, \quad (42)$$

where $\rho^* = \rho^*(q)$ is the solution of Equation 41, which is the (scaled) mean number of mutations in a local maximum. We will call ρ^* the mean genotypic distance. This solution is not available in closed form, but it can be shown that $\rho^* = \frac{1}{2} + \frac{q^2}{2} + O(q^4)$ and $\Sigma(\rho^*) = \ln 2 - q^2 + O(q^4)$ for small q. Figure 6 compares Equation 42 with the mean number of local maximum obtained by numerical simulations for various q's with $n = 1$, to show an excellent agreement even for $L = 10$.

It is obvious that $\Sigma(\rho)$ will eventually be negative as q increases for any value of ρ, and this must be true also for the maximum value $\Sigma(\rho^*)$. Indeed, we found the threshold $q_c \approx 0.924\,809$, above which $\Sigma(\rho^*)$ is negative. This signals a phase transition in the landscape properties. Inspection of Equation 40 shows that the transition is driven by a competition between the abundance of genotypes with a certain number of mutations and their likelihood to bring the phenotype close to the optimum. The first two terms in the expression for $\Sigma(\rho)$ are the standard sequence entropy [see, for example, (Schmitt and Herzel 1997)] which is maximal at $\rho = 1/2$ ($s = L/2$), whereas the last term represents the statistical cost associated with "stretching" the phenotype towards to origin. With increasing q, the genotypes contributing to the formation of local maxima become increasingly atypical, in the sense that they contain more than the typical fraction $\rho = 1/2$ of mutations, and ρ^* increases. For $q > q_c$ the cost can no longer be compensated by the entropy term and $\Sigma(\rho^*)$ becomes negative. In this regime $\mathcal{N}_<$ *decreases* exponentially with L, and therefore the total number of fitness maxima $\mathcal{N}$, which by construction cannot be smaller than 1, must be dominated by the second contribution $\mathcal{N}_>$.

Regime II. We defer the detailed derivation of $\mathcal{N}_>$ to Appendix C and here only report the final result obtained in the limit $L \to \infty$, which is independent of L and reads

$$\mathcal{N}_> \approx \left(\frac{q-q_0}{q}\exp\left[\frac{1}{q/q_0-1}\right]\right)^{n-1}. \quad (43)$$

This expression is valid for $q > q_0 = \frac{1}{\sqrt{2\pi}} \approx 0.399$, but it dominates the contribution $\mathcal{N}_<$ for large L only when $q > q_c$. Figure 7 indeed shows that Equation 43 approximates the mean number of local maxima for $q > q_c$, that is, $\mathcal{N}$ converges to $\mathcal{N}_>$ for large L. This figure also shows, as is clear by Equation 43, that $\mathcal{N}$ is a increasing (decreasing) function of n (q) for fixed value of q (n). The expected number of maxima is small in absolute terms in this regime, which can be attributed to the fact that the expression inside the parentheses in Equation 43 takes the value $1.214\ldots$ at $q = q_c$ and decreases rapidly towards unity for larger q.

To understand the appearance of q_0, we refer to **Model**, where it was argued that $2q_0s$ is the maximal distance towards the origin, which can be covered by a phenotype made up of s typical mutation vectors. Correspondingly, the analysis in Appendix C shows that the main contribution to $\mathcal{R}_s^>(L)$ comes from phenotypes located at distance $z = 2s(q - q_0)$ from the origin, i.e., at distance $2sq_0$ from the wild type. The sum over s
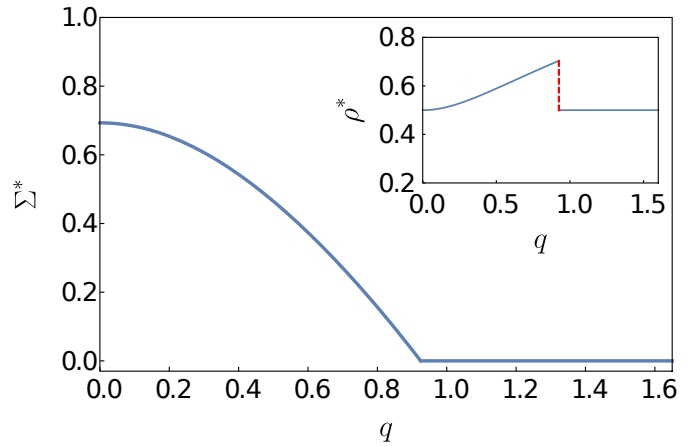

Figure 8 Plot of the genotypic complexity Σ^* as a function of the scaled phenotypic wild type distance q. Here the phenotypic dimension n is kept finite while taking the genotypic dimension L to infinity. The complexity vanishes at the phase transition point $q = q_c \approx 0.924\,809$. Inset: Plot of the mean genotypic distance ρ^* of local maxima from the wild type as a function of q. Starting from $1/2$, ρ^* increases with q for $q < q_c$ and remains at $\frac{1}{2}$ for $q > q_c$.

in Equation 33 is dominated by typical genotypes with $s = L/2$, and therefore the main contribution to $\mathcal{N}_>$ comes from phenotypes at distance $z = (q - q_0)L$ from the origin. The seeming divergence of $\mathcal{N}_>$ as $q \to q_0^+$ is an artifact of the approximation scheme, which assumes that the main contribution comes from the region where $z \sim O(L)$; clearly this assumption becomes invalid when $q \to q_0^+$. We note that for very large q and large n Equation 43 reduces to the expression $\mathcal{N}_{\mathrm{wt}}$ obtained in Equation 11 on the basis of Fisher's formula for the fraction of beneficial mutations from the wild type phenotype.

Phase transition. To sum up, the leading behavior of $\mathcal{N}$ is

$$\mathcal{N} = \begin{cases} \mathcal{N}_<, & q < q_c, \\ \mathcal{N}_>, & q \geq q_c \end{cases} \quad (44)$$

with $\mathcal{N}_<$ and $\mathcal{N}_>$ given by Equation 42 and Equation 43, respectively. Since $\mathcal{N}_<$ decreases to zero with L in a power-law fashion at $q = q_c$, the dominant contribution at this value is $\mathcal{N}_>$. At $q = q_c$ the mean genotypic distance ρ^* jumps discontinuously from $\rho^*(q_c) \approx 0.7035$ to $\rho^* = 1/2$, and the mean phenotypic distance z^* which is defined as the averaged magnitude of phenotype vectors for local maxima jumps from $z^* \approx 0$ to $z^* = (q_c - q_o)L$. The genotypic complexity Σ^* defined in Equation 13 is given by

$$\Sigma^* = \begin{cases} \Sigma(\rho^*), & q < q_c, \\ 0, & q \geq q_c, \end{cases} \quad (45)$$

where ρ^* is the solution of Equation 41, and hence vanishes continuously at $q = q_c$. These results are graphically represented in Figure 8. Recall that the value $\Sigma^* = \ln 2$ attained at $q = 0$ is the largest possible, because the total number of genotypes is $2^L = \exp(L \ln 2)$. Remarkably, these leading order results are independent of the phenotypic dimension. A dependence on n emerges at the subleading order, and it affects the number of fitness maxima in qualitatively different ways in the two phases. For $q < q_c$ the pre-exponential factor in Equation 42 is a power
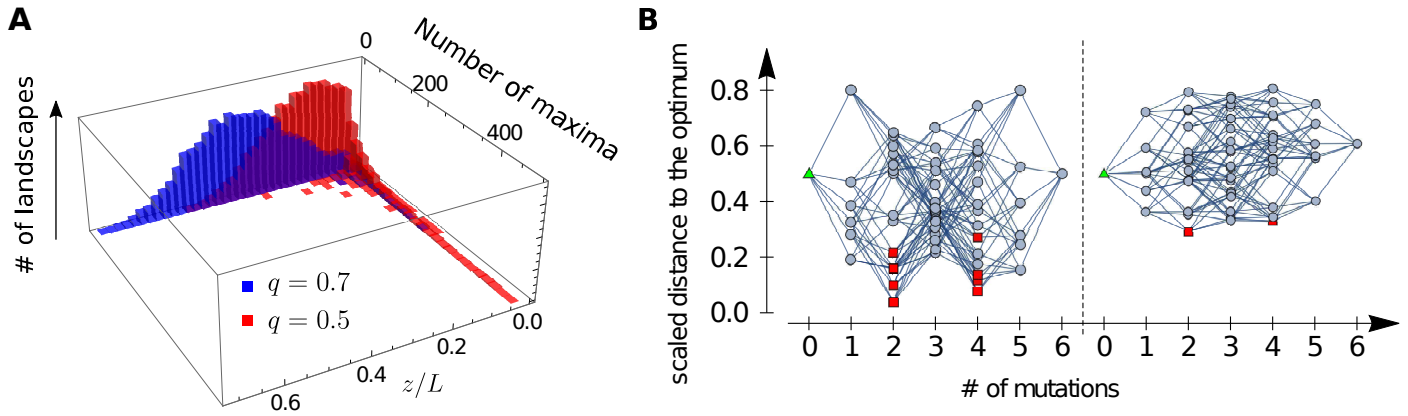
A

B

Figure 9 Coexistence of the two mechanisms I and II for $q_0 < q < q_c$. (**A**) Two dimensional histogram of the number of fitness maxima and the average phenotypic distance of the maxima to the optimum within a *single* realization. Here $L = 15$ and $n = 2$ are used and 10^4 different landscapes are randomly generated for each value of q. Only a small number of realizations have a small average distance but these contribute an exceptionally large number of fitness peaks. (**B**) Two examples of genotype-phenotype maps selected from realizations with $q = 0.5$, $L = 6$, and $n = 2$. The wild type phenotype is marked by a green triangle and local fitness maxima by red squares. When the phenotypes of the local fitness maxima are close to (far away from) the origin, the number of maxima is large (small), which corresponds to mechanism I (II).

law in L with exponent $1 + n/2$ and hence decreases with increasing n, whereas the expression in Equation 43 describing the regime $q > q_c$ increases exponentially with n.

Interpretation. The phase transition reflects a shift between two distinct mechanisms for generating genotypic complexity in FGM, which are analogous to the two origins of pairwise sign epistasis that were identified by Blanquart *et al.* (2014) and discussed above in ***Sign epistasis***. In regime I ($q < q_c$) the mutant phenotype closely approaches the origin and multiple fitness maxima are generated by overshooting the phenotypic optimum. By contrast, in regime II ($q > q_c$) the phenotypic optimum cannot be reached and the genotypic complexity arises from the local curvature of the fitness isoclines. These two situations are exemplified by the two panels of Figure 1. For the sake of brevity, in the following discussion we will refer to the two mechanisms as mechanism I and mechanism II, respectively.

The approach to the origin in regime I is a largely one-dimensional phenomenon governed by the components of the mutation vector along the direction of the wild type phenotype $\vec{Q}$, which explains why the leading order behavior of the genotypic complexity is independent of n. For $q < q_c$, the n-dependence of the pre-exponential factor in Equation 42 arises from the increasing difficulty of the random walk formed by the mutational vectors to locate the origin in high dimensions. By contrast, mechanism II operating for $q > q_c$ relies on the existence of the transverse dimensions, which is the reason why $\mathcal{N}_>$ in Equation 43 is an increasing function of n with $\mathcal{N}_> = 1$ for $n = 1$.

When $q_0 < q < q_c$, both mechanisms seem to be present simultaneously. As our analysis is restricted to the average number of local maxima, at this point we cannot decide whether both mechanisms appear in a single realization of the fitness landscape, or if one of them dominates for a given realization. To answer this question, we generated 10^4 fitness landscapes randomly for given parameter sets and identified all local maxima for each landscape. We then determined the number of local maxima and averaged the phenotypic distance of the local maxima to the optimum for each realization. This mean

distance will be denoted by $\bar{z}$ and is itself a random variable; it should not be confused with the mean phenotypic distance z^*, which is calculated by taking an average over all fitness peaks in all realizations, giving the same weight to each peak. The results are depicted as a two-dimensional histogram in Figure 9(A).

The figure shows that the marginal distribution of $\bar{z}$ displays a pronounced peak around $\bar{z}/L \approx q - q_0$, which corresponds to the behavior that is typical of mechanism I. For most realizations $\bar{z}/L$ deviates significantly from zero, and only a small number of landscapes have local maxima near $z = 0$. However, these landscapes have many more maxima than typical landscapes and therefore dominantly contribute to the mean number of maxima $\mathcal{N}$. This shows that within a single realization the two mechanisms are not operative together and only a single mechanism exists. Since most realizations exhibit mechanism II, whereas the mean number of local maxima grows exponentially as expected for mechanism I, we conclude that mechanism I occurs rarely but once it does, it generates a huge number of local maxima, which compensates the low probability of occurrence. We may thus say that both mechanisms coexist for $q_0 < q < q_c$ and q_0 can be regarded as the threshold of coexistence. Two fitness landscape realizations generated for the same value of q located in the coexistence region that exemplify the two mechanisms are shown in Figure 9(B).

Summary 2. If the dimension n of phenotypic space is much smaller than the dimension L of genotypic space, there exists a threshold q_c of the scaled wild distance q to the phenotypic optimum below which the mean number $\mathcal{N}$ of local maxima in a genotypic fitness landscape increases exponentially with L and above which it saturates to a finite value. The genotypic complexity Σ^*, which is defined as the exponential growth rate of $\mathcal{N}$ with L, is a decreasing function of q but does not depend on n. On the other hand, $\mathcal{N}$ decreases with n for $q < q_c$ yet increases with n for $q > q_c$. Figure 8 depicts Σ^* and the mean genotypic distance ρ^* as functions of q. For $q_0 < q < q_c$, where $q_0 = 1/\sqrt{2\pi}$, $\mathcal{N}$ is dominated by a small fraction of landscape realizations that display an exceptionally large number of max-
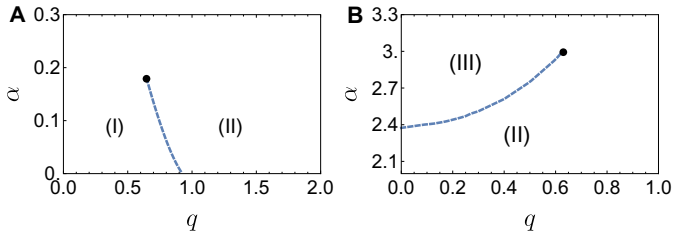
Figure 10 Phase diagrams in the parameter space (q, α). Here, $q = Q/L$ is the scaled distance of the wildtype phenotype from the origin and $\alpha = n/L$ is the ratio of phenotypic dimension to genotypic dimension. Dashed lines are phase boundaries at which the mean genotypic and phenotypic distances change discontinuously. (**A**) The phase boundary separating regimes I and II starts at $(q, \alpha) \simeq (0.925, 0)$ and continues to exist until approximately $\alpha \simeq 0.18$. (**B**) The phase boundary separating regimes II and III starts at $(q, \alpha) \simeq (0, 2.38)$ and continues to exist until approximately $q \simeq 0.62$.

ima. If the pleiotropic scaling is assumed to follow the total effects model, we need to specify how the unscaled wild type distance d in Equation 2 depends on L. Assuming that $d = d_0 L$ where d_0 is independent of n (Orr 2000), the scaled wild type distance $q = Q/L = d_0\sqrt{n}$ becomes an increasing function of n, and therefore the relation $q < q_c$ for regime I is never realized when n is sufficiently large.

Genotypic complexity in the joint limit

In the previous subsection, we have calculated the mean number of local fitness maxima $\mathcal{N}$ at a fixed phenotypic dimension n, assuming that the genotypic dimension L is much larger than n ($L \gg n$). However, in applications of FGM one often expects that both L and n are large and possibly of comparable magnitude. In this case the results derived above can be unreliable for large n, as exemplified by the fact that the subleading correction to Equation 42 is of the order of $O(L^{-1/n})$ (see Appendix C).

To obtain a reliable expression for $\mathcal{N}$ that is valid when both n and L are large, we now consider the joint limit $n, L \to \infty$ at fixed ratio $\alpha = n/L$. This will allow us to find the leading behavior of the mean number of local maxima with a correction of order $O(1/L)$. Furthermore, we will clarify the role of the phenotypic dimension in the two phases described in the previous subsection, and we will uncover a third phase that appears at large α (see Figure 10).

The number of local fitness maxima. We relegate the detailed calculation to Appendix D and directly present our final expression for the mean number of local maxima,

$$\mathcal{N} = \mathcal{C}(a^*, b^*, g^*) e^{L\Sigma^{\mathrm{red}}(a^*, b^*, g^*)} \left(1 + O\left(\frac{1}{L}\right)\right), \qquad (46)$$

where the function $\Sigma^{\mathrm{red}}(a, b, g)$ in the exponent is given by

$$\Sigma^{\mathrm{red}}(a, b, g) = -\frac{\alpha}{2} \ln\left(\frac{\alpha(\alpha + g)}{2(ac(g) + b^2)}\right) + \frac{\alpha + 2b + g}{2} - \ln 2$$
$$+ \ln\left(e^{-2c(g)}\left[\mathrm{erf}\left(\frac{\alpha + 2b}{\sqrt{2a}}\right) + 1\right] + \mathrm{erf}\left(\frac{\alpha}{\sqrt{2a}}\right) + 1\right) \qquad (47)$$
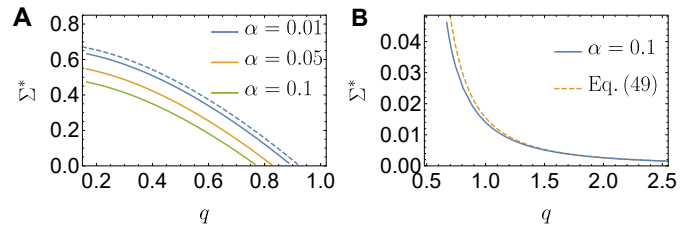
Figure 11 Convergence of the complexity to the fixed n case for small α. (**A**) The solid lines depict numerical solutions of Equation 48 for values of α belonging to regime I. The convergence to Equation 42 (dashed line) is clearly seen as $\alpha \to 0$. (**B**) The blue solid line depicts the numerical solution of Equation 48 for $\alpha = 0.1$ belonging to regime II. Except for a slight devitation detectable when q is close to q_0, Equation 49 (dashed line) remains a good approximation.

with $c(g) = \frac{\alpha^2 - g^2}{16q^2}$. As before, the starred variables a^*, b^* and g^* denote the solution of the extremum condition

$$\nabla \Sigma^{\mathrm{red}}(a, b, g)|_{(a,b,g)=(a^*,b^*,g^*)} = (0, 0, 0), \qquad (48)$$

where ∇ is the gradient with respect to the three variables (a, b, g). When several solutions of Equation 48 exist, the one giving the largest value of Σ^{red} is chosen. The prefactor $\mathcal{C}(a^*, b^*, g^*)$, which is independent of L, can be determined from Equation D17 presented in Appendix D. Even though the variables (a, b, g) lack a direct intepretation in terms of the original setting of FGM, we show in Appendix E that a^* is related to the mean phenotypic distance z^* by the equation $z^* = L\sqrt{a^*}/2$.

An immediate consequence of Equation 46 is that the number of local maxima increases exponentially in L for any value of q and α without algebraic corrections of the kind found in Equation 42. Obtaining closed form solutions of Equation 48, which ultimately determine the functional dependence of the complexity Σ^* on α and q, seems to be a formidable task. Instead, we resort to numerical methods by sweeping through the most interesting intervals, $q \in (0, 2)$ and $\alpha \in (0, 3)$. Surprisingly, we find three independent branches of solutions that correspond to distinct phases. In order to acquire a qualitative understanding of these branches, it is instructive to first focus on the small α behavior, where one expects a smooth continuation to the results of Equation 42 and Equation 43 as $\alpha \to 0$.

Small α behavior. In contrast to the fixed n case where two separate analyses were carried out for the two regimes $q < q_c$ and $q > q_c$, the present approach yields a single expression describing the genotypic complexity for arbitrary values of q and α. Consistently with the fixed n analysis, only two out of the three branches of solutions that were found in the numerical analysis exist for sufficiently small α, and they are separated by a phase transition as shown in the phase diagram in Figure 10(A). By extrapolating the behavior of Σ^* towards $\alpha \to 0$ as shown in Figure 11, we are able to identify the correct counterparts for each of the two previously found regimes.

The extrapolation is straightforward in regime II, where the replacement $n \to \alpha L$ in Equation 43 yields an exponential dependence of $\mathcal{N}$ on L with the growth rate

$$\Sigma^{*(\mathrm{II})}_{\mathrm{approx}} = \alpha \ln\left(\frac{q - q_0}{q} \exp\left[(q/q_0 - 1)^{-1}\right]\right). \qquad (49)$$
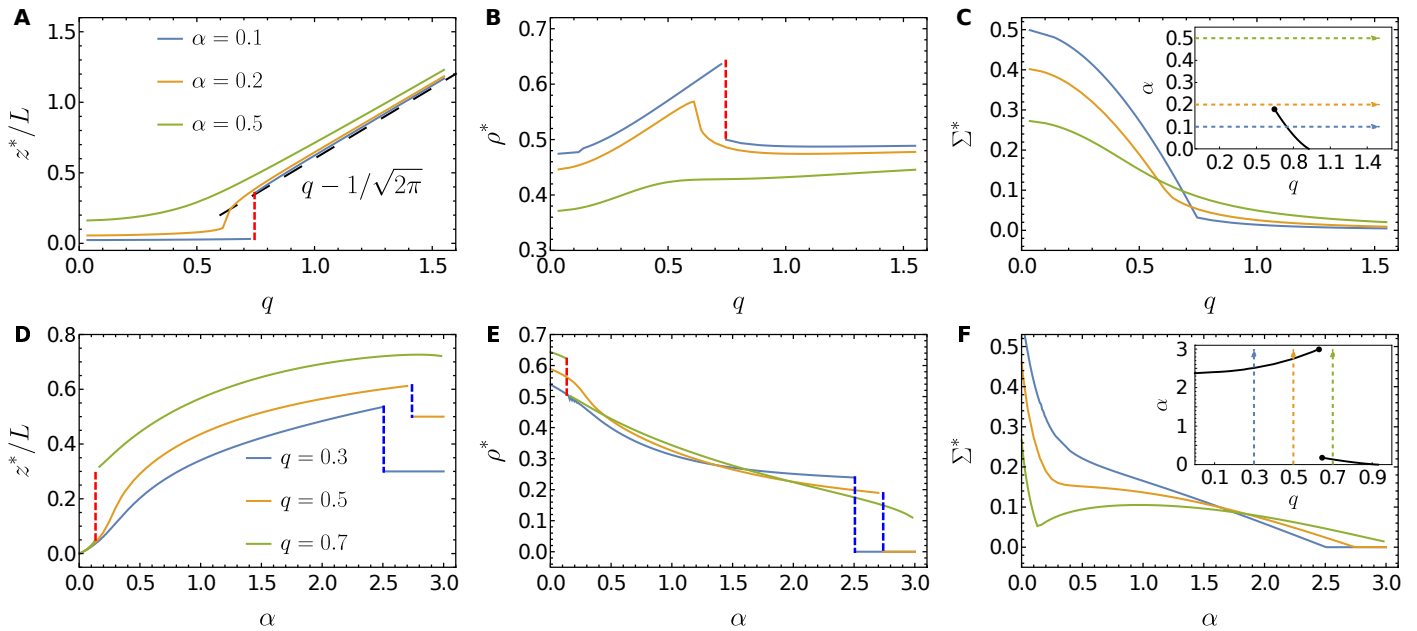
Figure 12 Plots of scaled mean phenotypic distance z^*/L (left column), mean genotypic distance ρ^* (middle column), and genotypic complexity Σ^* (right column) against q for fixed α (top row) and against α for fixed q (bottom row). The curves in the top (bottom) panels are drawn along the arrows in the inset of panel **C** (**F**). Top row (**A,B,C**): When α is small, the landscape behaves similar to the fixed n case which effectively corresponds to $\alpha = 0$. In this case z^* and ρ^* for large q are well approximated by $q - q_0$ and $1/2$, respectively. As α increases beyond the transition line, the first order transition visualized by the red dashed lines disappears and all quantities change smoothly with q. Bottom row (**D,E,F**): As α increases for small q, another phase transition with discontinuities in z^* and ρ^* (blue dashed lines) signals the appearance of regime III. The genotypic maxima in regime III are located very close to the wild type position, $z^*/L \simeq q$ and $\rho^* \simeq 0$. This transition ceases to exist when q exceeds approximately 0.62. Note that the dependence of Σ^* on α is non-monotonic for $q = 0.7$ (panel **F**).

This crude approximation turns out to be remarkably accurate even at $\alpha = 0.1$, as illustrated in Figure 11(B). By contrast, in regime I the naive replacement of n by αL in Equation 42 yields an expression that vanishes faster than exponential in L, as $\exp[-(\alpha/2)L\ln L]$. This reflects the fact that the mean phenotypic distance z^* moves away from the origin for any $\alpha > 0$ and hence the complexity cannot be derived only by inspecting Equation 26 around $z = 0$ (see Figure 12 (A, D)). At the same time the mean genotypic distance ρ^* decreases with increasing α and eventually falls below the value $\rho^* = 1/2$ favored by the sequence entropy (Figure 12 (B, E)).

Both trends can be attributed to the increasing role of the perpendicular mutational displacements that make up the second term on the right hand side of Equation 5. Under the scaling of the joint limit, this term is of order $\rho L(n-1) \approx \rho\alpha L^2$ and hence comparable to the first term originating from the parallel displacements. The perpendicular displacements always increase the phenotypic distance to the origin, and they are present even when $q = 0$. The additional cost to reduce the perpendicular contribution results in a smaller value of Σ^* compared to the case of fixed n. Moreover, whereas the parallel contribution is minimized (for $q > q_0$) by making ρ as large as possible, the reduction of the perpendicular displacements requires small ρ.

In the fixed n analysis the number of fitness maxima was found to decrease (increase) with n in regime I (II) and this tendency is recovered from the joint limit case when α is not too large (Figure 12 (C)). Because of these opposing trends of Σ^* in the two regimes, the location of the phase transition separating them is expected to decrease with increasing α, as can be seen in Figure 10(A). If one ignores the contribution from the

perpendicular displacements, the phenotypic position of the fitness maxima is expected to jump from $z^* = 0$ to $z^* = q_c - q_0$ at the transition, and thus the jump size should decrease as q_c decreases. This observation suggests that the two branches should merge into one when q_c reaches q_0. With the additional contribution of perpendicular dimensions, we numerically found that this critical end point at which the phases I and II merge occurs even earlier, at $\alpha \simeq 0.18$ and $q \simeq 0.62 > q_0$ (Figure 10). For α larger than 0.18, ρ^* does not show any discontinuity for any q as long as the parameters are in the regime II.

Large α behavior and regime III. In order to develop some intuition about the FGM fitness landscape in the regime where $\alpha = n/L \gg 1$, we revisit the results obtained in ***Sign epistasis***, where pairs of mutations were considered. Two conclusions can be drawn about the typical shape of these small genotypic landscapes (of size $L = 2$) in the limit $n \to \infty$. First, the probability that the wild type is a genotypic maximum tends to unity according to Equation 10. Second, the joint distribution given in Equation 15 enforces additivity of mutational effects for large n, and correspondingly the probability for sign epistasis vanishes. Thus for large n the two-dimensional genotypic landscape becomes smooth with a single maximum located at the wild type. Assuming that this picture holds more generally whenever the limit $n \to \infty$ is taken at finite L, we expect the following asymptotic behaviors of the quantifiers of genotypic complexity for large α: (i) $\mathcal{N} \to 1$, $\Sigma^* \to 0$ (unique genotypic optimum); (ii) $z^*/L \to q$, $\rho^* \to 0$ (location of the maximum at the wild type phenotype and genotype).

This expectation is largely borne out by the numerical results

shown in the bottom panels of Figure 12. However, depending on the value of q, the approach to the limit of a smooth landscape can be either continuous (for large q) or display characteristic jumps indicated by the blue dashed lines in Figure 12 (D) and (E). These jumps as well as the discontinuity in the slope of Σ^* as a function of α in Figure 12 (F) are hallmarks of the phase transition to the new regime III, which is represented by the dashed line in Figure 10 (B).

Fortunately, the solution of Equation 48 describing the new phase can be obtained analytically from Equation 47 or Equation F3 as a series expansion. The derivation presented in Appendix G yields

$$a^* = 4q^2 - \left(\frac{16\sqrt{\frac{2}{\pi}}q^3}{\alpha^2} + O(q^4/\alpha^3) \right) \epsilon + O(\epsilon^2),$$

$$b^* = -\alpha + \frac{\alpha\epsilon}{\sqrt{2\pi}q} + O(\epsilon^2),$$

$$g^* = \alpha + O(\epsilon^2), \quad \text{and} \quad \rho^* = \frac{\sqrt{\frac{2}{\pi}}q\epsilon}{\alpha} + O(\epsilon^2), \quad (50)$$

where the expansion parameter $\epsilon = e^{-\frac{\alpha^2}{8q^2}}$ decays rapidly with increasing α/q. The corresponding genotypic complexity can also be evaluated in a series expansion,

$$\Sigma^{(\mathrm{III})}(a^*, b^*, c^*) = \frac{\alpha\epsilon^2}{4\pi q^2} + O(\epsilon^3), \quad (51)$$

which shows that Σ^* is positive but vanishingly small in this regime. We note that using Equation E5, the expression for a^* in Equation 50 amounts to

$$\frac{z^*}{L} \simeq q - \frac{2\sqrt{\frac{2}{\pi}}q^2}{\alpha^2}\epsilon, \quad (52)$$

implying that the small number of local maxima that exist in this phase are located very close to the wild type phenotype.

To first order in ϵ, the results for ρ^* and z^* in Equation 50 and Equation 52 can be easily derived from the idea that mutational effects become approximately additive for large α, thus providing further support for this assumption. If mutational effects are strictly additive, the probability for a genotype contaning s mutations to be a local fitness maximum is given by

$$\mathcal{R}_s^{\mathrm{add}} = P_b^s(1 - P_b)^{L-s}, \quad (53)$$

where P_b is the probability for a mutation to be beneficial. Equation 53 expresses the condition that reverting any one of the s mutations contained in the genotype as well as adding one of the unused $L - s$ mutations should lower the fitness. Using Fisher's Equation 1, the probability for a beneficial mutation is $P_b \approx \sqrt{\frac{2}{\pi}}(q/\alpha)\epsilon$ for large α. Thus to linear order in ϵ or P_b, the expected number of mutations contributing to such a genotype is $LP_b = L\rho^* = L\sqrt{\frac{2}{\pi}}(q/\alpha)\epsilon$, which is consistent with Equation 50.

The phenotypic location of a local maximum deviates from $\vec{Q}$ in those rare instances where one of the mutations from the wild type is beneficial, which happens with probability P_b. To estimate the corresponding shift in z^*, we refer to the results of subsection **Sign epistasis**, where it was shown that the squared phenotypic displacement R_1 defined in Equation 14 has a Gaussian distribution with mean 1 and variance $1/x^2 = 4q^2/\alpha^2$ for
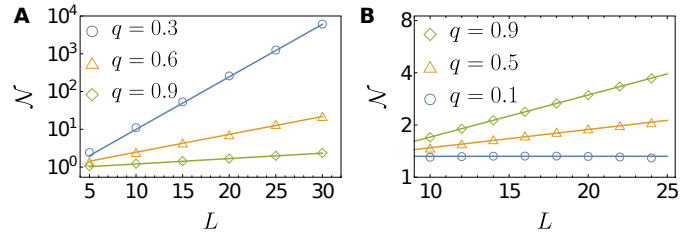
Figure 13 Semi-logarithmic plots of the mean number of local maxima $\mathcal{N}$ vs. the genotypic dimension L for (**A**) $\alpha = 0.2$ and (**B**) $\alpha = 2.5$ and for various values of q. Each symbol represents the average over 10^5 randomly generated landscapes, and lines depict the analytic approximation of Equation D17. The approximation is good even for moderate L.

large n. Using this, it is straightforward to show that the expected value of R_1 conditioned on the mutation to be beneficial ($R_1 < 0$) is $\bar{R}_1 = -4q^2/\alpha^2$ to leading order. Multiplying this by the expected number of mutations LP_b we obtain the relation

$$L\rho^*\bar{R}_1 \approx \frac{(z^*)^2 - Q^2}{n}, \quad (54)$$

which yields the same leading behavior for z^*/L as in Equation 52.

As previously observed for the transition between regimes I and II, the phase boundary separating regimes II and III terminates at a point where the two solutions defining the regimes merge (Figure 10 (B)). Beyond this point the jumps in z^* and ρ^* seen in Figure 12 (D) and (E) disappear and all quantities approach smoothly to their asymptotic values. A surprising feature of the large α behavior that persists also for larger q is that the complexity becomes an *increasing* function of q when $\alpha > 1.7$ (Figure 12 (F)). In Figure 13 we verify this behavior using direct simulations of FGM. These simulations also show that the predictions based on Equation 47 are remarkably accurate already for moderate values of L and n.

Summary 3. When the dimension n of the phenotypic trait space and the dimension L of the genotypic space are large and comparable, the genotypic complexity Σ^* is always nonzero and depends on the ratios $\alpha = n/L$ and $q = Q/L$. There are three regimes where the behavior of the genotypic complexity and the mean genotypic distance ρ^* (the average number of mutations in a local maximum divided by L) are qualitatively different. In regime I which is roughly characterized by small q and small α, there are many local maxima in the region located far away from the wild type but close to the phenotypic optimum, and the fitness landscape is quite rugged. In regime II which is roughly characterized by large q and small α, there is an appreciable number of local maxima, though smaller than in regime I, and typically half of the L mutations contribute to the corresponding genotypes. In regime III which is roughly characterized by large α, the genetic complexity is very small, though nonzero. Also ρ^* is close to zero, which means that the wild type has a high probability to be the global fitness maximum. An overview of the three regimes is found in Table 2.

Discussion

Fisher's geometric model (FGM) provides a simple yet generic scenario for the emergence of complex epistatic interactions from a nonlinear mapping of an additive, mul-

Table 2 Characteristics of the three regimes in the joint limit

Regime	Condition	Σ^*	ρ^*	landscape
I	$q \ll 1, \alpha \ll 1$	> 0	$> 1/2$	rugged
II	$q \gg 1, \alpha \ll 1$	$\simeq 0$	$\approx 1/2$	intermediate
III	$\alpha \gg 1$	$\simeq 0$	$\simeq 0$	almost smooth

tidimensional phenotype onto fitness. Its role in the theory of adaptation may be aptly described as that of a "proof of concept model" (Servedio *et al.* 2014), and as such it is widely used in fundamental theoretical studies (Blanquart *et al.* 2014; Chevin *et al.* 2010; Fraïsse *et al.* 2016; Gros *et al.* 2009; Martin 2014; Moura de Sousa *et al.* 2016) as well as for the parametrization and interpretation of empirical data (Bank *et al.* 2014; Blanquart and Bataillon 2016; Martin *et al.* 2007; Perfeito *et al.* 2014; Schoustra *et al.* 2016; Velenich and Gore 2013; Weinreich and Knies 2013). Rather than tracing the mutational effects and their interactions to the underlying molecular basis, the model aims at identifying robust features of the adaptive process that can be expected to be shared by large classes of organisms.

To give an example of such a feature that is of central importance in the present context, it was pointed out by Blanquart *et al.* (2014) that pairwise sign epistasis is generated in FGM through two distinct mechanisms. In one case the mutational displacements overshoot the phenotypic optimum, whereas in the other case the displacements are directed approximately perpendicular to the direction of the optimum and sign epistasis arises because the fitness isoclines are curved. The first mechanism is obviously operative also in a one-dimensional phenotype space, but in the second case [termed antagonistic pleiotropy by Blanquart *et al.* (2014)] at least two phenotypic dimensions are required. Interestingly, both mechanisms have been invoked in empirical studies where a nonlinear phenotype-fitness map was used to model epistatic interactions between multiple mutations. In one study, Rokyta *et al.* (2011) explained the pairwise epistatic interactions between 9 beneficial mutations in the ssDNA bacteriophage ID11 by assuming that fitness is a single-peaked nonlinear function of a one-dimensional additive phenotype. In the second study the genotypic fitness landscapes based on all combinations of two groups of four antibiotic resistance mutations in the enzyme β-lactamase were parametrized by a nonlinear function mapping a two-dimensional phenotype to resistance (Schenk *et al.* 2013). The fitted function was in fact monotonic and did not possess a phenotypic optimum, which makes it clear that the epistatic interactions arose solely from antagonistic pleiotropy in this case.

In this work we have shown that the two mechanisms described by Blanquart *et al.* (2014) lead to distinct regimes or phases in the parameter space of FGM, where the genotypic fitness landscapes display qualitatively different properties (Figure 10(A)). When the phenotypic dimension n is much smaller than the genotypic dimension L, the two regimes are separated by a sharp phase transition where the average number and location of genotypic fitness maxima changes abruptly as the distance q of the wild type phenotype from the optimum is varied. In regime I ($q < q_c$) the phenotypic optimum is reachable at least by some combinations of mutational displacements. Overshooting of the optimum is therefore possible and sign epistasis is strong, leading to rugged genotypic landscapes with a large
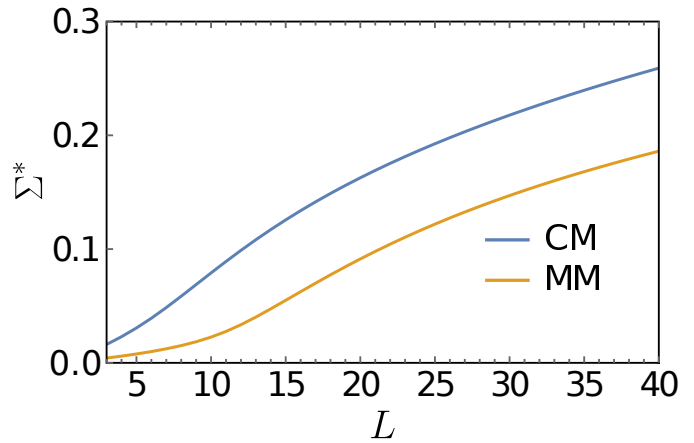

Figure 14 The logarithm of the number of local fitness maxima divided by the number of loci L is shown as a function of L for FGM with the parameter values $n = 19.3, Q = 6.89$ and $n = 34.8, Q = 9.81$ obtained by Schoustra *et al.* (2016) for the fungus *A. nidulans* growing in complete (CM) and minimal medium (MM), respectively. For the evaluation of $\mathcal{N}$ Equation 47 was used.

number of local fitness maxima that grows exponentially with L. By contrast, in regime II ($q > q_c$) only antagonistic pleiotropy is operative and the number of fitness maxima is much smaller. More precisely, for finite n the number tends to a finite limit for $L \to \infty$, but the limiting value is an exponentially growing function of n.

An important consequence of our results is that the dependence of the fitness landscape ruggedness on the phenotypic dimension n is remarkably complicated. For $n \ll L$, landscapes become less rugged with increasing n in regime I ($q < q_c$) but display increasing ruggedness in regime II ($q > q_c$). When $n \gg L$ the ruggedness decreases with n for all q and the landscapes become approximately additive (regime III). In particular, the probability of sign epistasis vanishes algebraically with n in this regime. Thus n cannot in general be regarded as a measure of "phenotypic complexity", as a larger value of n does not imply that the corresponding fitness landscape is more complex.

This observation is relevant for the interpretation of experiments where the parameters of FGM are estimated from data. In recent work, FGM was used to analyze data on pairwise epistasis between beneficial mutations in the filamentous fungus *Aspergillus nidulans* growing in two different media (Schoustra *et al.* 2016). The estimates obtained for the phenotypic dimension and the distance of the wildtype phenotype from the optimum were $n = 19.3, Q = 6.89$ in complete medium and $n = 34.8, Q = 9.81$ in minimal medium, which, surprisingly, may seem to suggest a higher "phenotypic complexity" in the minimal medium. Using the results derived in this paper we can translate the estimated parameter values into the average number of maxima that a genotypic fitness landscape of a given dimension L would have. As can be seen in Figure 14, with respect to this measure the fitness landscape of the fungus growing in complete medium is actually more rugged. This is consistent with experiments using *Escherichia coli*, which found a greater heterogeneity of fitness trajectories in complete medium (Rozen *et al.* 2008), and indicates that the complete medium allows for a greater diversity of paths to

adaptation than the minimal medium.

We hope that the results presented here will promote the use of FGM as part of the toolbox of probabilistic models that are currently available for the analysis of empirical fitness landscapes (Bank *et al.* 2016; Blanquart and Bataillon 2016; de Visser and Krug 2014; Hayashi *et al.* 2006; Neidhart *et al.* 2014; Szendro *et al.* 2013). Compared to purely genotype-based models such as the NK and Rough-Mount-Fuji (RMF) models, FGM is arguably more realistic in that it introduces an explicit phenotypic layer mediating between genotypes and fitness (Martin 2014). Somewhat similar to the RMF model, the fitness landscapes of FGM are anisotropic and display a systematic change of properties as a function of the distance to the optimal phenotype (FGM) or the reference sequence (RMF), respectively (Neidhart *et al.* 2014). The idea that fitness landscape ruggedness increases systematically and possibly abruptly when approaching the optimum has been proposed previously in the context of *in vitro* evolution of proteins (Hayashi *et al.* 2006). If this is indeed a generic pattern, it may have broader implications. For example, de Visser *et al.* (2009) showed that the evolutionary benefits of recombination are severely limited by the presence of multiple peaks. If such peaks are rare far away from the optimum, the benefits of recombination would be most pronounced for particularly maladapted populations.

A recent investigation of 26 published empirical fitness landscapes using Approximate Bayesian Computation concluded that FGM could account for the full structure of the landscapes only in a minority of cases (Blanquart and Bataillon 2016). One of the features of the empirical landscapes that prevented a close fit to FGM was the occurrence of sign epistasis far away from the phenotypic optimum. Our analysis confirms that this is an unlikely event in FGM, and precisely quantifies the corresponding probability through Equation 16 and Equation 17. Blanquart and Bataillon (2016) also found that the phenotypic dimension is particularly difficult to infer from realizations of genotypic fitness landscapes, which matches our observation that the structure of the landscape depends only weakly on n when $n \ll L$. We expect that our results will help to further clarify which features of an empirical fitness landscape make it more or less amenable to a phenotypic description in terms of FGM or some generalization thereof.

We conclude by mentioning some open questions that should be addressed in future theoretical work on FGM. First, a significant limitation of our results lies in their restriction to the *average* number of local fitness maxima. The number of maxima induced by a given realization of mutational displacements is a random variable, and unless the distribution of this variable is well concentrated, the average value may not reflect the typical behavior. The large fluctuations between different realizations of fitness landscapes generated by FGM were noticed already by Blanquart *et al.* (2014) on the basis of small-scale simulations, and they clearly contribute to the difficulty of inferring the parameters of FGM from individual realizations that was reported by Blanquart and Bataillon (2016). In the light of our analysis this pronounced heterogeneity can be attributed to the existence of multiple phases in the model, and it is exemplified by the simulation results in Figure 9. To quantitatively characterize the fluctuations between different realizations a better understanding of the distribution of the number of fitness maxima and its higher moments is required.

Second, the consequences of relaxing some of the assumptions underlying the formulation of FGM used in this work

should be explored. The level of pleiotropy can be reduced by restricting the effects of mutational displacements to a subset of traits (Chevin *et al.* 2010; Moura de Sousa *et al.* 2016), and it would be interesting to see how this affects the ruggedness of the fitness landscape. However, the most critical and empirically poorly motivated assumption of FGM is clearly the absence of epistatic interactions on the level of phenotypes. It would therefore be important to understand how robust the results presented here are with respect to some level of phenotypic epistasis, which should ideally arise from a realistic model of phenotypic networks (Martin 2014).

Third, a natural extension of the present study is to consider multi-allelic genetic sequences. An immediate generalization keeping the additivity of mutational effects on the level phenotypes is to consider the following genotype-phenotype map

$$\vec{z}(\tau) = \vec{Q} + \sum_{i=1}^{L} \sum_{k=1}^{A} \tau_{ik} \vec{\xi}_{ik}, \qquad (55)$$

where A is the size of the alphabet from which the sequence elements are drawn (e.g., $A = 4$ for DNA or RNA and $A = 20$ for proteins), $\tau_{ik} = 1$ (0) if the allele at site i is (is not) k, and the $\vec{\xi}_{ik}$ are uncorrelated random vectors. Clearly, our results for pairwise epistasis remain the same for this generalized model because they only concern mutations at different sites. However, the condition for a local fitness maximum now involves mutations to different alleles at the same site, which may lead to a non-trivial dependence on A. On the basis of a recent study of evolutionary accessibility in multi-allelic sequence spaces one may expect the fitness landscapes to become less rugged with increasing A (Zagorski *et al.* 2016), but this conjecture would have to be corroborated by a detailed analysis.

Finally, whereas the present work focused on the structure of the fitness landscapes induced by FGM, it is of obvious importance to understand how the adaptative process actually proceeds on such a landscape (Orr 2005). A simple framework that allows to address this question is provided by adaptive walks following Gillespie's strong selection/weak mutation dynamics (Gillespie 1983, 1984; Orr 2002). In a pioneering study, Orr (1998) considered adaptive walks in FGM assuming that the number L of possible mutations is unlimited. In this setting, any population not located precisely at the phenotypic optimum has a nonzero probability of generating another beneficial mutation and the adaptive walker never stops; see Park and Krug (2008) for a related analysis of adaptation in the house-of-cards landscape. For finite but large L, an interesting question concerns the number of steps until the population finds a local fitness maximum when the adaptive dynamics is random (Kauffman and Levin 1987; Park *et al.* 2015; Park and Krug 2016) or greedy (Orr 2003; Park *et al.* 2016). This problem is currently under investigation.

Acknowledgments

Literature Cited

Bank, C., R. T. Hietpas, A. Wong, D. N. Bolon, and J. D. Jensen, 2014 A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: Uncovering the potential for adaptive walks in challenging environments. Genetics **196**: 841–852.

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

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

Blanquart, F. and T. Bataillon, 2016 Epistasis and the structure of fitness landscapes: are experimental fitness landscapes compatible with Fisher's model? Genetics **203**: 847–862.

Bray, A. J. and M. A. Moore, 1980 Metastable states in spin glasses. J. Phys. C: Solid State Phys. **13**: L469–L476.

Chevin, L.-M., G. Martin, and T. Lenormand, 2010 Fisher's model and the genomics of adaptation: Restricted pleiotropy, heterogeneous mutation, and parallel evolution. Evolution **64**: 3213–3231.

Crona, K., D. Greene, and M. Barlow, 2013 The peaks and geometry of fitness landscapes. J. Theor. Biol. **317**: 1–10.

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

de Visser, J. A. G. M., S.-C. Park, and J. Krug, 2009 Exploring the effect of sex on empirical fitness landscapes. Am. Nat. **174**: S15–S30.

Durrett, R. and V. Limic, 2003 Rigorous results for the NK model. Ann. Prob. **31**: 1713–1753.

Evans, S. N. and D. Steinsaltz, 2002 Estimating some features of NK fitness landscapes. Ann. Appl. Probab. **12**: 1299–1321.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fraïsse, C., P. A. Gunnarsson, D. Roze, N. Bierne, and J. J. Welch, 2016 The genetics of speciation: Insights from Fisher's geometric model. Evolution **70**: 1450–1464.

Franke, J., A. Klözer, J. A. G. M. de Visser, and J. Krug, 2011 Evolutionary accessibility of mutational pathways. PLoS Comp. Biol. **7**: e1002134.

Gillespie, J. H., 1983 A simple stochastic gene substitution model. Theor. Popul. Biol. **23**: 202–215.

Gillespie, J. H., 1984 Molecular evolution over the mutational landscape. Evolution **38**: 1116–1129.

Gros, P.-A., H. Le Nagard, and O. Tenaillon, 2009 The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. Genetics **182**: 277–293.

Hayashi, Y., T. Aita, H. Toyota, Y. Husimi, I. Urabe, and T. Yomo, 2006 Experimental rugged fitness landscape in protein sequence space. PLoS ONE **1**: e96.

Hermisson, J. and A. P. McGregor, 2008 Pleiotropic scaling and QTL data. Nature **456**: E3.

Kauffman, S. and S. Levin, 1987 Towards a general theory of adaptive walks on rugged landscapes. J. Theor. Biol. **128**: 11–45.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Lande, R., 1980 The genetic covariance between characters maintained by pleiotropic mutations. Genetics **94**: 203–215.

Limic, V. and R. Pemantle, 2004 More rigorous results on the Kauffman-Levin model of evolution. Ann. Prob. **32**: 2149–2178.

Martin, G., 2014 Fisher's geometric model emerges as a property of complex integrated phenotypic networks. Genetics **197**: 237–255.

Martin, G., S. F. Elena, and T. Lenormand, 2007 Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nat. Gen. **39**: 555–560.

Moura de Sousa, J. A., J. Alpendrinha, P. R. A. Campos, and I. Gordo, 2016 Competition and fixation of cohorts of adaptive mutations under Fisher geometrical model. PeerJ **4**: e2256.

Neidhart, J., I. G. Szendro, and J. Krug, 2014 Adaptation in tunably rugged fitness landscapes: The Rough Mount Fuji Model. Genetics **198**: 699–721.

Orr, H. A., 1998 The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. Evolution **52**: 935–949.

Orr, H. A., 2000 Adaptation and the cost of complexity. Evolution **54**: 13–20.

Orr, H. A., 2002 The population genetics of adaptation: the adaptation of DNA sequences. Evolution **56**: 1317–1330.

Orr, H. A., 2003 A minimum on the mean number of steps taken in adaptive walks. J. Theor. Biol. **220**: 241–247.

Orr, H. A., 2005 The genetic theory of adaptation: A brief history. Nat. Rev. Genet. **6**: 119–127.

Park, S.-C. and J. Krug, 2008 Evolution in random fitness landscapes: the infinite sites model. J. Stat. Mech.:Theory Exp. p. P04014.

Park, S.-C. and J. Krug, 2016 δ-exceedance records and random adaptive walks. J. Phys. A: Math. Theor. **49**: 315601.

Park, S.-C., J. Neidhart, and J. Krug, 2016 Greedy adaptive walks on a correlated fitness landscape. J. Theor. Biol. **397**: 89–102.

Park, S.-C., I. G. Szendro, J. Neidhart, and J. Krug, 2015 Phase transition in random adaptive walks on correlated fitness landscapes. Phys. Rev. E **91**: 042707.

Perfeito, L., A. Sousa, T. Bataillon, and I. Gordo, 2014 Rates of fitness decline and rebound suggest pervasive epistasis. Evolution **68**: 150–162.

Poelwijk, F. J., D. J. Kiviet, D. M. Weinreich, and S. J. Tans, 2007 Empirical fitness landscapes reveal accessible evolutionary paths. Nature **445**: 383–386.

Poelwijk, F. J., S. Tănase-Nicola, D. J. Kiviet, and S. J. Tans, 2011 Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. J. Theor. Biol. **272**: 141–144.

Ram, Y. and L. Hadany, 2015 The probability of improvement in Fisher's geometric model: A probabilistic approach. Theor. Popul. Biol. **99**: 1–6.

Rokyta, D., P. Joyce, S. Caudle, C. Mille, C. Beisel, and H. Wichman, 2011 Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. PLoS Genet. **7**: e1002075.

Rozen, D. E., H. M. G. J. L., H. Andreas, and de Visser J. Arjan G. M., 2008 Heterogeneous Adaptive Trajectories of Small Populations on Complex Fitness Landscapes. PLoS ONE **3**: e1715.

Schenk, M. F., I. G. Szendro, M. L. Salverda, J. Krug, and J. A. G. de Visser, 2013 Patterns of epistasis between beneficial mu-

tations in an antibiotic resistance gene. Mol. Biol. Evol. **30**: 1779–1787.

Schimitt, A. O. and H. Herzel, 1997 Estimating the entropy of DNA sequences. J. Theor. Biol. **188**: 369–377.

Schoustra, S., S. Hwang, J. Krug, and J. A. G. de Visser, 2016 Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. Proc. R. Soc. Lond. Ser. B **283**: 20161376.

Servedio, M. R., Y. Brandvain, S. Dhole, C. L. Fitzpatrick, E. E. Goldberg, C. A. Stern, J. V. Cleve, and D. J. Yeh, 2014 Not just a theory-the utility of mathematical models in evolutionary biology. PLoS Biol. **12**: e1002017.

Szendro, I. G., M. F. Schenk, J. Franke, J. Krug, and J. A. G. de Visser, 2013 Quantitative analyses of empirical fitness landscapes. J. Stat. Mech.:Theory Exp. p. P01005.

Tanaka, F. and S. F. Edwards, 1980 Analytic theory of the ground state properties of a spin glass. I. Ising spin glass. J. Phys. F: Met. Phys. **10**: 2769.

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

Velenich, A. and J. Gore, 2013 The strength of genetic interactions scales weakly with mutational effects. Genome Biology **14**: R76.

Wagner, G. P., J. P. Kenney-Hunt, M. Pavlicev, J. R. Peck, D. Waxman, and J. M. Cheverud, 2008 Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature **452**: 470–472.

Waxman, D. and J. J. Welch, 2005 Fisher's microscope and Haldane's ellipse. Am. Nat. **166**: 447–457.

Weinberger, E. D., 1991 Local properties of Kauffman's N-k model: A tunably rugged energy landscape. Phys. Rev. A **44**: 6399–6413.

Weinreich, D. M. and J. L. Knies, 2013 Fisher's geometric model of adaptation meets the functional synthesis: Data on pairwise epistasis for fitness yields insights into the shape and size of phenotype space. Evolution **67**: 2957–2972.

Weinreich, D. M., Y. Lan, C. S. Wylie, and R. B. Heckendorn, 2013 Should evolutionary geneticists worry about higher-order epistasis? Curr. Opin. Genet. Dev. **23**: 700–707.

Weinreich, D. M., R. A. Watson, and L. Chao, 2005 Sign epistasis and genetic constraint on evolutionary trajectories. Evolution **59**: 1165–1174.

Zagorski, M., Z. Burda, and B. Waclaw, 2016 Beyond the Hypercube: Evolutionary Accessibility of Fitness Landscapes with Realistic Mutational Networks. PLoS Comp. Biol. **12**: e1005218.

Appendix A: Derivation of the joint probability density $\mathcal{P}(R_1, R_2, R)$

For the purpose of this calculation it will turn out to be convenient to locate the wild type phenotype on the diagonal of the trait space, i.e. to set $\vec{Q} = \frac{Q}{\sqrt{n}}(1, 1, 1, \dots,)$. The probability density $\mathcal{P}(R_1, R_2, R)$ can then be formally defined as

$$\mathcal{P}(R_1, R_2, R) = n^3 \left\langle \delta\left(nR_1 - \sum_{i=1}^n \left[(\xi_i + Q_i)^2 - Q_i^2\right]\right) \delta\left(nR_2 - \sum_{i=1}^n \left[(\eta_i + Q_i)^2 - Q_i^2\right]\right) \delta\left(nR - \sum_{i=1}^n \left[(\xi_i + \eta_i + Q_i)^2 - Q_i^2\right]\right) \right\rangle_{\vec{\xi}, \vec{\eta}},$$
(A1)

where $Q_i = Q/\sqrt{n}$ and $\langle \cdots \rangle_{\vec{\xi}, \vec{\eta}}$ stands for the average over the distribution of $\vec{\xi}$ and $\vec{\eta}$. Using the integral representation of the delta function, we can write

$$\mathcal{P}(R_1, R_2, R) = \frac{n^3}{(2\pi)^3} \int d\vec{k} e^{ik_1 nR_1 + ik_2 nR_2 + iknR} \prod_i \left\langle e^{-ik_1(\xi_i + Q_i)^2 - ik_2(\eta_i + Q_i)^2 - ik(\xi_i + \eta_i + Q_i)^2 + iQ_i^2(k_1 + k_2 + k_3)} \right\rangle_{\vec{\xi}, \vec{\eta}},$$
(A2)

where $d\vec{R}$ and $d\vec{k}$ stand for $dR_1 dR_2 dR$ and $dk_1 dk_2 dk$, respectively and we factorized the average by taking into account that the ξ_i's and η_i's are all independent and identically distributed. The average in Equation A2 is readily calculated as

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\eta \exp\left[-\frac{\eta^2}{2} - \frac{\xi^2}{2} - ik_1\xi^2 - ik_2\eta^2 - ik(\xi + \eta)^2 - 2ik_1\frac{Q}{\sqrt{n}}\xi - 2ik_2\frac{Q}{\sqrt{n}}\eta - 2ik\frac{Q}{\sqrt{n}}(\xi + \eta)\right]$$

$$= \sqrt{\frac{1}{(1 + 2ik_1)(1 + 2ik_2) - 4k(k_1 + k_2 - i)}} \exp\left[\frac{2Q^2}{n} \frac{2i(k + k_1)(k + k_2)(k_1 + k_2 - i) + (k_1 - k_2)^2}{4k(k_1 + k_2 - i) + (2k_1 - i)(2k_2 - i)}\right],$$
(A3)

which gives

$$\mathcal{P}(R_1, R_2, R) = \frac{n^3}{(2\pi)^3} \int d\vec{k} \frac{e^{ik_1 nR_1 + ik_2 nR_2 + iknR}}{[(1 + 2ik_1)(1 + 2ik_2) - 4k(k_1 + k_2 - i)]^{n/2}} \exp\left[\frac{n^2}{2x^2} \frac{2i(k + k_1)(k + k_2)(k_1 + k_2 - i) + (k_1 - k_2)^2}{4k(k_1 + k_2 - i) + (2k_1 - i)(2k_2 - i)}\right].$$
(A4)

In the limit $n \to \infty$, the integral is dominated by contributions from the vicinity of the extremum of the exponent, which can be algebraically determined to be $k = k_1 = k_2 = 0$. By expanding the argument of the exponential function up to the second order around this point and performing the Gaussian integral, we obtain

$$\mathcal{P}(R_1, R_2, R) \approx \frac{n^3}{(2\pi)^3} \int d\vec{k} \exp\left[-\frac{n^2}{2x^2}\left\{(k_1 + k_2 + k)^2 - 2k_1 k_2\right\}\right.$$

$$\left. - n\left\{4k^2 + k_1^2 + k_2^2 - ik_1(R_1 - 1) - ik_2(R_2 - 1) + k(2i + 2k_1 + 2k_2 - iR)\right\}\right]$$

$$= \frac{\sqrt{n}x^2}{4\sqrt{2}\pi^{3/2}}(1 + O(n^{-1})) \exp\left[-\frac{n}{8}(R - R_1 - R_2)^2 - \frac{x^2}{2}\left\{(R_1 - 1)^2 + (R_2 - 1)^2\right\}\right],$$
(A5)

which is Equation 15.

Appendix B: Probability of sign epistasis

In this appendix, we present the mathematical details of the derivation of the probabilities P_r and P_s of observing reciprocal sign epistasis (RSE) and simple sign epistasis (SSE), respectively. As in the main text, let us assume $R_1 < R_2$. In calculating the probabilities, the integral over R takes one of three forms

$$\int_{-\infty}^{R_1} \sqrt{\frac{n}{8\pi}} e^{-\frac{n}{8}(-R + R_1 + R_2)^2} dR = \frac{1}{2}\text{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right), \qquad \int_{R_2}^{\infty} \sqrt{\frac{n}{8\pi}} e^{-\frac{n}{8}(-R + R_1 + R_2)^2} dR = \frac{1}{2}\text{erfc}\left(-\frac{\sqrt{n}R_1}{2\sqrt{2}}\right),$$

$$\int_{R_1}^{R_2} \sqrt{\frac{n}{8\pi}} e^{-\frac{n}{8}(-R + R_1 + R_2)^2} dR = \frac{1}{2}\left[\text{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) - \text{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right)\right].$$
(B1)

First, we consider RSE which corresponds to the two domains

$$D_1 = \{(R_1, R_2, R) | R_1 < R_2 < R, R_2 < 0\}, \quad D_2 = \{(R_1, R_2, R) | R < R_1 < R_2, R_1 > 0\},$$
(B2)

as illustrated in Figure 2. The probability of being in D_1 is

$$\text{Pr}(D_1) = \int_{-\infty}^{0} dR_1 \int_{R_1}^{0} dR_2 \int_{R_2}^{\infty} dR \mathcal{P}(R_1, R_2, R_3) = \frac{x^2}{2\pi} \int_{-\infty}^{0} dR_1 \int_{R_1}^{0} dR_2 \exp\left[-\frac{x^2}{2}(R_1 - 1)^2 - \frac{x^2}{2}(R_2 - 1)^2\right] \text{erfc}\left(-\frac{\sqrt{n}R_1}{2\sqrt{2}}\right)$$

$$= \frac{x^2}{2\pi} \int_0^{\infty} dR_1 \int_0^{R_1} dR_2 \exp\left[-\frac{x^2}{2}(R_1 + 1)^2 - \frac{x^2}{2}(R_2 + 1)^2\right] \text{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right),$$
(B3)

where we have changed variables $R_i \mapsto -R_i$. Since $\mathrm{erfc}(y) \sim e^{-y^2}/(y\sqrt{\pi})$ for $y \gg 1$, the above integral is dominated by the region $R_1 \ll 1$ for large n. Thus, it is sufficient to approximate $\exp[-x^2((R_1-1)^2+(R_2-1)^2)/2] \approx e^{-x^2}$, which yields

$$\Pr(D_1) \approx \frac{x^2 e^{-x^2}}{4\pi} \int_0^\infty dR_1 \int_0^{R_1} dR_2 \mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) \approx \frac{x^2 e^{-x^2}}{4\pi} \int_0^\infty dR_1 R_1 \mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) = \frac{2x^2 e^{-x^2}}{n\pi} \int_0^\infty dy\, y\,\mathrm{erfc}(y) = \frac{x^2}{2n\pi} e^{-x^2}. \quad \text{(B4)}$$

The probability of being in D_2 has the same leading behavior,

$$\Pr(D_2) = \int_0^\infty dR_2 \int_0^{R_2} dR_1 \int_{-\infty}^{R_1} dR \mathcal{P}(R_1,R_2,R_3) = \frac{x^2}{2\pi} \int_0^\infty dR_2 \int_0^{R_2} dR_1 \exp\left[-\frac{x^2}{2}(R_1-1)^2 - \frac{x^2}{2}(R_2-1)^2\right] \mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right)$$

$$\approx \frac{x^2}{2\pi} \int_0^\infty dR_1 \int_0^{R_1} dR_2 e^{-x^2} \mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) = \frac{x^2}{2n\pi} e^{-x^2}, \quad \text{(B5)}$$

where we have exchanged the variables $R_1 \leftrightarrow R_2$. Due to the symmetrical roles of R_1 and R_2, the total probability of RSE is

$$P_r \equiv 2\sum_{i=1}^{2} \Pr(D_i) \approx \frac{2x^2}{n\pi} e^{-x^2}. \quad \text{(B6)}$$

We can use a similar approximation scheme to calculate the probability of SSE. There are four domains contributing to SSE (see Figure 2),

$$D_3 = \{(R_1,R_2,R)|R < R_1 < 0 < R_2\}, \quad D_4 = \{(R_1,R_2,R)|R_1 < 0 < R_2 < R\},$$
$$D_5 = \{(R_1,R_2,R)|R_1 < R < R_2 < 0\}, \quad D_6 = \{(R_1,R_2,R)|0 < R_1 < R < R_2\}. \quad \text{(B7)}$$

As we will see, all integrals can be represented by the functions

$$G_1(a,b) = \frac{x^2}{4\pi} \int_0^\infty dR_1 \int_0^{R_1} dR_2 \exp\left[-\frac{x^2}{2}(R_1+a)^2 - \frac{x^2}{2}(R_2+b)^2\right] \mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right), \quad \text{(B8)}$$

$$G_2(a,b) = \frac{x^2}{4\pi} \int_0^\infty dR_1 \int_0^{R_1} dR_2 \exp\left[-\frac{x^2}{2}(R_1+a)^2 - \frac{x^2}{2}(R_2+b)^2\right] \mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right)$$

$$= \frac{x}{4\sqrt{2\pi}} \mathrm{erfc}\left(\frac{bx}{\sqrt{2}}\right) \int_0^\infty dR_1 \exp\left[-\frac{x^2}{2}(R_1+a)^2\right] \mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) - G_1(b,a), \quad \text{(B9)}$$

where $a,b = \pm 1$ and we have used that

$$\int_0^\infty dy \int_y^\infty dz f(y,z) = \int_0^\infty dy \int_0^y dz f(z,y). \quad \text{(B10)}$$

To be specific, we write the probabilities of being in each domain as

$$\Pr(D_3) = \frac{x^2}{4\pi} \int_{-\infty}^0 dR_1 \int_0^\infty dR_2 \exp\left[-\frac{x^2}{2}(R_1-1)^2 - \frac{x^2}{2}(R_2-1)^2\right] \mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right) = G_1(1,-1) + G_2(-1,1),$$

$$\Pr(D_4) = \frac{x^2}{4\pi} \int_{-\infty}^0 dR_1 \int_0^\infty dR_2 \exp\left[-\frac{x^2}{2}(R_1-1)^2 - \frac{x^2}{2}(R_2-1)^2\right] \mathrm{erfc}\left(-\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) = G_1(-1,1) + G_2(1,-1),$$

$$\Pr(D_5) = \frac{x^2}{4\pi} \int_{-\infty}^0 dR_1 \int_{R_1}^0 dR_2 \exp\left[-\frac{x^2}{2}(R_1-1)^2 - \frac{x^2}{2}(R_2-1)^2\right] \left[\mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) - \mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right)\right] = G_1(1,1) - G_2(1,1),$$

$$\Pr(D_6) = \frac{x^2}{4\pi} \int_0^\infty dR_1 \int_{R_1}^\infty dR_2 \exp\left[-\frac{x^2}{2}(R_1-1)^2 - \frac{x^2}{2}(R_2-1)^2\right] \left[\mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) - \mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right)\right] = G_1(-1,-1) - G_2(-1,-1),$$

where we have changed negative integral domains into positive domains and made use of Equation B10. Using the approximation scheme explained above, we get

$$G_1(a,b) = \frac{x}{4\sqrt{2\pi}} \int_0^\infty dR_2 e^{-x^2(R_2+b)^2/2} \mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right) \mathrm{erfc}\left(\frac{x(R_2+a)}{\sqrt{2}}\right)$$

$$\approx \frac{x}{4\sqrt{2\pi}} \mathrm{erfc}\left(\frac{ax}{\sqrt{2}}\right) e^{-x^2/2} \int_0^\infty dR_2 \mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right) = \frac{x}{2\sqrt{n\pi}} \mathrm{erfc}\left(\frac{ax}{\sqrt{2}}\right) e^{-x^2/2} + O(1/n). \quad \text{(B11)}$$

Since

$$\frac{x}{4\sqrt{2\pi}} \mathrm{erfc}\left(\frac{bx}{\sqrt{2}}\right) \int_0^\infty dR_1 \exp\left[-\frac{x^2}{2}(R_1+a)^2\right] \mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) \approx \frac{x}{2\sqrt{n\pi}} \mathrm{erfc}\left(\frac{bx}{\sqrt{2}}\right) e^{-x^2/2} + O(1/n), \quad \text{(B12)}$$

we conclude that $G_2(a,b) = O(1/n)$. Using $\mathrm{erfc}(y) + \mathrm{erfc}(-y) = 2$, we finally obtain

$$P_s \equiv 2\sum_{i=3}^{6} \Pr(D_i) \approx \frac{4x}{\sqrt{n\pi}} e^{-x^2/2}. \quad \text{(B13)}$$

20   S. Hwang *et al.*

Appendix C: Large L behavior of $\mathcal{R}_s(L)$ for fixed n

In this appendix, we calculate the asymptotic behavior of the probability $\mathcal{R}_s(L)$ for a genotype with s mutations to be a local fitness maximum in the limit where L is large and the phenotype dimensions n is fixed. As explained in the main text, this probability has two contributions which arise from expanding the function $F(\vec{k}, \vec{z})$ defined in Equation 30 near $|\vec{z}| = 0$ and $|\vec{z}| = z^* \sim L$, respectively (see Equation 31).

First, we consider the contribution from the region $|\vec{z}| \ll 1$. In this case, we can approximate $F(\vec{k}, \vec{z})$ as

$$F(\vec{k}, \vec{z}) = \int_n e^{-i\vec{k}\cdot\vec{\xi}} p(\vec{\xi}) d\vec{\xi} - \int_{\mathcal{D}^c(\vec{z})} e^{-i\vec{k}\cdot\vec{\xi}} p(\vec{\xi}) d\vec{\xi} \approx e^{-k^2/2} - \int_{\mathcal{D}^c(\vec{z})} p(0) d\vec{\xi} = e^{-k^2/2} - A_n z^n \approx \exp\left[-\frac{k^2}{2} - A_n z^n e^{k^2/2} \right], \quad \text{(C1)}$$

where $k = |\vec{k}|$, $z = |\vec{z}|$, $\mathcal{D}^c(\vec{z}) = \{\vec{y}||\vec{y} - \vec{z}| \le z\}$ which is the complement of $\mathcal{D}(\vec{z})$, $A_n = p(0)S_{n-1}/n$ with $S_{n-1} = 2\pi^{n/2}/\Gamma(n/2)$ being the surface area of the unit sphere in $(n-1)$ dimensions, and $p(0) = (2\pi)^{-n/2}$. Note that the error of the above approximation is $O(z^{n+1})$. Thus, setting $\rho \equiv s/L$ we can approximate

$$\mathcal{R}_s^<(L) \approx \int_n \frac{d\vec{z}d\vec{k}}{(2\pi)^n} \exp\left[i\vec{k}\cdot\vec{z} + LH_1(\vec{k},\vec{z}) \right], \quad H_1(\vec{k},\vec{z}) \equiv -i\vec{k}\cdot\vec{q} - \rho\frac{k^2}{2} - \rho A_n z^n e^{k^2/2} - (1-\rho)A_n z^n, \quad \text{(C2)}$$

where $\vec{q} = \vec{Q}/L$. Since L is large, we can employ the saddle point approximation. One can easily see that the saddle point solving the equations $\partial_{k_j} H_1 = \partial_{z_k} H_1 = 0$ is at $\vec{z} = 0$ and $k_j = -iq_j/\rho$. Around the saddle point, we expand

$$H_1 \approx -\frac{q^2}{2\rho} - \frac{\rho}{2}\left(\vec{k} + i\vec{q}/\rho \right)^2 - A_n z^n \left[\rho e^{-q^2/(2\rho^2)} + (1-\rho) \right], \quad \text{(C3)}$$

which gives

$$\mathcal{R}_s^<(L) \approx \exp\left(-L\frac{q^2}{2\rho} \right) \int_n d\vec{z} \exp\left[-LA_n z^n \left\{ \rho e^{-q^2/(2\rho^2)} + (1-\rho) \right\} \right] \int_n \frac{d\vec{k}}{(2\pi)^n} \exp\left[-\frac{L\rho}{2}\left(\vec{k} + i\vec{q}/\rho \right)^2 \right]$$

$$= \frac{\exp\left(-L\frac{q^2}{2\rho} \right)}{(2\pi L\rho)^{n/2}} \int_0^\infty S_{n-1} z^{n-1} dz \exp\left[-LA_n z^n \left\{ \rho e^{-q^2/(2\rho^2)} + (1-\rho) \right\} \right] = \frac{s^{-n/2}\exp\left[-Q^2/(2s) \right]}{s\exp[-Q^2/(2s^2)] + L - s}\left[1 + O(L^{-1/n}) \right], \quad \text{(C4)}$$

and the last step involves a change of variables $z \to t = S_{n-1} z^n/n$. Since L appears in the integrand in the combination Lz^n, the error that arises from neglecting terms of $O(z^{n+1})$ is $L^{-1/n}$. The leading order of Equation C4 was reported in Equation 34.

Now we move on to the calculation of $\mathcal{R}_s^>(L)$, where the dominant contribution to $F(\vec{k}, \vec{z})$ comes from a region where $z \sim O(L)$. Using $\int d\vec{\xi} p(\vec{\xi}) \exp(-i\vec{k}\cdot\vec{\xi}) = \exp(-k^2/2)$, we calculate the integral $I \equiv \exp(-k^2/2) - F(\vec{k}, \vec{z})$ as

$$I = \frac{1}{(2\pi)^{n/2}} \int_0^{2z} d\xi_n e^{-ik_n\xi_n - \xi_n^2/2} \int_{B(2z,\xi_n)} d\vec{\xi}_\perp e^{-i\vec{k}_\perp\cdot\vec{\xi}_\perp - \vec{\xi}_\perp^2/2}$$

$$= \frac{1}{(2\pi)^{n/2}} \int_0^{2z} d\xi_n e^{-ik_n\xi_n - \xi_n^2/2} \left[\int_{\mathbb{R}^{n-1}} d\vec{\xi}_\perp e^{-i\vec{k}_\perp\cdot\vec{\xi}_\perp - \vec{\xi}_\perp^2/2} - \int_{B^c(2z,\xi_n)} d\vec{\xi}_\perp e^{-i\vec{k}_\perp\cdot\vec{\xi}_\perp - \vec{\xi}_\perp^2/2} \right]$$

$$= \frac{e^{-\vec{k}_\perp^2/2}}{\sqrt{2\pi}} \left[\int_0^\infty d\xi_n e^{-ik_n\xi_n - \xi_n^2/2} - \int_{2z}^\infty d\xi_n e^{-ik_n\xi_n - \xi_n^2/2} \right] - \frac{1}{(2\pi)^{n/2}} \int_0^{2z} d\xi_n e^{-ik_n\xi_n - \xi_n^2/2} \int_{B^c(2z,\xi_n)} d\vec{\xi}_\perp e^{-i\vec{k}_\perp\cdot\vec{\xi}_\perp - \vec{\xi}_\perp^2/2}$$

$$= \frac{e^{-k^2/2}}{2}\left(1 - \text{erf}\left(\frac{ik_n}{\sqrt{2}} \right) \right) - \frac{e^{-\vec{k}_\perp^2/2}}{\sqrt{2\pi}} \int_{2z}^\infty d\xi_n e^{-ik_n\xi_n - \xi_n^2/2} - C_1(\vec{k}, z), \quad \text{(C5)}$$

where we set $\vec{z} = z\vec{e}_n$, $\vec{\xi}_\perp = \vec{\xi} - \xi_n\vec{e}_n$ and $\vec{k}_\perp = \vec{k} - k_n\vec{e}_n$ with $\vec{e}_n = (0,\ldots,0,1)$, $B(2z,\xi_n)$ is an $(n-1)$ dimensional ball with radius $\sqrt{\xi_n(2z - \xi_n)}$ whose center is located at the origin, B^c is the relative complement of B with respect to $\mathbb{R}^{n-1}$, and $\text{erf}(z) = 2\int_0^z e^{-t^2} dt/\sqrt{\pi}$ is the error function. The definition of C_1 is self-explanatory. Since

$$\left| \int_{2z}^\infty d\xi_n e^{-ik_n\xi_n - \xi_n^2/2} \right| \le \int_{2z}^\infty d\xi_n e^{-\xi_n^2/2} = 2z \int_1^\infty dy e^{-2z^2 y^2} \approx \frac{e^{-2z^2}}{2z}, \quad \text{(C6)}$$

where we used the Laplace method for the asymptotic expansion, the leading finite z correction is expected to come from C_1 for $n > 1$. Note that C_1 is identically zero for $n = 1$. Thus we get

$$F(\vec{k}, \vec{z}) \approx \frac{1}{2}e^{-k^2/2}\left(1 + \text{erf}\left(\frac{i\vec{k}\cdot\vec{z}}{\sqrt{2}z} \right) \right) + C_1(\vec{k}, z), \quad \text{(C7)}$$

where k_n is written as a projection of $\vec{k}$ along the $\vec{z}$ direction, $k_n = \vec{k}\cdot\vec{z}/z$. Since

$$\left| C_1(\vec{k}, z) \right| \le \frac{1}{(2\pi)^{n/2}} \int_0^{2z} d\xi_n e^{-\xi_n^2/2} \int_{B^c(2z,\xi_n)} d\vec{\xi}_\perp e^{-\vec{\xi}_\perp^2/2} = C_1(0, z), \quad \text{(C8)}$$

it is sufficient to find an approximate formula for $C_1(0,z)$ to determine the z dependence of $C_1(\vec{k},z)$. Using spherical coordinates in $\mathbb{R}^{n-1}$, we get

$$
\begin{aligned}
C_1(0,z) &= \frac{S_{n-2}}{(2\pi)^{n/2}} \int_0^{2z} dy\, e^{-y^2/2} \int_{\sqrt{y(2z-y)}}^\infty dx\, x^{n-2} e^{-x^2/2} \\
&= \frac{S_{n-2}}{(2\pi)^{n/2}} \left[\int_0^{2z} dy\, e^{-y^2/2} \int_z^\infty dx\, x^{n-2} e^{-x^2/2} + \int_0^z dx\, x^{n-2} e^{-x^2/2} \left\{ \int_0^{M_-(x,z)} dy\, e^{-y^2/2} + \int_{M_+(x,z)}^z dy\, e^{-y^2/2} \right\} \right],
\end{aligned}
\tag{C9}
$$

where $S_{n-2} = 2\pi^{(n-1)/2}/\Gamma[(n-1)/2]$ is the surface area of the unit $(n-2)$-sphere. In the second term on the second line the order of integration was reversed and the integration boundaries $M_\pm(x,z) = z \pm \sqrt{z^2 - x^2}$ were introduced. Since the first integral ($\int_z^\infty dz$) and the third integral ($\int_{M_+}^z dy$) decrease exponentially with z, the main contribution to $C_1(0,z)$ comes from the second integral. Thus,

$$
\begin{aligned}
C_1(0,z) &\approx \frac{S_{n-2}}{(2\pi)^{n/2}} \int_0^z dx\, x^{n-2} e^{-x^2/2} \int_0^{M_-(x,z)} dy\, e^{-y^2/2} = \frac{S_{n-2} z^n}{(2\pi)^{n/2}} \int_0^1 dx\, x^{n-2} e^{-z^2 x^2/2} \int_0^{M_-(x,1)} dy\, e^{-z^2 y^2/2} \\
&= \frac{S_{n-2} z^{n-1}}{(2\pi)^{n/2}} \sqrt{\frac{\pi}{2}} \int_0^1 dx\, x^{n-2} e^{-z^2 x^2/2} \mathrm{erf}(M_-(x,1)z/\sqrt{2}).
\end{aligned}
\tag{C10}
$$

Since the last integral is dominated by the region $xz \leq 1$, we can approximate $M_-(x,1)z \approx x^2 z/2 \sim O(1/z)$ and $\mathrm{erf}(M_-(x,1)z/\sqrt{2}) = x^2 z/\sqrt{2\pi}$. Finally, we get

$$
C_1(0,z) \approx \frac{S_{n-2} z^n}{(2\pi)^{n+1/2}} \sqrt{\frac{\pi}{2}} \int_0^1 dx\, x^n e^{-z^2 x^2/2} \approx \frac{S_{n-2} z^n}{(2\pi)^{n+1/2}} \sqrt{\frac{\pi}{2}} \int_0^\infty dx\, x^n e^{-z^2 x^2/2} = \frac{n-1}{2\sqrt{2\pi}z},
\tag{C11}
$$

which also implies that $C_1(\vec{k},z) \sim O(z^{-1})$.

If we write

$$
F(\vec{k},\vec{z}) = \frac{1}{2} e^{-k^2/2} \left(1 + \mathrm{erf}\left(\frac{i\vec{k}\cdot\vec{r}}{\sqrt{2}r} \right) \right) \left(1 + \frac{1}{L} f(\vec{r},\vec{k}) + O(z^{-2}) \right),
\tag{C12}
$$

where $\vec{r} = \vec{z}/L$ and $r = z/L$, then comparison with Equation C7 and Equation C11 shows that $f(\vec{r},0) = (n-1)/(\sqrt{2\pi}r)$. Inserting Equation C12 into Equation 32, it follows that

$$
\mathcal{R}_s^>(L) \approx \frac{L^n}{2^L} \int \frac{d\vec{r}d\vec{k}}{(2\pi)^n} \exp\left[\rho f(\vec{r},\vec{k}) + (1-\rho) f(\vec{r},0) \right] \exp\left[L H_2(\vec{k},\vec{r}) \right],
\tag{C13}
$$

with

$$
H_2(\vec{k},\vec{r}) = i\vec{k}\cdot(\vec{r}-\vec{q}) - \rho\frac{k^2}{2} + \rho \ln\left(1 + \mathrm{erf}\left(\frac{i\vec{k}\cdot\vec{r}}{\sqrt{2}r} \right) \right).
\tag{C14}
$$

Now we employ the steepest descent method. For convenience, we set $\vec{q} = (q,0,\ldots,0)$. The saddle point satisfies the equations

$$
\frac{\partial H_2}{\partial r_j} = ik_j + i\frac{\sqrt{2}\rho}{\sqrt{\pi}r^3} \exp\left(\frac{(\vec{k}\cdot\vec{r})^2}{2r^2} \right) \left[k_j r^2 - r_j(\vec{k}\cdot\vec{r}) \right] \left(1 + \mathrm{erf}\left(\frac{i\vec{k}\cdot\vec{r}}{\sqrt{2}r} \right) \right)^{-1} = 0,
\tag{C15}
$$

$$
\frac{\partial H_2}{\partial k_j} = i(r_j - q\delta_{j1}) - \rho k_j + i\frac{\sqrt{2}\rho}{\sqrt{\pi}r} \exp\left(\frac{(\vec{k}\cdot\vec{r})^2}{2r^2} \right) r_j \left(1 + \mathrm{erf}\left(\frac{i\vec{k}\cdot\vec{r}}{\sqrt{2}r} \right) \right)^{-1} = 0,
\tag{C16}
$$

with the solution $\vec{k}^* = 0$ and $\vec{r}^* = (q - \rho\sqrt{2/\pi}, 0, \ldots, 0)$. Note that there is no solution if $q < \rho\sqrt{2/\pi}$, so the valid range of ρ has the upper boundary $\rho_c(q) = \min(1, \sqrt{\pi/2}q)$. The matrix of second derivatives around the saddle point $\vec{k}^*, \vec{r}^*$ is

$$
\left.\frac{\partial^2 H_2}{\partial r_l \partial r_j}\right|_* = 0, \quad \left.\frac{\partial^2 H_2}{\partial k_m \partial k_j}\right|_* = -\rho\delta_{mj}\left(1 - \frac{2}{\pi}\delta_{l1}\right), \quad \left.\frac{\partial^2 H_2}{\partial r_m \partial k_j}\right|_* = i\delta_{jl}\left[1 + (1-\delta_{m1})\frac{\sqrt{2}\rho}{\sqrt{\pi}q - \sqrt{2}\rho}\right].
\tag{C17}
$$

Thus, we get

$$
\mathcal{R}_s^>(L) \approx \frac{L^n}{2^L} e^{f(\vec{r}^*,0)} \int \frac{d\vec{y}d\vec{k}}{(2\pi)^n} \exp\left[-L\frac{\rho}{\pi}(\pi-2)k_1^2 - L\rho\vec{k}_\perp^2 + iLk_1 y_1 + iL\frac{\sqrt{\pi}q}{\sqrt{\pi}q - \sqrt{2}\rho}\vec{k}_\perp \cdot \vec{y}_\perp \right],
\tag{C18}
$$

where $\vec{y} = \vec{r} - \vec{r}^*$, $\vec{y}_\perp = (0, y_2, \ldots, y_n)$, and $\vec{k}_\perp = (0, k_2, \ldots, k_n)$. If we perform the integration over $\vec{y}$ first, we obtain delta functions which make the integral over $\vec{k}$ trivial. Finally, we arrive at

$$
\mathcal{R}_s^>(L) \approx 2^{-L} \theta(\rho_c - \rho)\left(\frac{q - 2\rho/\sqrt{2\pi}}{q} \exp\left(\frac{1}{\sqrt{2\pi}q - 2\rho} \right) \right)^{n-1}
\tag{C19}
$$

22   S. Hwang *et al.*

where $\theta(x)$ is the Heaviside step function defined by $\theta(x \geq 0) = 1$ and $\theta(x < 0) = 0$.

To evaluate the corresponding contribution to the number of fitness maxima, $\mathcal{N}_>$, we replace the summation over s in Equation 33 by an integral over $\rho = s/L$ and use Stirling's formula to approximate the binomial coefficients. This yields

$$\mathcal{N}_> = \sqrt{L} \int_0^1 d\rho \frac{\exp[L\{-\rho \ln \rho - (1-\rho)\ln(1-\rho) - \ln 2\}]}{\sqrt{2\pi\rho(1-\rho)}} \left(\frac{q - 2\rho/\sqrt{2\pi}}{q} \exp\left(\frac{1}{\sqrt{2\pi}q - 2\rho} \right) \right)^{n-1} \theta(\rho_c - \rho). \quad \text{(C20)}$$

If $\rho_c < \frac{1}{2}$ or $q < \sqrt{2\pi}^{-1} = q_0$, the integral is dominated around $\rho \approx \rho_c$, which results in an exponential decrease with L. On the other hand, if $\rho_c > \frac{1}{2}$, the integral is dominated around $\rho \approx \frac{1}{2}$, which gives

$$\mathcal{N}_> \approx \sqrt{\frac{2L}{\pi}} \int_{-\infty}^\infty dx e^{-2Lx^2} \left[\frac{q - 1/\sqrt{2\pi}}{q} \exp\left(\frac{1}{\sqrt{2\pi}q - 1} \right) \right]^{n-1} = \left[\frac{q - 1/\sqrt{2\pi}}{q} \exp\left(\frac{1}{\sqrt{2\pi}q - 1} \right) \right]^{n-1} \quad \text{(C21)}$$

as reported in Equation 43.

Appendix D: Derivation of Equation 47

In this appendix, we calculate the average number of fitness maxima $\mathcal{N}$ in the limit $n, L \to \infty$ at fixed ratio $\alpha \equiv n/L$. To this end, we write I_τ, the probability for the genotype τ to be a local fitness maximum, using the Heaviside step function as

$$I_\tau = \int \prod_{k=1}^L \left[d\vec{\xi}_k p(\vec{\xi}_k) \theta\left(\frac{1}{L}(\vec{z} + (1 - 2\tau_k)\vec{\xi}_k)^2 - \frac{1}{L}|\vec{z}|^2 \right) \right] \equiv \int \mathcal{D}\xi \prod_k \theta\left(\mathcal{E}_k\right), \quad \text{(D1)}$$

where $\vec{z}$ is determined by τ through Equation 3, $\int \mathcal{D}\xi \equiv \int \prod_k d\vec{\xi}_k p(\vec{\xi}_k)$, and $\mathcal{E}_k$ is defined as

$$\mathcal{E}_k = \frac{1}{L}\left(\vec{z} + \vec{\xi}_k(1 - 2\tau_k) \right)^2 - \frac{1}{L}|\vec{z}|^2 = \frac{1}{L}\left(2\vec{z} \cdot \vec{\xi}_k(1 - 2\tau_k) + \left|\vec{\xi}\right|^2 \right) = \frac{1}{L}\left[2\left(\vec{Q} + \sum_j \vec{\xi}_j \tau_j \right) \cdot \vec{\xi}_k(1 - 2\tau_k) + |\vec{\xi}_k|^2 \right]. \quad \text{(D2)}$$

Note that the prefactor $1/L$ is introduced to make $\mathcal{E}_k$ finite in the limit $L \to \infty$ and we have used that $(1 - 2\tau_k)^2 = 1$. Applying the identity (Tanaka and Edwards 1980; Bray and Moore 1980)

$$\theta(\mathcal{E}_k) = \int_0^\infty d\lambda_k \delta(\lambda_k - \mathcal{E}_k) = \int_0^\infty d\lambda_k \int_{-\infty}^\infty \frac{d\phi_k}{2\pi} \exp\left[i\phi_k(\lambda_k - \mathcal{E}_k)\right] \quad \text{(D3)}$$

to Equation D1, the expected number of local fitness maxima reads

$$\mathcal{N} = \sum_\tau \int \mathcal{D}\xi \prod_{k=1}^L \left[\int_0^\infty d\lambda_k \int_{-\infty}^\infty \frac{d\phi_k}{2\pi} e^{i\phi_k(\lambda_k - \mathcal{E}_k)} \right] = \sum_\tau \int \mathcal{D}\xi \mathcal{D}\lambda \mathcal{D}\phi \exp\left[\sum_{k=1}^L \left\{ i\phi_k\lambda_k + \frac{i}{L}\phi_k\left(2\vec{\xi}_k \cdot \sum_{j=1}^L \vec{\xi}_j \tau_j + 2\vec{\xi}_k \cdot \vec{Q} - |\vec{\xi}_k|^2 \right) \right\} \right]$$

$$= \sum_\tau \int \mathcal{D}\xi \mathcal{D}\lambda \mathcal{D}\phi \exp\left[\sum_{k=1}^L \left\{ i\phi_k\lambda_k + \frac{i}{L}\phi_k\left(2\vec{\xi}_k \cdot \vec{Q} - |\vec{\xi}_k|^2 \right) \right\} \right] \exp\left[\frac{i}{L}\sum_{k=1}^L \phi_k\vec{\xi}_k \cdot \sum_{j=1}^L \vec{\xi}_j (2\tau_j) \right], \quad \text{(D4)}$$

where $\int \mathcal{D}\lambda \equiv \int_0^\infty \prod_k d\lambda_k$, $\int \mathcal{D}\phi \equiv \int_{-\infty}^\infty \prod_k \frac{d\phi_k}{2\pi}$, and we made the change of variables $(2\tau_k - 1)\vec{\xi}_k \mapsto \vec{\xi}_k$ to arrive at the second equality. Using the identity

$$\exp\left(\frac{i}{L}\vec{X} \cdot \vec{Y} \right) = L^n \int_n d\vec{v}\delta\left(L\vec{v} - \vec{X} \right) \exp\left(i\vec{Y} \cdot \vec{v} \right) = \left(\frac{L}{2\pi} \right)^n \int_n d\vec{\mu} d\vec{v} \exp\left[iL\vec{\mu} \cdot \vec{v} - i\vec{X} \cdot \vec{\mu} + i\vec{Y} \cdot \vec{v} \right] \quad \text{(D5)}$$

which is valid for any n-dimensional real vectors $\vec{X}$ and $\vec{Y}$, we can write the last term of Equation D4 as

$$\exp\left[\frac{i}{L}\sum_k \phi_k\vec{\xi}_k \cdot \sum_j \vec{\xi}_j(2\tau_j) \right] = \left(\frac{L}{2\pi} \right)^n \int_n d\vec{\mu} d\vec{v} \exp\left[iL\vec{\mu} \cdot \vec{v} + i\sum_k \vec{\xi}_k \cdot \left(-\phi_k\vec{\mu} + 2\tau_k\vec{v} \right) \right], \quad \text{(D6)}$$

which gives

$$\mathcal{N} = \sum_\tau \int \mathcal{D}\lambda \mathcal{D}\phi \mathcal{D}\mu \mathcal{D}v e^{i\phi \cdot \lambda + iL\vec{\mu} \cdot \vec{v}} \prod_{k=1}^L \prod_{s=1}^n \int \frac{d\xi_{ks}}{\sqrt{2\pi}} \exp\left[i\left\{ \frac{1}{L}\phi_k\left(2\xi_{ks}Q_s - \xi_{ks}^2 \right) + \xi_{ks}(-\phi_k\mu_s + 2\tau_k v_s) \right\} - \frac{\xi_{ks}^2}{2} \right], \quad \text{(D7)}$$

where $\int \mathcal{D}\mu\mathcal{D}\nu \equiv \left(\frac{L}{2\pi}\right)^n \int_n d\vec{\mu}d\vec{\nu}$, $\phi \cdot \lambda \equiv \sum_{k=1}^L \phi_k \lambda_k$, and ξ_{ks}, Q_s, μ_s, ν_s are the s-th components of the vectors $\vec{\xi}_k$, $\vec{Q}$, $\vec{\mu}$, $\vec{\nu}$, respectively. Note that the integrals over the ξ_{ks}'s become independent of each other. If we choose $Q_s = Q/\sqrt{n} = q\sqrt{L/\alpha}$ for all s and define $\chi \equiv 2Q_s/L$, the integral over ξ_{ks} becomes

$$\int_{-\infty}^{\infty} \frac{d\xi_{ks}}{\sqrt{2\pi}} \exp\left[-\frac{\xi_{ks}^2}{2}(1+2i\phi_k/L) + i\xi_{ks}\{\phi_k(\chi-\mu_s) + 2\tau_k\nu_s\}\right] = \frac{1}{\sqrt{1+2i\phi_k/L}} \exp\left[-\frac{\{2\nu_s\tau_k + (\chi-\mu_s)\phi_k\}^2}{2(1+2i\phi_k/L)}\right]$$

$$= \frac{1}{\sqrt{1+2i\phi_k/L}} \exp\left[\frac{-(\mu_s-\chi)^2\phi_k^2 + 4\tau_k(\mu_s-\chi)\nu_s\phi_k - 4\nu_s^2\tau_k}{2(1+2i\phi_k/L)}\right], \tag{D8}$$

which, in turn, gives

$$\mathcal{N} = \sum_\tau \int \mathcal{D}\lambda\mathcal{D}\phi\mathcal{D}\mu\mathcal{D}\nu e^{i\phi\cdot\lambda + iL\vec{\mu}\cdot\vec{\nu}} \prod_{k=1}^L \frac{1}{(1+2i\phi_k/L)^{n/2}} \exp\left[\frac{-\phi_k^2\sum_s(\mu_s-\chi)^2 + 4\tau_k\phi_k\sum_s(\mu_s-\chi)\nu_s - 4\tau_k\sum_s\nu_s^2}{2(1+2i\phi_k/L)}\right]. \tag{D9}$$

If we now insert the identity

$$1 = \int_0^\infty da \int dbdc\delta\left(a - \sum_s(\mu_s-\chi)^2\right)\delta\left(b - \sum_s(\mu_s-\chi)\nu_s\right)\delta\left(c - \sum_s \nu_s^2\right)$$

$$= \int_0^\infty da \int dbdc \int \frac{dA}{2\pi/L}\frac{dB}{2\pi/L}\frac{dC}{2\pi/L} \exp\left[iAL\left\{a - \sum_s(\mu_s-\chi)^2\right\} + iBL\left\{b - \sum_s(\mu_s-\chi)\nu_s\right\} + iCL\left\{c - \sum_s \nu_s^2\right\}\right], \tag{D10}$$

we can write

$$\mathcal{N} = \sum_\tau \int \mathcal{D}\lambda\mathcal{D}\phi\mathcal{D}\mu\mathcal{D}\nu e^{i\phi\cdot\lambda + iL\vec{\mu}\cdot\vec{\nu}} \int_0^\infty da \int \frac{dAdbdBdcdC}{(2\pi/L)^3} \prod_k (1+2i\phi_k/L)^{-L\alpha/2} \exp\left[\frac{-a\phi_k^2 + 4b\tau_k\phi_k - 4c\tau_k}{2(1+2i\phi_k/L)}\right]$$

$$\times \exp\left[iAL\left\{a - \sum_s(\mu_s-\chi)^2\right\} + iBL\left\{b - \sum_s(\mu_s-\chi)\nu_s\right\} + iCL\left\{c - \sum_s \nu_s^2\right\}\right] \tag{D11}$$

where we have replaced n by $L\alpha$. The integral domain of a is restricted to the positive real axis in order to ensure that the integral with respect to ϕ_k in Equation D9 continues to be well-defined after the substitution. Performing the integrals over μ_s and ν_s, we get

$$\frac{L}{2\pi} \int d\mu_s d\nu_s e^{iL\mu_s\nu_s - iLA(\mu_s-r)^2 - iLB(\mu_s-r)\nu_s - iLC\nu_s^2} = \exp\left[-\frac{1}{2}\left(\text{Ln}\left\{\frac{((B-1)^2 - 4AC)}{Ai}\right\} + \text{Ln}(Ai)\right) + \frac{i4q^2A/\alpha}{4AC - (B-1)^2}\right], \tag{D12}$$

where $\text{Ln}(x)$ is the principal value of the logarithm with argument in the interval $(-\pi, \pi]$ and the branch cut lies on the negative real axis.

Subsequently, the remaining integral over ϕ_i and λ_i can be readily evaluated as follows:

$$\frac{1}{2\pi} \int_0^\infty d\lambda_k \int d\phi_k (1+2i\phi_k/L)^{-L\alpha/2} \exp\left[-\frac{a\phi_k^2 + 4b\tau_k\phi_k - 4c\tau_k}{2(1+2i\phi_k/L)} + i\phi_k\lambda_k\right] = T(a, bi, c, \tau_k) + \frac{1}{L}U(a, bi, c, \tau_k) + O(1/L^2), \tag{D13}$$

where

$$T(a, b, c, \tau) = \frac{1}{2}e^{-2c\tau}\left(\text{erf}\left(\frac{\alpha + 2b\tau}{\sqrt{2a}}\right) + 1\right),$$

$$U(a, b, c, \tau) = -\frac{4ac\tau + a + 2b\tau(\alpha + 2b\tau)}{\sqrt{2\pi}a^{3/2}} \exp\left[-\frac{(\alpha + 2b\tau)^2}{2a} - 2c\tau\right]. \tag{D14}$$

After summing over the τ_k's, we arrive at the equation

$$\mathcal{N} = \int_0^\infty da \int \frac{dAdbdBdcdC}{(2\pi/L)^3} \exp\left[\frac{U(a, bi, c, 1) + U(a, bi, c, 0)}{T(a, bi, c, 1) + T(a, bi, c, 0)}\right] \exp\left(L\Sigma(a, bi, c, Ai, B, Ci)\right) \tag{D15}$$

where

$$\Sigma(a, b, c, A, B, C) = aA + bB + cC - \frac{1}{2}\alpha\ln\left(4AC + (B-1)^2\right) - \frac{4Aq^2}{4AC + (B-1)^2} + \ln(T(a, b, c, 1) + T(a, b, c, 0)). \tag{D16}$$

The remaining integrals are hard to evaluate analytically. Instead, we resort to the saddle point method to obtain an asymptotic expansion of the integral. Since Σ is the exponential growth factor of the number of local maximam which must be a real number, one expects that the saddle points of Equation D16 are formed for the real arguments of Σ. This suggests that we should make the changes of variables $b \to b/i$, $A \to A/i$ and $C \to C/i$. For large L, the integrals are then dominated by the saddle point $(a^*, b^*, c^*, A^*, B^*, C^*)$

24   S. Hwang *et al.*

of $\Sigma(a, b, c, A, B, C)$. If there is more than one saddle point, the one giving the largest value of $\Sigma(a, b, c, A, B, C)$ has to be chosen. Then, the leading behavior of the number of maxima can be expressed in terms of the saddle point as

$$\mathcal{N} = \frac{1}{\sqrt{|\det H(\Sigma)|}} \exp\left[\frac{U(a^*, b^*, c^*, 1) + U(a^*, b^*, c^*, 0)}{T(a^*, b^*, c^*, 1) + T(a^*, b^*, c^*, 0)}\right] \exp\left(L\Sigma(a^*, b^*, c^*, A^*, B^*, C^*)\right), \tag{D17}$$

where $H(\Sigma)$ is the Hessian matrix around the saddle point. The reader may have noticed that the two principal values of the logarithm defined in Equation D12 are replaced by a real-valued logarithm in Equation D16, which can be dangerous in general. However, it can be shown that this substitution is indeed correct by verifying that $(B^* - 1)^2 + 4A^*C^*$ is always positive for all saddle points of Equation D16, and thus the imaginary arguments always cancel each other out.

Now, let us evaluate the saddle point conditions. The derivatives of Σ with respect to A, B, C are

$$\frac{\partial \Sigma}{\partial A} = a - 2\frac{\alpha C[4AC + (B-1)^2] + 2q^2(B-1)^2}{[4AC + (B-1)^2]^2}, \qquad \frac{\partial \Sigma}{\partial B} = b - \frac{(B-1)[\alpha(B-1)^2 + A(4C\alpha - 8q^2)]}{[4AC + (B-1)^2]^2},$$

$$\frac{\partial \Sigma}{\partial C} = c - \frac{2A[\alpha(B-1)^2 + A(4C\alpha - 8q^2)]}{[4AC + (B-1)^2]^2}. \tag{D18}$$

By requiring that the above three equations are zero at the saddle point, we get

$$A = \frac{\alpha c}{2(ac + b^2)}, \quad B - 1 = \frac{\alpha b}{ac + b^2}, \quad C = \frac{1}{4}\left(\frac{2a\alpha}{ac + b^2} + \frac{-\alpha \pm \sqrt{\alpha^2 - 16cq^2}}{c}\right). \tag{D19}$$

The two solutions of C force us to perform a two-fold analysis for the remaining integrals since we cannot *a priori* determine which solution will yield the correct saddle point. Instead, we introduce another real number $g = \pm\sqrt{\alpha^2 - 16cq^2}$ which is allowed to take both signs. Then, by imposing the functional relation $c(g) = \frac{\alpha^2 - g^2}{16q^2}$, both solutions are covered by a single analysis. In this way, the saddle point is obtained in terms of g instead of c. Finally, substituting this solution into Equation D17 gives Equation 47.

Appendix E: Mean phenotypic distance z^* in the joint limit

In this appendix, we will associate the fixed point value a^* of the variable a entering the complexity function Equation 47 with the mean phenotypic distance z^*. To this end, we first consider the probability density $P(\tau, a)$ that a genotype τ whose phenotypic vector is of squared magnitude $L^2a/4$ is a local maximum. Formally, we can write

$$P(\tau, a) = \int D\xi \prod_k \theta(\mathcal{E}_k)\delta\left(a - \frac{4}{L^2}\left(\vec{Q} + \sum_k \vec{\xi}_k \tau_k\right)^2\right) = \int \frac{L}{2\pi}dA \int D\xi \prod_k \theta(\mathcal{E}_k) \exp\left[iLaA - i\frac{4A}{L}\left(\vec{Q} + \sum_k \vec{\xi}_k \tau_k\right)^2\right]$$

$$= \frac{L}{2\pi}\int dA \int D\xi \prod_k \theta(\mathcal{E}_k) \int \mathcal{D}\psi \exp\left[iLaA + i\frac{L}{16A}\vec{\psi}^2 + i\vec{\psi} \cdot \left(\vec{Q} + \sum_k \vec{\xi}_k \tau_k\right)\right], \tag{E1}$$

where we have used the identity $\int_{-\infty}^{\infty} dx \exp(ipx^2 + iqx) = \sqrt{\pi/p}e^{-i\pi/4}\exp[-iq^2/(4p)]$ for $p > 0$, $\mathcal{D}\psi = \prod_s(\sqrt{L}e^{i\pi/4}d\psi_s)/(4\sqrt{\pi A})$, and the notation is the same as in Appendix D. Following the same procedure in the previous appendix, we get

$$P(\tau, a) = \int \mathcal{D}\xi \mathcal{D}\lambda \mathcal{D}\phi \mathcal{D}\mu \mathcal{D}v \, e^{i\phi \cdot \lambda + iL\vec{\mu} \cdot \vec{v}} \exp\left[\sum_k \frac{i}{L}\phi_k\left(2\vec{\xi}_k \cdot \vec{Q} - |\vec{\xi}_k|^2\right) + i\sum_k \vec{\xi}_k \cdot (-\phi_k\vec{\mu} + 2\tau_k\vec{v})\right]$$

$$\times \frac{L}{2\pi}\int dA \int \mathcal{D}\psi \exp\left[iLaA + i\frac{L}{16A}\vec{\psi}^2 + i\vec{\psi} \cdot (\vec{Q} + \sum_i \vec{\xi}_i \tau_i)\right]. \tag{E2}$$

By shifting $\vec{v} \to \vec{v} - \vec{\psi}/2$ and integrating over $\vec{\psi}$, we have

$$P(\tau, a) = \int \mathcal{D}\xi \mathcal{D}\lambda \mathcal{D}\phi \mathcal{D}\mu \mathcal{D}v \, e^{i\phi \cdot \lambda + iL\vec{\mu} \cdot \vec{v}} \exp\left[\sum_k \frac{i}{L}\phi_k\left(2\vec{\xi}_k \cdot \vec{Q} - |\vec{\xi}_k|^2\right) + i\sum_k \vec{\xi}_k \cdot (-\phi_k\vec{\mu} + 2\tau_k\vec{v})\right]$$

$$\times \frac{L}{2\pi}\int dA \int \mathcal{D}\psi \exp\left[iLaA + i\frac{L}{16A}\vec{\psi}^2 + i\vec{\psi} \cdot (\vec{Q} - L\vec{\mu}/2)\right]$$

$$= \int \mathcal{D}\xi \mathcal{D}\lambda \mathcal{D}\phi \mathcal{D}\mu \mathcal{D}v \, e^{i\phi \cdot \lambda + iL\vec{\mu} \cdot \vec{v}} \exp\left[\sum_k \frac{i}{L}\phi_k\left(2\vec{\xi}_k \cdot \vec{Q} - |\vec{\xi}_k|^2\right) + i\sum_k \vec{\xi}_k \cdot (-\phi_k\vec{\mu} + 2\tau_k\vec{v})\right]$$

$$\times \int \frac{dA}{2\pi/L} \exp\left\{iLA\left[a - \left(\frac{2\vec{Q}}{L} - \vec{\mu}\right)^2\right]\right\}. \tag{E3}$$

Since we have set $Q_s = Q/\sqrt{n}$ for $s = 1, \ldots, n$, the last integral becomes

$$\int \frac{dA}{2\pi/L} \exp\left\{ iLA \left[a - \left(\frac{2\vec{Q}}{L} - \vec{\mu} \right)^2 \right] \right\} = \int \frac{dA}{2\pi/L} \exp\left\{ iLA \left[a - \sum_s (\mu_s - \chi)^2 \right] \right\}. \tag{E4}$$

Since $\mathcal{N} = \sum_\tau \int_0^\infty da P(\tau, a)$, by applying the manipulations of Appendix D to Equation E3 we arrive at the same integral form as in Equation D10. Since $\sum_\tau P(\tau, a)$ is the mean number of local maxima whose phenotypic vectors have squared magnitude a, we see that the saddle point a^* of Equation D16 determines the mean phenotypic distance z^* through

$$z^* = L \frac{\sqrt{a^*}}{2}. \tag{E5}$$

This shows in particular that z^* is linear in L.

Appendix F: Mean genotypic distance ρ^* in the joint limit

To have access to the information about the typical value of the genotypic (Hamming) distance of a local fitness maximum from the wild type, we rewrite Equation D15 as

$$\mathcal{N} = \int_0^\infty da \int \frac{dA\,db\,dB\,dc\,dC}{(2\pi/L)^3} \sum_{s=0}^L \binom{L}{s} \left(T(a,b,c,1) + \frac{U(a,b,c,1)}{L} \right)^s \left(T(a,b,c,0) + \frac{U(a,b,c,0)}{L} \right)^{L-s}$$

$$\approx \int_0^\infty da \int \frac{dA\,db\,dB\,dc\,dC}{(2\pi/L)^3} \int_0^1 d\rho \frac{e^{L\Sigma(a,b,c,A,B,C,\rho)}}{\sqrt{2\pi L\rho(1-\rho)}} \exp\left[\rho \left(1 + \frac{U(a,b,c,1)}{T(a,b,c,1)} \right) + (1-\rho) \left(1 + \frac{U(a,b,c,0)}{T(a,b,c,0)} \right) \right], \tag{F1}$$

where we have rearranged the summation $\sum_\tau$ as $\sum_{s=0}^L \binom{L}{s}$ taking advantage of the inherent permutation symmetry, Stirling's formula has been used to evaluate the binomial coefficients, $\sum_s$ is approximated as $L \int_0^1 d\rho$ with $s = L\rho$, and

$$\Sigma(a,b,c,A,B,C,\rho) \equiv aA + bB + cC - \frac{1}{2}\alpha \ln\left(4AC + (B-1)^2 \right) - \frac{4Aq^2}{4AC + (B-1)^2}$$

$$+ \rho \ln T(a,b,c,1) + (1-\rho) \ln T(a,b,c,0) - \rho \ln \rho - (1-\rho) \ln(1-\rho). \tag{F2}$$

The saddle point equations for this expression involve seven variables including ρ. Since the saddle point equations for A, B, C are the same as Equation D18, we may again insert Equation D19 into Equation F2, which yields

$$\Sigma^{\mathrm{red}}(a,b,g,\rho) = -\ln 2 + \frac{\alpha}{2}\left[1 - \ln\frac{\alpha}{2} \right] - \frac{\alpha}{2}\ln\left(\frac{\alpha+g}{ac(g) + b^2} \right) + b + \frac{g}{2} - \rho \ln \rho - (1-\rho)\ln(1-\rho)$$

$$+ (1-\rho)\ln\left(\mathrm{erf}\left(\frac{\alpha}{\sqrt{2a}} \right) + 1 \right) + \rho \left(\ln\left[\left(\mathrm{erf}\left(\frac{\alpha+2b}{\sqrt{2a}} \right) + 1 \right) \right] - 2c(g) \right). \tag{F3}$$

Since

$$\frac{\partial \Sigma}{\partial \rho} = \ln \frac{T(a,b,c,1)}{T(a,b,c,0)} - \ln \rho + \ln(1-\rho), \tag{F4}$$

the saddle point value of ρ^* is

$$\rho^* = \frac{T(a^*,b^*,c^*,1)}{T(a^*,b^*,c^*,1) + T(a^*,b^*,c^*,0)} = \left\{ 1 + e^{2c^*} \left[\frac{\mathrm{erf}\left(\alpha/\sqrt{2a^*} \right) + 1}{\mathrm{erf}\left((\alpha+2b^*)/\sqrt{2a^*} \right) + 1} \right] \right\}^{-1}. \tag{F5}$$

By inserting ρ^* into the saddle point equations for a, b, c, one can easily see that the final equations are the same as those derived from Equation D16.

Appendix G: Derivation of Equation 50

The determination of the solution describing regime III relies on the intuition that as α becomes large, the fitness landscape is asymptotically linear with the wild type being the global fitness maximum, as demonstrated in **Sign epistasis** for $L = 2$. This suggests an ansatz where a^* is close to $4q^2$, which corresponds to the wild type phenotypic distance as shown in Equation E5. Given this clue, one can additionally find that

$$\frac{\partial}{\partial a} \Sigma^{\mathrm{red}}(a,b,g) = 0 \tag{G1}$$

26   S. Hwang *et al.*

is solved by $a = 4q^2$, $b = -\alpha$ and $g = \alpha$. Furthermore, if we evaluate the remaining saddle point conditions around this point, we find that this solution fails to solve them by a slight margin,

$$\frac{\partial}{\partial b} \Sigma^{\mathrm{red}}(a, b, g)\bigg|_{a=4q^2, b=-\alpha, g=\alpha} = \frac{e^{-\frac{\alpha^2}{8q^2}}}{\sqrt{2\pi}q} \tag{G2}$$

and

$$\frac{\partial}{\partial c} \Sigma^{\mathrm{red}}(a, b, g)\bigg|_{a=4q^2, b=-\alpha, g=\alpha} = \frac{\alpha}{8q^2}\mathrm{erfc}\left(\frac{\alpha}{2\sqrt{2}q}\right). \tag{G3}$$

Given the fact that $\mathrm{erfc}(x) = e^{-x^2}\left(\frac{1}{\sqrt{\pi}x} + O\left(x^{-3}\right)\right)$, these non-vanishing terms are seen to be of the order of $\epsilon = e^{-\frac{\alpha^2}{8q^2}}$. Hence, it is sufficient to consider an expansion around the zeroth order solution of the form $\Sigma^{\mathrm{red}}(4q^2 + A_1\epsilon, -\alpha + A_2\epsilon, \alpha + A_3\epsilon)$ to show that Equation 50 satisfies the saddle point conditions Equation 48. To this end, we first focus on the derivatives with respect to A_1 and A_2,

$$\frac{1}{\epsilon}\frac{\partial}{\partial A_1}\Sigma^{\mathrm{red}}(4q^2 + A_1\epsilon, -\alpha + A_2\epsilon, \alpha + A_3\epsilon) = -\frac{A_3\epsilon}{16q^2} + O(\epsilon^2),$$

$$\frac{1}{\epsilon}\frac{\partial}{\partial A_2}\Sigma^{\mathrm{red}}(4q^2 + A_1\epsilon, -\alpha + A_2\epsilon, \alpha + A_3\epsilon) = \left(\frac{1}{\sqrt{2\pi}q} - \frac{2A_2 + A_3}{2\alpha}\right)\epsilon + O(\epsilon^2). \tag{G4}$$

The vanishing contributions in ϵ imply that the zeroth order solution $(4q^2, -\alpha, \alpha)$ satisfies the first two saddle point conditions. Additionally, we find that the corrections of the order $O(\epsilon)$ are $A_3 = 0$ and $A_2 = \frac{\alpha}{\sqrt{2\pi}q}$. Since $A_3 = 0$ to leading order, the saddle point equation with respect to g should be evaluated to order $O(\epsilon^2)$. This yields

$$\frac{1}{\epsilon^2}\frac{\partial}{\partial B_3}\Sigma^{\mathrm{red}}(4q^2 + A_1\epsilon, -\alpha + A_2\epsilon, \alpha + B_3\epsilon^2) = \left(-\frac{A_1}{16q^2} - \frac{\sqrt{\frac{2}{\pi}}q}{\alpha^2} + O\left(\frac{q^3}{\alpha^4}\right)\right)\epsilon + O(\epsilon^2), \tag{G5}$$

and subsequently, A_1 is solved to be $A_1 = \left(\frac{16\sqrt{\frac{2}{\pi}}q^3}{\alpha^2} + O(q^4/\alpha^3)\right)\epsilon + O(\epsilon^2)$. Finally, by inserting the solutions A_1, A_2 and A_3 as well as the zeroth order solutions into Equation F5, the solution for ρ^* is found to be

$$\rho^* = \left(\frac{\sqrt{\frac{2}{\pi}}q\epsilon}{\alpha} + O\left(\frac{q^3}{\alpha^3}\right)\right) + O(\epsilon^2). \tag{G6}$$