bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 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 algebraically with increasing phenotypic dimension n, and varies non-monotonically with the distance from the phenotypic optimum. We then derive asymptotic 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 genotypic landscape. The dependence of the complexity on the parameters of the model is found to be surprisingly rich, and three distinct phases characterized by different landscape structures are identified. The complexity generally decreases with increasing phenotypic dimension, but a non-monotonic dependence on n is found in certain regimes. Our results 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 (or cannot) 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 $\frac{1}{\sqrt{n}}$ 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).

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

Copyright © 2016 by the Genetics Society of America
doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Thursday 22nd December, 2016%
[1]Corresponding author: Department of Physics, The Catholic University of Korea, Bucheon 14662, Republic of Korea. E-mail:spark0@catholic.ac.kr

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 epistastic 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). 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). 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 general 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, 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 study 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. Some implications of our study for the theory of adaptation and the interpretation of empirical data will be elaborated in the **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 $x =$

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$\frac{r\sqrt{n}}{2d}$. 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 $Q = |\vec{Q}| = d/\sigma$, and draw the displacement vectors $\vec{\xi}$ from the n-dimensional Gaussian $p(\vec{\xi})$ with unit covariance matrix.

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. As originally proposed by Martin *et al.* (2007), we assume 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 landscapes 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

$$\vec{z}(\tau) = \vec{Q} + \sum_{i=1}^{L} \tau_i \vec{\xi}_i. \tag{2}$$

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

$$-|\vec{z}(\tau)|^2 = -|\vec{Q}|^2 - 2\sum_{i=1}^{L}(\vec{Q}\cdot\vec{\xi}_i)\tau_i - \sum_{i,j=1}^{L}(\vec{\xi}_i\cdot\vec{\xi}_j)\tau_i\tau_j, \tag{3}$$

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

It is instructive to decompose Equation 3 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 3

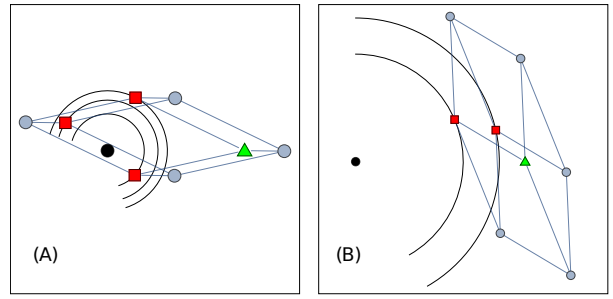

Figure 1 Two three-dimensional genotypic fitness landscapes induced by FGM with two phenotypic dimensions. The panels show the projection of the discrete space of genotypes onto the phenotype plane, where the phenotypic optimum is represented by a black dot. The wild type phenotype is marked by a green triangle, and nodes represented by red squares are local fitness maxima of the genotypic landscape, as can be seen from the contour lines of constant fitness.

can be recast into the form

$$-|\vec{z}(\tau)|^2 = -\left(Q + \sum_{i=1}^{L}\xi_i^{\parallel}\tau_i\right)^2 - \sum_{i,j=1}^{L}(\vec{\xi}_i^{\perp}\cdot\vec{\xi}_j^{\perp})\tau_i\tau_j. \tag{4}$$

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

We now show how the first term on the right hand side of Equation 4 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

$$2\int_{-\infty}^{0} dx\, \frac{x}{\sqrt{2\pi}} e^{-x^2/2} = -\sqrt{\frac{2}{\pi}} \equiv -2q_0, \tag{5}$$

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 4 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 4 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

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 at a nonzero fraction of beneficial mutations for $n \to \infty$. In our notation Fisher's parameter is

$$x = \frac{n}{2Q} \tag{6}$$

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

$$q = \frac{Q}{L}. \tag{7}$$

From a biological point of view one expects that $L \gg n \gg 1$, which motivates to consider this limit at constant phenotypic dimension n. Under this scaling the first term on the right hand side of Equation 4 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 4 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

$$\alpha = \frac{n}{L}. \tag{8}$$

Combining this with the limit $Q \to \infty$ at fixed $q = Q/L$, both terms on the right hand size of Equation 4 are of the same order $\sim L^2$. Fisher's parameter Equation 6 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 the **Results** section, 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 $P_{\rm wt}$ for the wild type to be a local fitness maxima, which is equal to the probability that all the L mutations are deleterious. Since mutations are statistically independent, we have

$$P_{\rm wt} = [1 - P_b]^L = 2^{-L}\left[1 + \mathrm{erf}(x/\sqrt{2})\right]^L, \tag{9}$$

where erf $= 1 -$ 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

$$\mathcal{N}_{\rm wt} = 2^L P_{\rm wt} = \left[1 + \mathrm{erf}(x/\sqrt{2})\right]^L \tag{10}$$

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 $\mathrm{erf}(y) \approx \frac{2}{\sqrt{\pi}} y$ we obtain

$$\mathcal{N}_{\rm wt} \approx \left[1 + \frac{2x}{\sqrt{2\pi}}\right]^L \to \exp\left(\frac{q_0 n}{q}\right) \tag{11}$$

for $L \to \infty$, where $q_0 = \frac{1}{\sqrt{2\pi}}$ was defined in Equation 5. 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 10 in the scaling limit (iii), where $x = \alpha/2q$ is a nonzero 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 10 will again turn out to be valid only when q is very large. In particular, whereas Equation 10 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 4, which grows with α and makes it increasingly more difficult for the mutant phenotypes to closely approach the origin.

Data availability

All simulations described in this work were implemented in Mathematica and C++. All relevant source codes are available from S.-C.P. on request.

Results

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

$$R_1 \equiv \frac{1}{n}\left(|\vec{\xi} + \vec{Q}|^2 - Q^2\right), \quad R_2 \equiv \frac{1}{n}\left(|\vec{\eta} + \vec{Q}|^2 - Q^2\right),$$
$$R \equiv \frac{1}{n}\left(|\vec{\xi} + \vec{\eta} + \vec{Q}|^2 - Q^2\right), \tag{12}$$

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 12. Without loss of generality we assume $R_1 < R_2$, and consider

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

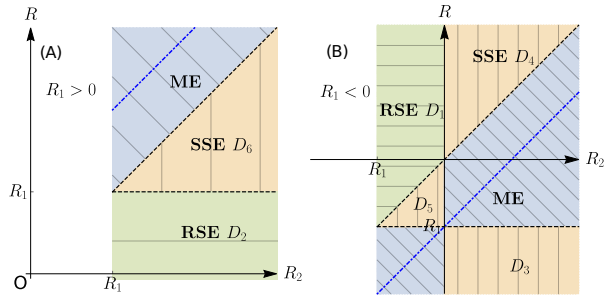

Figure 2 Domains of integration in the (R_2, R)-plane contributing to different types of epistasis: magnitude epistasis (ME), simple sign epistasis (SSE) and reciprocal sign epistasis (RSE). The two panels illustrate the two cases (A) $R_1 > 0$ and (B) $R_1 < 0$. Fitness effects are additive along the blue dash-dotted lines $R = R_1 + R_2$. The labeling of the domains $D_1, \ldots, D_6$ is used in the derivation in appendix B.

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 $R_2 > 0 > R_1$.

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*, SSE), or whether the interaction is reciprocal (*reciprocal sign epistasis*, RSE). The condition for 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 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 $R_1 \leftrightarrow R_2$.

The goal of this section is to compute the probability of observing each type of epistasis between randomly chosen mutations for fixed x and large n. For this purpose we require the joint probability density $\mathcal{P}(R_1, R_2, R)$. In appendix A it is shown that

$$\mathcal{P}(R_1, R_2, R) = \frac{x^2 n^{1/2}}{4\sqrt{2}\pi^{3/2}} e^{-\frac{1}{8}n(-R+R_1+R_2)^2 - \frac{x^2}{2}\left((R_1-1)^2+(R_2-1)^2\right)} \times \left(1 + O\left(\frac{1}{n}\right)\right), \tag{13}$$

a result that 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 that $\langle R_i \rangle = 1$, $\langle R_i^2 \rangle - \langle R_i \rangle^2 = 1/x^2$ for $i = 1, 2$ and $\langle R_1 R_2 \rangle - \langle R_1 \rangle \langle R_2 \rangle = 0$. Similarly, $\langle R - R_1 - R_2 \rangle = 0$, $\langle (R - R_1 - R_2)^2 \rangle - \langle R - R_1 - R_2 \rangle^2 = 4/n$ and $\langle (R - R_1 - R_2) R_1 \rangle - \langle R - R_1 - R_2 \rangle \langle R_1 \rangle = 0$. With the appropriate normalization constant, this leads directly to Equation 13.

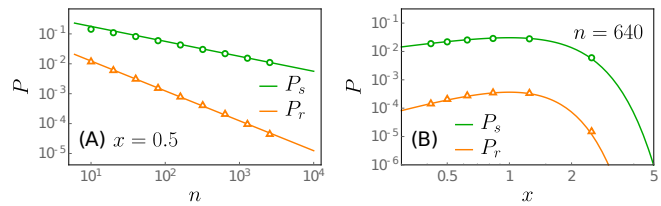

Figure 3 The probabilities of simple (P_s) and reciprocal (P_r) sign epistasis between two randomly chosen mutations. The curves are the asymptotic behaviors in Equation 14 and Equation 15 and symbols are simulation results. The asymptotic expressions provide accurate approximations for $n > 10$.

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

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

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 13 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

$$P_r = \frac{2x^2}{\pi n} e^{-x^2} + O(n^{-3/2}), \tag{14}$$

$$P_s = \frac{4x}{\pi\sqrt{n}} e^{-x^2/2} + O\left(n^{-1}\right), \tag{15}$$

for the probabilities of reciprocal (P_r) and simple (P_s) sign epistasis. The probability of magnitude epistasis is correspondingly given by $P_m = 1 - P_r - P_s$. Figure 3 compares analytic results to simulations, showing excellent agreement. As can be seen in Figure 3(B), the probability of sign epistasis varies non-monotonically with x and hence with the distance Q of the wild type to the phenotypic optimum.

Similarly, we can calculate the probabilities of sign epistasis conditioned on both mutations being beneficial, which in our setting implies $R_2 < 0$. 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. Hence

$$P_r^b = \frac{2\Pr(D_1)}{P_b^2} \approx \frac{4x^2}{\pi n \mathrm{erfc}\left(x/\sqrt{2}\right)^2} e^{-x^2} \tag{16}$$

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and

$$P_s^b = \frac{2\Pr(D_5)}{P_b^2} \approx \frac{4x}{\pi\sqrt{n}\mathrm{erfc}\left(x/\sqrt{2}\right)} e^{-x^2/2}, \tag{17}$$

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 13, 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

$$\tilde{P}_r(\tilde{R}) = \frac{\int_{\tilde{R}}^{\infty} \mathcal{P}(\tilde{R}, \tilde{R}, R)\, dR}{\int_{-\infty}^{\infty} \mathcal{P}(\tilde{R}, \tilde{R}, R)\, dR} = \frac{1}{2}\mathrm{erfc}\left(-\frac{\sqrt{n}\tilde{R}}{2\sqrt{2}}\right). \tag{18}$$

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 $\tilde{P}_r$ in terms of the selection coefficient of the single mutations we introduce a Gaussian phenotypic fitness function of the form

$$W(\vec{y}) = W_0 \exp(-\lambda|\vec{y}|^2), \tag{19}$$

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

$$S = \ln\left[\frac{W(\vec{Q}+\vec{\xi})}{W(\vec{Q})}\right] = -\lambda\left(|\vec{Q}+\vec{\xi}|^2 - |\vec{Q}|^2\right) = -\lambda n\tilde{R}. \tag{20}$$

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 18 becomes

$$\tilde{P}_r(S) = \frac{1}{2}\mathrm{erfc}\left(\frac{n^{3/2}}{8\sqrt{2}x^2}\frac{|S|}{S_0}\right). \tag{21}$$

The probability of sign epistasis conditioned on selection strength takes on its maximal value $\tilde{P}_r = \frac{1}{2}$ in the neutral limit $S \to 0$ and decreases monotonically with $|S|$. Similar to the results of Equation 14, Equation 15 and Equation 16 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 $\tilde{P}_r$ 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 $x = \frac{n}{2Q}$, 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 21 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 $\tilde{P}_r = 0$ to $\tilde{P}_r = 1$ at $S/S_0 = 8/9 = 0.888\ldots$.

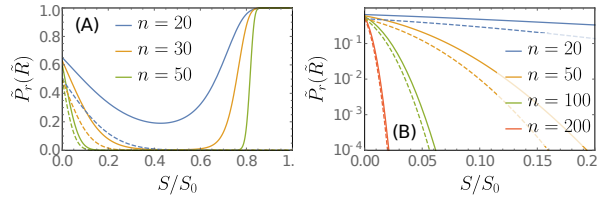

Figure 4 The probability of (reciprocal) sign epistasis $\tilde{P}_r$ conditioned on the selection coefficients of the two single mutations to be $S > 0$. Dashed lines show the asymptotic expression Equation 21 and full lines were obtained numerically using the Gaussian approximation for the distribution of epistasis developed by Schoustra *et al.* (2016).

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 2 satisfies

$$|\vec{z}(\tau)| < |\vec{z}(\tau) + (1-2\tau_i)\vec{\xi}_i| \tag{22}$$

for all $1 \le i \le L$. 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 Equation 2 ($\tau_i = 1$) or by adding a mutation vector that was not previously present ($\tau_i = 0$). The condition Equation 22 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 22 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

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

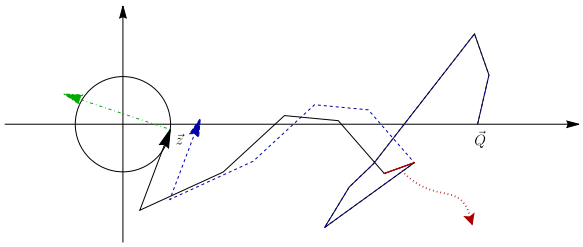

Figure 5 The condition Equation 22 for a genotype to be a local fitness maximum is illustrated for two phenotypic dimensions. The black circle encloses phenotypes that have higher fitness than the focal phenotype $\vec{z}(\tau)$. For τ to be a genotypic fitness maximum, neither the phenotypes obtained from $\vec{z}$ by adding a further mutation (dash-dotted green arrow) nor the phenotypes obtained by removing one of the mutations that constitute $\vec{z}$ (red segment and dotted blue arrows) should lie inside the circle.

are statistically equivalent. The expected number of fitness maxima can therefore be decomposed as

$$\mathcal{N} = \sum_{s=0}^{L} \binom{L}{s} \mathcal{R}_s(L), \tag{23}$$

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

$$\mathcal{R}_s(L) = \int_n d\vec{z} \left[\prod_{i=s+1}^{L} \int_{\mathcal{D}(-\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right] \times \left[\prod_{i=1}^{s} \int_{\mathcal{D}(\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right] \delta\left(\vec{z} - \vec{Q} - \sum_{i=1}^{s} \vec{\xi}_i \right), \tag{24}$$

with

$$\mathcal{D}(\vec{x}) \equiv \{\vec{y} \in \mathbb{R}^n | |\vec{y} - \vec{x}| > |\vec{x}|\}. \tag{25}$$

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

Equation 24 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 2. Next, the integration domains of the $\vec{\xi}_i$'s reflect the condition in Equation 22. 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

$$\delta(\vec{x}) = \frac{1}{(2\pi)^n} \int_n d\vec{k} \exp\left(i\vec{k} \cdot \vec{x}\right), \tag{26}$$

we can write

$$\mathcal{R}_s(L) = \int_n \int_n \frac{d\vec{z} d\vec{k}}{(2\pi)^n} \exp\left[i\vec{k} \cdot \left(\vec{z} - \vec{Q}\right)\right] F(\vec{k}, \vec{z})^s F(0, -\vec{z})^{L-s}, \tag{27}$$

where

$$F(\vec{k}, \vec{z}) \equiv \int_{\mathcal{D}(\vec{z})} d\vec{\xi} p(\vec{\xi}) \exp\left(-i\vec{k} \cdot \vec{\xi}\right). \tag{28}$$

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 27 comes from the region around the origin $\vec{z} = 0$ when q is small, but shifts to a distance $z^* \sim L$ along the $\vec{Q}$-axis 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

$$\mathcal{R}_s(L) = \mathcal{R}_s^{<}(L) + \mathcal{R}_s^{>}(L), \tag{29}$$

where

$$\mathcal{R}_s^{<}(L) = \int_{|\vec{z}|<z_0} d\vec{z} \int_n \frac{d\vec{k}}{(2\pi)^n} e^{i\vec{k}\cdot(\vec{z}-\vec{Q})} F(\vec{k}, \vec{z})^s F(0, -\vec{z})^{L-s},$$
$$\mathcal{R}_s^{>}(L) = \int_{|\vec{z}|>z_0} d\vec{z} \int_n \frac{d\vec{k}}{(2\pi)^n} e^{i\vec{k}\cdot(\vec{z}-\vec{Q})} F(\vec{k}, \vec{z})^s F(0, -\vec{z})^{L-s}, \tag{30}$$

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

$$\mathcal{N}_< = \sum_s \binom{L}{s} \mathcal{R}_s^{<}(L), \quad \mathcal{N}_> = \sum_s \binom{L}{s} \mathcal{R}_s^{>}(L). \tag{31}$$

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

$$\mathcal{R}_s^{<}(L) \approx \frac{s^{-n/2} \exp\left[-Q^2/(2s)\right]}{s \exp[-Q^2/(2s^2)] + L - s}. \tag{32}$$

For an interpretation of Equation 32 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

$$\text{Prob}(|\vec{z}| < \zeta) \approx \frac{V_n}{(2\pi)^{n/2}} s^{-n/2} \exp\left[-Q^2/(2s)\right], \tag{33}$$

where $V_n(r) \sim \zeta^n$ 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

$$\tilde{p}_s(\vec{\xi}) \approx \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\left|\vec{\xi} + \frac{\vec{Q}}{s}\right|^2\right). \tag{34}$$

For $\vec{z}$ to be a local optimum, 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 happens with probability

$$\left(\int_{\mathcal{D}(\vec{z})} d\vec{\xi} \tilde{p}_s(\vec{\xi})\right)^s \left(\int_{\mathcal{D}(-\vec{z})} d\vec{\xi} p(\vec{\xi})\right)^{L-s}$$
$$\approx \left\{1 - P_\zeta \exp[-Q^2/(2s^2)]\right\}^s \left(1 - P_\zeta\right)^{L-s}$$
$$\approx \exp\left[-P_\zeta \left\{s \exp[-Q^2/(2s^2)] + L - s\right\}\right], \tag{35}$$

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

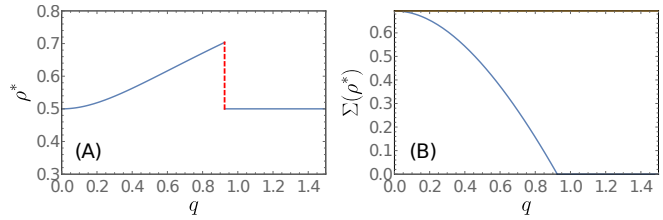

Figure 6 The most probable genotypic distance ρ^* and the genotypic complexity $\Sigma^* = \Sigma(\rho^*(q))$ as a function of q determined by Equation 38 and Equation 39. Starting from $1/2$, ρ^* increases with increasing q. The complexity vanishes at the phase transition point $q \approx q_c \simeq 0.924\,809$.

where $P_\zeta \equiv V_n/(2\pi)^{n/2}$. Thus, we can estimate the typical value of P_ζ as

$$P_\zeta = \frac{V_n}{(2\pi)^{n/2}} \approx \left[s\exp[-Q^2/(2s^2)] + L - s\right]^{-1}, \tag{36}$$

which, combined with Equation 33, indeed gives Equation 32.

To find the asymptotic behavior of $\mathcal{N}_<$ for large L, we use Stirling's formula in Equation 31 and approximate the summation over s by an integral over $\rho \equiv s/L$. This yields

$$\mathcal{N}_< \approx \int_0^1 d\rho \frac{1}{L^{n/2}\rho^{n/2}} \frac{e^{L\Sigma(\rho)}}{\sqrt{2\pi L\rho(1-\rho)}} \frac{1}{1-\rho+\rho e^{-\frac{q^2}{2\rho^2}}}, \tag{37}$$

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

$$\Sigma(\rho) \equiv -\rho\ln\rho - (1-\rho)\ln(1-\rho) - q^2/2\rho. \tag{38}$$

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

$$0 = \frac{\partial}{\partial\rho}\Sigma(\rho)\Big|_{\rho=\rho^*} = \frac{q^2}{2(\rho^*)^2} - \ln\frac{\rho^*}{1-\rho^*}. \tag{39}$$

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

$$\mathcal{N}_< \approx \frac{1}{L^{1+n/2}}\sqrt{\frac{1}{1+(1-\rho^*)\,(q/\rho^*)^2}}\,\frac{(\rho^*)^{-n/2}e^{L\Sigma(\rho^*)}}{1-\rho^*+\rho^* e^{-\frac{q^2}{2(\rho^*)^2}}}, \tag{40}$$

where $\rho^* = \rho^*(q)$ is the solution of Equation 39. 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 shows the behavior of ρ^* and $\Sigma(\rho^*)$ as a function of q obtained from a numerical solution of Equation 39. The most striking feature of these results is that Σ^* becomes negative for $q > q_c \approx 0.924\,809$, which signals a phase transition in the landscape properties. Inspection of Equation 38 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 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

$$\mathcal{N}_> \approx \left(\frac{q-q_0}{q}\exp\left[\frac{1}{q/q_0-1}\right]\right)^{n-1}. \tag{41}$$

This expression is valid for $q > q_0 = \frac{1}{\sqrt{2\pi}} \approx 0.398\,942$, but it dominates the contribution $\mathcal{N}_<$ for large L only when $q > q_c$. To understand the appearance of q_0, we refer to section **Model**, where it was argued that $2q_0 s$ is the maximal distance towards the origin that 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 in Equation 31 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 $z^* \sim O(L)$; clearly this assumption becomes invalid when $q \to q_0^+$. We note that for very large q and large n Equation 41 reduces to the expression $\mathcal{N}_{\rm wt}$ obtained in Equation 10 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

$$\mathcal{N} = \begin{cases}\mathcal{N}_<, & q < q_c,\\ \mathcal{N}_>, & q \geq q_c\end{cases} \tag{42}$$

with $\mathcal{N}_<$ and $\mathcal{N}_>$ given by Equation 40 and Equation 41, 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 typical *genotypic* location of the fitness maxima jumps discontinuously from $\rho^*(q_c) \approx 0.7035$ to $\rho^* = 1/2$ (Figure 6 (A)), and the *phenotypic* position jumps from $z^* \approx 0$ to $z^* = (q_c - q_o)L$. The *genotypic complexity* defined as

$$\Sigma^* = \lim_{L\to\infty}\frac{\ln\mathcal{N}}{L}, \tag{43}$$

is given by

$$\Sigma^* = \begin{cases}\Sigma(\rho^*), & q < q_c,\\ 0, & q \geq q_c,\end{cases} \tag{44}$$

where ρ^* is the solution of Equation 39, and hence vanishes continuously at $q = q_c$ (Figure 6 (B)). Note that the value $\Sigma^* = \ln 2$ attained at $q = 0$ is the largest possible, because the total number of genotypes is $2^L = e^{L\ln 2}$. 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 40 is a power law in L with exponent $1 + n/2$ and hence decreases with increasing n, whereas the expression in

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

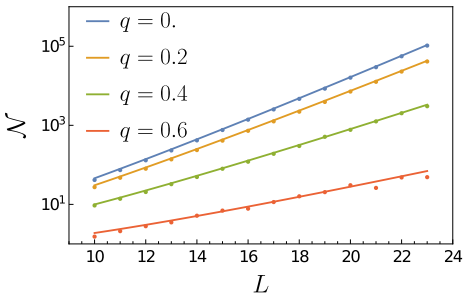

Figure 7 Mean number of maxima $\mathcal{N}$ is shown as a function of the genotypic dimension L for $q = 0, 0.2, 0.4$, and 0.6 with $n = 1$. Data from numerical simulations are represented as dots and the analytical prediction Equation 40 is shown as solid lines.

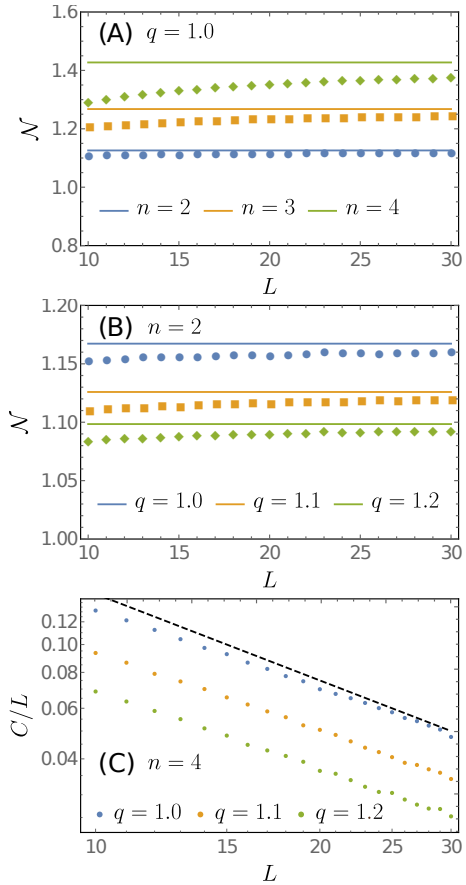

Figure 8 (A), (B) Average number of fitness maxima $\mathcal{N}$ as a function of the genotypic dimension L in regime II from the analytical prediction Equation 41 (solid lines) and numerical simulation (dots) averaged over 2×10^6 realizations for $q = 1.0, 1.1, 1.2$ and $n = 2, 3, 4$. Already for moderate L, the average number of local maxima converges to the theoretical lines. The deviations from the theoretical lines appear to be larger in (B) because of the finer resolution of the y-axis. (C) To verify that the numerical data converges to the theoretical values, we estimated the sub-leading correction to Equation 41 by calculating $1 - \frac{\mathcal{N}_{\text{data}}}{\mathcal{N}_{\text{theory}}}$ for $n = 4$ where the largest deviations are observed. The black dashed line with slope -1 shows that this correction is of the form C/L and vanishes for $L \to \infty$.

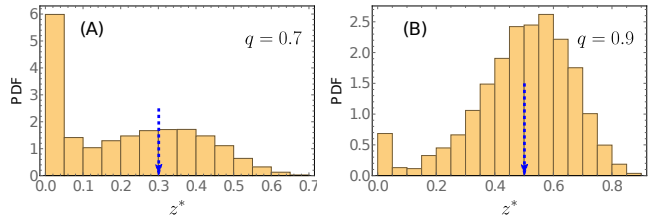

Figure 9 Histograms of the average phenotypic distance over fitness landscape realizations with $L = 15$ and $n = 2$, weighted by the number of local maxima in each realization. The bimodal structure implies that within one realization, only one of two mechanisms for generating multiple fitness maxima is operative. The blue dashed arrows indicate the mean phenotypic position $z^* = q - q_0$ predicted in regime II.

Equation 41 describing the regime $q > q_c$ increases exponentially with n.

To confirm our results, we compared the analytical predictions with simulations. In Figure 7, we compare $\mathcal{N}$ for $q < q_c$ with Equation 40 and in Figure 8, we compare $\mathcal{N}$ for $q > q_c$ with Equation 41. In both cases we see an excellent agreement even for moderately large L. In particular, for $q > q_c$ the number of maxima converges for large L to a value that is independent of L and increases with n (Figure 8(B)). 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 41 takes the value 1.2142...at $q = q_c$ and decreases rapidly towards unity for larger q.

Interpretation. The phase transition reflects a shift between two distinct mechanisms for generating genotypic complexity in the 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 approaches closely 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.

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 40 arises from the increasing difficulty of the random walk formed by the mutational vectors to locate the origin in high dimensions. By contrast, the mechanism operating for $q > q_c$ relies on the existence of the transverse dimensions, which is the reason why $\mathcal{N}_>$ in Equation 41 is an increasing function of n with $\mathcal{N}_> = 1$ for $n = 1$.

As our analysis is restricted to the average number of local maxima, at this point we cannot decide whether both mechanisms described above 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 10^6 realizations sharing the same parameters and calculated the average phenotypic position of the local fitness maxima present in each realization. Then, a histogram is constructed from these averages with a weight defined by the total number of local maxima in each

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

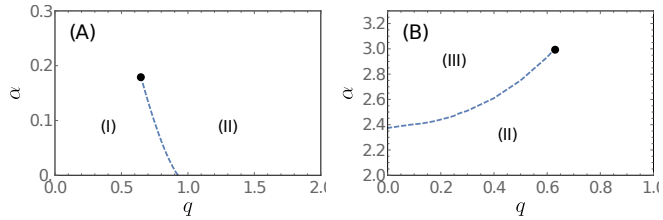

Figure 10 Phase diagrams illustrating the different regimes in the parameter space (q, α), where $q = Q/L$ is the scaled distance of the wildtype phenotype from the origin and $\alpha = n/L$ is the ratio of phenotypic and genotypic dimensions. Dashed lines are phase boundaries at which the typical genotypic and phenotypic locations of the local fitness maxima 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 $(q, \alpha) \simeq (0, 0.238)$ and continues to exist until approximately $q \simeq 0.62$.

realization. The resulting distributions shown in Figure 9 are clearly bimodal with peaks at $z^* \approx 0$ and $z^* \approx q - q_0$, and the weight of the second peak grows at the expense of the first with increasing q. This shows that within a single realization, most of the local optima belong to either of the two phases. The weighting by the number of maxima is indispensible, because realizations corresponding to regime I are very rare but contain a large number of maxima once they occur. Thus we conclude that the ensemble of fitness landscapes generated by FGM is highly heterogeneous, and the exponential dependence of the mean number of fitness maxima on L for $q < q_c$ is caused by a small number of exceptional realizations.

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 $L \gg n$. 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 sub-leading correction to Equation 40 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 $n/L = \alpha$. 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,

$$\mathcal{N} = \mathcal{C}(a^*, b^*, g^*) e^{L\Sigma^{\text{red}}(a^*, b^*, g^*)} \left(1 + O\left(\frac{1}{L}\right)\right), \tag{45}$$

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

$$\Sigma^{\text{red}}(a, b, g) = -\frac{\alpha}{2} \log\left(\frac{\alpha(\alpha + g)}{2(ac(g) + b^2)}\right) + \frac{\alpha + 2b + g}{2} + \log\left(\frac{1}{2}\left(e^{-2c(g)}\left(\text{erf}\left(\frac{\alpha + 2b}{\sqrt{2}\sqrt{a}}\right) + 1\right) + \text{erf}\left(\frac{\alpha}{\sqrt{2}\sqrt{a}}\right) + 1\right)\right) \tag{46}$$

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

$$\nabla\Sigma^{\text{red}}(a, b, g)|_{(a,b,g)=(a^*,b^*,g^*)} = (0, 0, 0) \tag{47}$$

where ∇ is the gradient with respect to the three variables (a, b, g). When several solutions of Equation 47 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 typical phenotypic distance z^* by the equation $z^* = L\sqrt{a^*}/2$.

An immediate consequence of Equation 45 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 40. Obtaining closed form solutions of Equation 47, which ultimately determine the functional dependence of the complexity 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 40 and Equation 41 as $\alpha \to 0$. For an overview of the behavior of the various quantities of interest as a function of q and α we refer to Figure 11.

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 $q > q_c$, respectively, 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 12, 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 $n \to L\alpha$ in Equation 41 yields an exponential dependence of $\mathcal{N}$ on L with the growth rate

$$\Sigma^{(\text{II})}_{\text{approx}} = \alpha \ln\left(\frac{q - q_0}{q} \exp\left[(q/q_0 - 1)^{-1}\right]\right). \tag{48}$$

This crude approximation turns out to be remarkably accurate even at $\alpha = 0.1$ as illustrated by Figure 12(B). By contrast, in regime I the naive replacement of n by αL in Equation 40 yields an expression that vanishes faster than exponential in L, as $\exp[-(\alpha/2)L \ln L]$. This reflects the fact that the typical phenotypic position z^* of the fitness maxima moves away from the origin for any $\alpha > 0$, and hence the complexity cannot be derived only by inspecting Equation 24 around $z = 0$ (see Figure

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

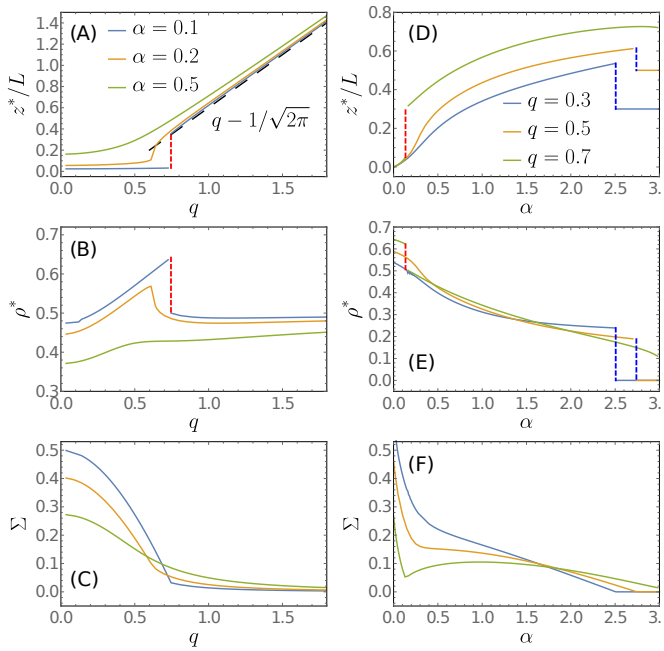

Figure 11 Number of local maxima and their mean phenotypic and genotypic positions are shown as a function of q (left column) and α (right column). (A-D): The phenotypic and genotypic distances behave similarly to the fixed n case which effectively corresponds to $\alpha = 0$. As α increases, the first order transition visualized by the red dashed lines disappears, and z^* and ρ^* vary continuously as a function of q. For sufficiently large α and small q, another phase transition to the regime III appears that was not observed in the previous limit (blue dashed lines in (B) and (D)). The genotypic maxima in regime III are located very close to the wild type position, $z^*/L = q$ and $\rho^* = 0$. This transition ceases to exist when q exceeds approximately 0.62. (E-F): The genotypic complexity is shown as a function of α or q. Note that the dependence on α is non-monotonic for $q = 0.7$.

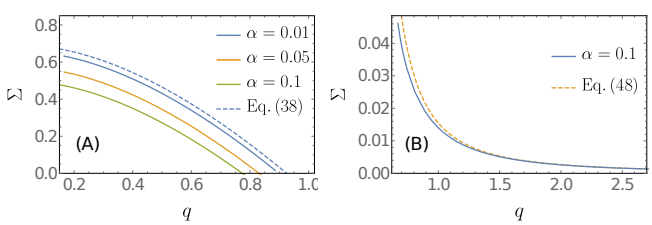

Figure 12 Convergence of the complexity to the fixed n case for small α. (A) The solid lines are constructed for $\alpha = 0.01, 0.05$ and 0.1 by numerically finding the solution of Equation 47 that belongs to Regime I. The convergence to Equation 40 (dashed line) is clearly seen as $\alpha \to 0$. (B) The blue solid line is similarly constructed from the second branch of solutions of Equation 47, whereas the dashed line is Equation 48 obtained from Equation 41. Even though a slight deviation is noticed as q approaches to q_0, Equation 48 remains a good approximation unless α becomes large.

11 (A, B)). At the same time the typical genotypic position ρ^* of the maxima, which is equal to the expected fraction of nonzero entries in the genotype sequence τ, decreases with increasing α and eventually falls below the value $\rho^* = 1/2$ favored by the sequence entropy (Figure 11 (C, D)).

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 4. 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), respectively, and this tendency is recovered from the joint approach when α is not too large (Figure 11 (E)). 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).

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 subsection **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 9. Second, the joint distribution given in Equation 13 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 additive 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 panels (B), (D) and (F) of Figure 11. However, depending on the value of q, the approach to the additive limit can be either smooth (for large q) or display characteristic jumps indicated by the blue dashed lines in Figure 11 (B) and (D). These jumps as well as the discontinuity in the slope of Σ as a function of α in Figure 11 (F) are hallmarks of the phase transition to the new regime III that is represented by the dashed line in Figure 10 (B). Fortunately, the solution of Equation 47 describing the new phase can be obtained analytically from Equation 46 or Equation F3 as a series expansion. The derivation outlined in

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Appendix G yields

$$a^* = 4q^2 - \left(\frac{16\sqrt{\frac{2}{\pi}}q^3}{\alpha^2} + O(q^4/\alpha^3)\right)\epsilon + O(\epsilon^2),$$
$$b^* = -\alpha + \frac{\alpha\epsilon}{\sqrt{2\pi}q} + O(\epsilon^2),$$
$$g^* = \alpha + O(\epsilon^2), \quad \text{and} \quad \rho^* = \frac{\sqrt{\frac{2}{\pi}}q\epsilon}{\alpha} + O(\epsilon^2), \tag{49}$$

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,

$$\Sigma^{(\mathrm{III})}(a^*, b^*, c^*) = \frac{\alpha\epsilon^2}{4\pi q^2} + O(\epsilon^3), \tag{50}$$

which shows that Σ is positive but vanishingly small in this regime. We note that using Equation E5, the expression for a^* in Equation 49 amounts to

$$\frac{z^*}{L} \simeq q - \frac{2\sqrt{\frac{2}{\pi}}q^2}{\alpha^2}\epsilon, \tag{51}$$

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 49 and Equation 51 can be easily derived from the idea that fitness 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

$$\mathcal{R}_s^{\mathrm{add}} = P_b^s(1-P_b)^{L-s} \tag{52}$$

where P_b is the probability for a mutation to be beneficial. Equation 52 expresses the condition that reverting 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 the same as Equation 49.

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 12 has a Gaussian distribution with mean 1 and variance $1/x^2 = 4q^2/\alpha^2$ for 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

$$L\rho^*\bar{R}_1 = \frac{(z^*)^2 - Q^2}{n}, \tag{53}$$

which yields the same leading behavior for z^*/L as in Equation 51.

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 11 (B) and (D) 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 11 (F)). In Figure 13 we verify this behavior using direct simulations of FGM. These simulations also show that the predictions based on Equation 46 are remarkably accurate already for moderate values of L and n.

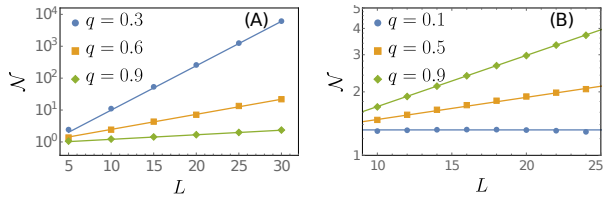

Figure 13 Numerical estimates of $\mathcal{N}$ as a function of L for $q = 0, 0.2, 0.4, 0.6$ and (A) $\alpha = 0.2$, (B) $\alpha =$???. The solid lines show the analytical prediction presented in Equation D17.

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, multidimensional 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; 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

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

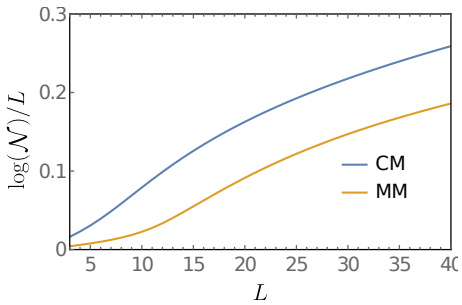

Figure 14 The logarithm of the number of local fitness maxima divided by the number of loci L is shown as a function of L for FGM with the parameter values $n = 19.3, Q = 6.89$ and $n = 34.8, Q = 9.81$ obtained by Schoustra *et al.* (2016) for the fungus *A. nidulans* growing in complete (CM) and minimal medium (MM), respectively. For the evaluation of $\mathcal{N}$ Equation 46 was used.

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 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 the common designation of n as "phenotypic complexity" may be somewhat misleading, 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, two of us used FGM 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 more rugged, as might be expected intuitively.

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 *e.g.* for the evolutionary benefits of recombination (de Visser *et al.* 2009).

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 14 and Equation 15. 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 understand-

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ing 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).

Acknowledgments

We thank Anton Bovier and David Dean for helpful discussions. This work was supported by DFG within SFB 680 *Molecular Basis of Evolutionary Innovations* and SPP 1590 *Probabilistic Structures in Evolution*. S-CP acknowledges the support by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (Grant No. 2014R1A- 1A2058694). S-CP would also like to thank Korea Institute for Advanced Study (KIAS) for its support and hospitality during his stay there on sabbatical leave (2016-2017).

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. Journal of Physics C: Solid State Physics **13**: L469.

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.

de Visser, J. A. G. M. and J. Krug, 2014 Empirical fitness landscapes and the predictability of evolution. Nature Reviews Genetics **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.

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

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.

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.

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., 2005 The genetic theory of adaptation: A brief history. Nature Reviews Genetics **6**: 119–127.

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.

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.

Schenk, M. F., I. G. Szendro, M. L. Salverda, J. Krug, and J. A. G. de Visser, 2013 Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. Mol. Biol. Evol. **30**: 1779–1787.

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. Journal of

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Physics F: Metal Physics **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.

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? Current Opinion in Genetics & Development **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.

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 $\vec{Q} = \frac{Q}{\sqrt{n}}(1,1,1,\ldots,)$. The probability density $\mathcal{P}(R_1, R_2, R)$ can then be formally defined as

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

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

$$\mathcal{P}(R_1,R_2,R) = \frac{n^3}{(2\pi)^3}\int d\vec{k} e^{ik_1nR_1+ik_2nR_2+iknR}\prod_i \left\langle e^{-ik_1(\xi_i+Q_i)^2-ik_2(\eta_i+Q_i)^2-ik(\xi_i+\eta_i+Q_i)^2+iQ_i^2(k_1+k_2+k_3)}\right\rangle_{\vec{\xi},\vec{\eta}}, \tag{A2}$$

where $d\vec{R}$ and $d\vec{k}$ stand for dR_1dR_2dR and dk_1dk_2dk, 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

$$\begin{aligned}&\frac{1}{2\pi}\int_{-\infty}^{\infty}d\xi\int_{-\infty}^{\infty}d\eta\exp\left[-\frac{\eta^2}{2}-\frac{\xi^2}{2}-ik_1\xi^2-ik_2\eta^2-ik(\xi+\eta)^2-2ik_1\frac{Q}{\sqrt{n}}\xi-2ik_2\frac{Q}{\sqrt{n}}\eta-2ik\frac{Q}{\sqrt{n}}(\xi+\eta)\right]\\ =&\sqrt{\frac{1}{(1+2ik_1)(1+2ik_2)-4k(k_1+k_2-i)}}\exp\left[\frac{2Q^2}{n}\frac{2i(k+k_1)(k+k_2)(k_1+k_2-i)+(k_1-k_2)^2}{4k(k_1+k_2-i)+(2k_1-i)(2k_2-i)}\right],\end{aligned} \tag{A3}$$

which gives

$$\mathcal{P}(R_1,R_2,R) = \frac{n^3}{(2\pi)^3}\int d\vec{k}\frac{e^{ik_1nR_1+ik_2nR_2+iknR}}{[(1+2ik_1)(1+2ik_2)-4k(k_1+k_2-i)]^{n/2}}\exp\left[\frac{n^2}{2x^2}\frac{2i(k+k_1)(k+k_2)(k_1+k_2-i)+(k_1-k_2)^2}{4k(k_1+k_2-i)+(2k_1-i)(2k_2-i)}\right]. \tag{A4}$$

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

$$\begin{aligned}\mathcal{P}(R_1,R_2,R) \approx \frac{n^3}{(2\pi)^3}\int d\vec{k}\exp\Bigg[&-\frac{n^2}{2x^2}\left\{(k_1+k_2+k)^2-2k_1k_2\right\}\\ &-n\left\{4k^2+k_1^2+k_2^2-ik_1(R_1-1)-ik_2(R_2-1)+k(2i+2k_1+2k_2-iR)\right\}\Bigg]\\ =\frac{\sqrt{n}x^2}{4\sqrt{2}\pi^{3/2}}&(1+O(n^{-1}))\exp\left[-\frac{n}{8}(R-R_1-R_2)^2-\frac{x^2}{2}\left\{(R_1-1)^2+(R_2-1)^2\right\}\right],\end{aligned} \tag{A5}$$

which is Equation 13.

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

$$\begin{aligned}&\int_{-\infty}^{R_1}\sqrt{\frac{n}{8\pi}}e^{-\frac{n}{8}(-R+R_1+R_2)^2}dR = \frac{1}{2}\mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right), \qquad \int_{R_2}^{\infty}\sqrt{\frac{n}{8\pi}}e^{-\frac{n}{8}(-R+R_1+R_2)^2}dR = \frac{1}{2}\mathrm{erfc}\left(-\frac{\sqrt{n}R_1}{2\sqrt{2}}\right),\\ &\int_{R_1}^{R_2}\sqrt{\frac{n}{8\pi}}e^{-\frac{n}{8}(-R+R_1+R_2)^2}dR = \frac{1}{2}\left[\mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right)-\mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right)\right].\end{aligned} \tag{B1}$$

First we consider RSE which corresponds to the two domains

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

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

$$\begin{aligned}\Pr(D_1) &= \int_{-\infty}^{0}dR_1\int_{R_1}^{0}dR_2\int_{R_2}^{\infty}dR\mathcal{P}(R_1,R_2,R_3) = \frac{x^2}{2\pi}\int_{-\infty}^{0}dR_1\int_{R_1}^{0}dR_2\exp\left[-\frac{x^2}{2}(R_1-1)^2-\frac{x^2}{2}(R_2-1)^2\right]\mathrm{erfc}\left(-\frac{\sqrt{n}R_1}{2\sqrt{2}}\right)\\ &= \frac{x^2}{2\pi}\int_{0}^{\infty}dR_1\int_{0}^{R_1}dR_2\exp\left[-\frac{x^2}{2}(R_1+1)^2-\frac{x^2}{2}(R_2+1)^2\right]\mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right),\end{aligned} \tag{B3}$$

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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,

$$\begin{aligned}\Pr(D_2) &= \int_0^\infty dR_2 \int_0^{R_2} dR_1 \int_{-\infty}^{R_1} dR\mathcal{P}(R_1,R_2,R_3) = \frac{x^2}{2\pi}\int_0^\infty dR_2 \int_0^{R_2} dR_1 \exp\left[-\frac{x^2}{2}(R_1-1)^2 - \frac{x^2}{2}(R_2-1)^2\right]\mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right)\\ &\approx \frac{x^2}{2\pi}\int_0^\infty dR_1 \int_0^{R_1} dR_2 e^{-x^2}\mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) = \frac{x^2}{2n\pi}e^{-x^2},\end{aligned} \quad \text{(B5)}$$

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),

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

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

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

$$\begin{aligned}G_2(a,b) &= \frac{x^2}{4\pi}\int_0^\infty dR_1 \int_0^{R_1} dR_2 \exp\left[-\frac{x^2}{2}(R_1+a)^2 - \frac{x^2}{2}(R_2+b)^2\right]\mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right)\\ &= \frac{x}{4\sqrt{2\pi}}\mathrm{erfc}\left(\frac{bx}{\sqrt{2}}\right)\int_0^\infty dR_1 \exp\left[-\frac{x^2}{2}(R_1+a)^2\right]\mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) - G_1(b,a),\end{aligned} \quad \text{(B9)}$$

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

$$\begin{aligned}\Pr(D_3) &= \frac{x^2}{4\pi}\int_{-\infty}^0 dR_1 \int_0^\infty dR_2 \exp\left[-\frac{x^2}{2}(R_1-1)^2 - \frac{x^2}{2}(R_2-1)^2\right]\mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right) = G_1(1,-1) + G_2(-1,1),\\ \Pr(D_4) &= \frac{x^2}{4\pi}\int_{-\infty}^0 dR_1 \int_0^\infty dR_2 \exp\left[-\frac{x^2}{2}(R_1-1)^2 - \frac{x^2}{2}(R_2-1)^2\right]\mathrm{erfc}\left(-\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) = G_1(-1,1) + G_2(1,-1),\\ \Pr(D_5) &= \frac{x^2}{4\pi}\int_{-\infty}^0 dR_1 \int_{R_1}^0 dR_2 \exp\left[-\frac{x^2}{2}(R_1-1)^2 - \frac{x^2}{2}(R_2-1)^2\right]\left[\mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) - \mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right)\right] = G_1(1,1) - G_2(1,1),\\ \Pr(D_6) &= \frac{x^2}{4\pi}\int_0^\infty dR_1 \int_{R_1}^\infty dR_2 \exp\left[-\frac{x^2}{2}(R_1-1)^2 - \frac{x^2}{2}(R_2-1)^2\right]\left[\mathrm{erfc}\left(\frac{\sqrt{n}R_1}{2\sqrt{2}}\right) - \mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right)\right] = G_1(-1,-1) - G_2(-1,-1),\end{aligned}$$

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

$$\begin{aligned}G_1(a,b) &= \frac{x}{4\sqrt{2\pi}}\int_0^\infty dR_2 e^{-x^2(R_2+b)^2/2}\mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right)\mathrm{erfc}\left(\frac{x(R_2+a)}{\sqrt{2}}\right)\\ &\approx \frac{x}{4\sqrt{2\pi}}\mathrm{erfc}\left(\frac{ax}{\sqrt{2}}\right)e^{-x^2/2}\int_0^\infty dR_2 \mathrm{erfc}\left(\frac{\sqrt{n}R_2}{2\sqrt{2}}\right) = \frac{x}{2\sqrt{n\pi}}\mathrm{erfc}\left(\frac{ax}{\sqrt{2}}\right)e^{-x^2/2} + O(1/n).\end{aligned} \quad \text{(B11)}$$

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

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

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 28 near $|\vec{z}|=0$ and $|\vec{z}|=z^*\sim L$, respectively (see Equation 29).

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

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

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

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

where $\vec{q}=\vec{Q}/L$ and the definition of H_1 is clear from the context. 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

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

which gives

$$\begin{aligned}\mathcal{R}_s^<(L)&\approx\exp\left(-L\frac{q^2}{2\rho}\right)\int_n d\vec{z}\exp\left[-LA_nz^n\left\{\rho e^{-q^2/(2\rho^2)}+(1-\rho)\right\}\right]\int_n\frac{d\vec{k}}{(2\pi)^n}\exp\left[-\frac{L\rho}{2}\left(\vec{k}+i\vec{q}/\rho\right)^2\right]\\&=\frac{\exp\left(-L\frac{q^2}{2\rho}\right)}{(2\pi L\rho)^{n/2}}\int_0^\infty S_{n-1}z^{n-1}dz\exp\left[-LA_nz^n\left\{\rho e^{-q^2/(2\rho^2)}+(1-\rho)\right\}\right]=\frac{s^{-n/2}\exp\left[-Q^2/(2s)\right]}{s\exp[-Q^2/(2s^2)]+L-s}\left[1+O(L^{-1/n})\right],\end{aligned} \tag{C4}$$

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 32.

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

$$\begin{aligned}I&=\frac{1}{(2\pi)^{n/2}}\int_0^{2z}d\xi_ne^{-ik_n\xi_n-\xi_n^2/2}\int_{B(2z,\xi_n)}d\vec{\xi}_\perp e^{-i\vec{k}_\perp\cdot\vec{\xi}_\perp-\vec{\xi}_\perp^2/2}\\&=\frac{1}{(2\pi)^{n/2}}\int_0^{2z}d\xi_ne^{-ik_n\xi_n-\xi_n^2/2}\left[\int_{\mathbb{R}^{n-1}}d\vec{\xi}_\perp e^{-i\vec{k}_\perp\cdot\vec{\xi}_\perp-\vec{\xi}_\perp^2/2}-\int_{B^c(2z,\xi_n)}d\vec{\xi}_\perp e^{-i\vec{k}_\perp\cdot\vec{\xi}_\perp-\vec{\xi}_\perp^2/2}\right]\\&=\frac{e^{-\vec{k}_\perp^2/2}}{\sqrt{2\pi}}\left[\int_0^\infty d\xi_ne^{-ik_n\xi_n-\xi_n^2/2}-\int_{2z}^\infty d\xi_ne^{-ik_n\xi_n-\xi_n^2/2}\right]-\frac{1}{(2\pi)^{n/2}}\int_0^{2z}d\xi_ne^{-ik_n\xi_n-\xi_n^2/2}\int_{B^c(2z,\xi_n)}d\vec{\xi}_\perp e^{-i\vec{k}_\perp\cdot\vec{\xi}_\perp-\vec{\xi}_\perp^2/2}\\&=\frac{e^{-k^2/2}}{2}\left(1-\mathrm{erf}\left(\frac{ik_n}{\sqrt{2}}\right)\right)-\frac{e^{-\vec{k}_\perp^2/2}}{\sqrt{2\pi}}\int_{2z}^\infty d\xi_ne^{-ik_n\xi_n-\xi_n^2/2}-C_1(\vec{k},z),\end{aligned} \tag{C5}$$

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 $\mathrm{erf}(z)=2\int_0^z e^{-t^2}dt/\sqrt{\pi}$ is the error function. The definition of C_1 is self-explanatory. Since

$$\left|\int_{2z}^\infty d\xi_ne^{-ik_n\xi_n-\xi_n^2/2}\right|\le\int_{2z}^\infty d\xi_ne^{-\xi_n^2/2}=2z\int_1^\infty dye^{-2z^2y^2}\approx\frac{e^{-2z^2}}{2z}, \tag{C6}$$

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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,

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

Since the last integral is dominated by the region $xz \le 1$, 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

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

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 30, 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

$$\mathcal{R}_s^>(L) \approx \frac{L^n}{2^L} e^{f(\vec{r}^*,0)} \int \frac{d\vec{y} d\vec{k}}{(2\pi)^n} \exp\left[-L\frac{\rho}{\pi}(\pi-2)k_1^2 - L\rho\vec{k}_\perp^2 + iLk_1y_1 + iL\frac{\sqrt{\pi}q}{\sqrt{\pi}q - \sqrt{2}\rho} \vec{k}_\perp \cdot \vec{y}_\perp \right], \tag{C18}$$

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

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

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 31 by an integral over $\rho = s/L$ and use Stirling's formula to approximate the binomial coefficients. This yields

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

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

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

as reported in Equation 41.

Derivation of Equation 46

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

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

where $\vec{z}$ is determined by τ through Equation 2, $\int \mathcal{D}\xi \equiv \int \prod_k d\vec{\xi}_k p(\vec{\xi}_k)$, and $\mathcal{E}_k$ is defined as

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

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)

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

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

$$\begin{aligned}\mathcal{N} &= \sum_\tau \int \mathcal{D}\xi \prod_{k=1}^{L}\left[\int_0^\infty d\lambda_k \int_{-\infty}^{\infty}\frac{d\phi_k}{2\pi}e^{i\phi_k(\lambda_k - \mathcal{E}_k)}\right] = \sum_\tau \int \mathcal{D}\xi\mathcal{D}\lambda\mathcal{D}\phi \exp\left[\sum_{k=1}^{L}\left\{i\phi_k\lambda_k + \frac{i}{L}\phi_k\left(2\vec{\xi}_k\cdot\sum_{j=1}^{L}\vec{\xi}_j\tau_j + 2\vec{\xi}_k\cdot\vec{Q} - |\vec{\xi}_k|^2\right)\right\}\right] \\ &= \sum_\tau \int \mathcal{D}\xi\mathcal{D}\lambda\mathcal{D}\phi \exp\left[\sum_{k=1}^{L}\left\{i\phi_k\lambda_k + \frac{i}{L}\phi_k\left(2\vec{\xi}_k\cdot\vec{Q} - |\vec{\xi}_k|^2\right)\right\}\right]\exp\left[\frac{i}{L}\sum_{k=1}^{L}\phi_k\vec{\xi}_k\cdot\sum_{j=1}^{L}\vec{\xi}_j\left(2\tau_j\right)\right],\end{aligned} \tag{D4}$$

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

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

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

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

which gives

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

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{d\xi_{ks}}{\sqrt{2\pi}}\exp\left[-\frac{\xi_{ks}^2}{2}\left(1+2i\phi_k/L\right)+i\xi_{ks}\left\{\phi_k(\chi-\mu_s)+2\tau_k\nu_s\right\}\right] &= \frac{1}{\sqrt{1+2i\phi_k/L}}\exp\left[-\frac{\{2\nu_s\tau_k+(\chi-\mu_s)\phi_k\}^2}{2(1+2i\phi_k/L)}\right]\\ &= \frac{1}{\sqrt{1+2i\phi_k/L}}\exp\left[\frac{-(\mu_s-\chi)^2\phi_k^2+4\tau_k(\mu_s-\chi)\nu_s\phi_k-4\nu_s^2\tau_k}{2(1+2i\phi_k/L)}\right],\end{aligned} \tag{D8}$$

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

$$\begin{aligned}1 &= \int_0^\infty da\int dbdc\,\delta\left(a-\sum_s(\mu_s-\chi)^2\right)\delta\left(b-\sum_s(\mu_s-\chi)\nu_s\right)\delta\left(c-\sum_s\nu_s^2\right)\\ &= \int_0^\infty da\int dbdc\int\frac{dA}{2\pi/L}\frac{dB}{2\pi/L}\frac{dC}{2\pi/L}\exp\left[iAL\left\{a-\sum_s(\mu_s-\chi)^2\right\}+iBL\left\{b-\sum_s(\mu_s-\chi)\nu_s\right\}+iCL\left\{c-\sum_s\nu_s^2\right\}\right],\end{aligned} \tag{D10}$$

we can write

$$\begin{aligned}\mathcal{N} = &\sum_\tau\int\mathcal{D}\lambda\mathcal{D}\phi\mathcal{D}\mu\mathcal{D}\nu e^{i\phi\cdot\lambda+iL\vec{\mu}\cdot\vec{\nu}}\int_0^\infty da\int\frac{dAdbdBdcdC}{(2\pi/L)^3}\prod_k(1+2i\phi_k/L)^{-L\alpha/2}\exp\left[\frac{-a\phi_k^2+4b\tau_k\phi_k-4c\tau_k}{2(1+2i\phi_k/L)}\right]\\ &\times\exp\left[iAL\left\{a-\sum_s(\mu_s-\chi)^2\right\}+iBL\left\{b-\sum_s(\mu_s-\chi)\nu_s\right\}+iCL\left\{c-\sum_s\nu_s^2\right\}\right]\end{aligned} \tag{D11}$$

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

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

where $\mathrm{Log}x$ 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

$$\begin{aligned}T(a,b,c,\tau) &= \frac{1}{2}e^{-2c\tau}\left(\mathrm{erf}\left(\frac{\alpha+2b\tau}{\sqrt{2a}}\right)+1\right),\\ U(a,b,c,\tau) &= -\frac{4ac\tau+a+2b\tau(\alpha+2b\tau)}{\sqrt{2\pi}a^{3/2}}\exp\left[-\frac{(\alpha+2b\tau)^2}{2a}-2c\tau\right].\end{aligned} \tag{D14}$$

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 $(a^*,b^*,c^*,A^*,B^*,C^*)$

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

$$\begin{aligned}
\frac{\partial\Sigma}{\partial A} &= a - 2\frac{\alpha C[4AC+(B-1)^2]+2q^2(B-1)^2}{[4AC+(B-1)^2]^2}, \qquad \frac{\partial\Sigma}{\partial B} = b - \frac{(B-1)[\alpha(B-1)^2+A(4C\alpha-8q^2)]}{[4AC+(B-1)^2]^2},\\
\frac{\partial\Sigma}{\partial C} &= c - \frac{2A[\alpha(B-1)^2+A(4C\alpha-8q^2)]}{[4AC+(B-1)^2]^2}.
\end{aligned} \tag{D18}$$

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 46.

Phenotypic location of typical local maxima

In this appendix, we will explain how the fixed point value a^* of the variable a entering the complexity function Equation 46 is related to the typical magnitude of the phenotype vectors corresponding to local genotypic fitness maxima, i.e. to the typical phenotypic location z^* of such maxima. 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

$$\begin{aligned}
P(\tau,a) &= \int D\xi \prod_k \theta(\mathcal{E}_k)\delta\left(a - \frac{4}{L^2}\left(\vec{Q}+\sum_k \vec{\xi}_k\tau_k\right)^2\right) = \int \frac{L}{2\pi}dA \int D\xi \prod_k \theta(\mathcal{E}_k)\exp\left[iLaA - i\frac{4A}{L}\left(\vec{Q}+\sum_k\vec{\xi}_k\tau_k\right)^2\right]\\
&= \frac{L}{2\pi}\int dA \int D\xi\prod_k\theta(\mathcal{E}_k)\int \mathcal{D}\psi \exp\left[iLaA + i\frac{L}{16A}\vec{\psi}^2 + i\vec{\psi}\cdot\left(\vec{Q}+\sum_k\vec{\xi}_k\tau_k\right)\right],
\end{aligned} \tag{E1}$$

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

$$\begin{aligned}
P(\tau,a) &= \int \mathcal{D}\xi\mathcal{D}\lambda\mathcal{D}\phi\mathcal{D}\mu\mathcal{D}v e^{i\phi\cdot\lambda+iL\vec{\mu}\cdot\vec{v}}\exp\left[\sum_k\frac{i}{L}\phi_k\left(2\vec{\xi}_k\cdot\vec{Q}-|\vec{\xi}_k|^2\right)+i\sum_k\vec{\xi}_k\cdot(-\phi_k\vec{\mu}+2\tau_k\vec{v})\right]\\
&\quad\times\frac{L}{2\pi}\int dA\int\mathcal{D}\psi\exp\left[iLaA+i\frac{L}{16A}\vec{\psi}^2+i\vec{\psi}\cdot(\vec{Q}+\sum_i\vec{\xi}_i\tau_i)\right]
\end{aligned} \tag{E2}$$

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

$$\begin{aligned}
P(\tau,a) &= \int \mathcal{D}\xi\mathcal{D}\lambda\mathcal{D}\phi\mathcal{D}\mu\mathcal{D}v e^{i\phi\cdot\lambda+iL\vec{\mu}\cdot\vec{v}}\exp\left[\sum_k\frac{i}{L}\phi_k\left(2\vec{\xi}_k\cdot\vec{Q}-|\vec{\xi}_k|^2\right)+i\sum_k\vec{\xi}_k\cdot(-\phi_k\vec{\mu}+2\tau_k\vec{v})\right]\\
&\quad\times\frac{L}{2\pi}\int dA\int\mathcal{D}\psi\exp\left[iLaA+i\frac{L}{16A}\vec{\psi}^2+i\vec{\psi}\cdot(\vec{Q}-L\vec{\mu}/2)\right]\\
&= \int \mathcal{D}\xi\mathcal{D}\lambda\mathcal{D}\phi\mathcal{D}\mu\mathcal{D}v e^{i\phi\cdot\lambda+iL\vec{\mu}\cdot\vec{v}}\exp\left[\sum_k\frac{i}{L}\phi_k\left(2\vec{\xi}_k\cdot\vec{Q}-|\vec{\xi}_k|^2\right)+i\sum_k\vec{\xi}_k\cdot(-\phi_k\vec{\mu}+2\tau_k\vec{v})\right]\\
&\quad\times\int\frac{dA}{2\pi/L}\exp\left\{iLA\left[a-\left(\frac{2\vec{Q}}{L}-\vec{\mu}\right)^2\right]\right\}
\end{aligned} \tag{E3}$$

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 distance z^* of typical local fitness maxima from the phenotypic optimum through

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

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

Typical number of mutations ρ^* of a local fitness maximum

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

$$\begin{aligned} \mathcal{N} &= \int_0^\infty da \int \frac{dAdbdBdcdC}{(2\pi/L)^3} \sum_{s=0}^{L} \binom{L}{s} \left(T(a,b,c,1) + \frac{U(a,b,c,1)}{L} \right)^s \left(T(a,b,c,0) + \frac{U(a,b,c,0)}{L} \right)^{L-s} \\ &\approx \int_0^\infty da \int \frac{dAdbdBdcdC}{(2\pi/L)^3} \int_0^1 d\rho \frac{e^{L\Sigma(a,b,c,A,B,C,\rho)}}{\sqrt{2\pi L\rho(1-\rho)}} \exp\left[\rho \left(1 + \frac{U(a,b,c,1)}{T(a,b,c,1)} \right) + (1-\rho) \left(1 + \frac{U(a,b,c,0)}{T(a,b,c,0)} \right) \right], \end{aligned} \tag{F1}$$

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

$$\begin{aligned} \Sigma(a,b,c,A,B,C,\rho) \equiv & aA + bB + cC - \frac{1}{2}\alpha \ln\left(4AC + (B-1)^2 \right) - \frac{4Aq^2}{4AC + (B-1)^2} \\ & + \rho \ln T(a,b,c,1) + (1-\rho) \ln T(a,b,c,0) - \rho \ln \rho - (1-\rho) \ln(1-\rho). \end{aligned} \tag{F2}$$

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

$$\begin{aligned} \Sigma^{\text{red}}(a,b,g,\rho) = & -\ln 2 + \frac{\alpha}{2}\left[1 - \ln\frac{\alpha}{2} \right] - \frac{\alpha}{2} \ln\left(\frac{\alpha + g}{ac(g) + b^2} \right) + b + \frac{g}{2} - \rho \ln \rho - (1-\rho)\ln(1-\rho) \\ & + (1-\rho) \ln\left(\operatorname{erf}\left(\frac{\alpha}{\sqrt{2a}} \right) + 1 \right) + \rho \left(\ln\left[\left(\operatorname{erf}\left(\frac{\alpha + 2b}{\sqrt{2a}} \right) + 1 \right) \right] - 2c(g) \right). \end{aligned} \tag{F3}$$

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

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

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.

Derivation of Equation 49

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

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

bioRxiv preprint doi: https://doi.org/10.1101/096438; this version posted December 23, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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,

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

and

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

Given the fact that $\text{erfc}(x) = e^{-x^2}\left(\frac{1}{\sqrt{\pi}x} + O\left(\left(\frac{1}{x}\right)^3\right)\right)$, 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 $\Sigma^{\text{red}}(4q^2 + A_1\epsilon, -\alpha + A_2\epsilon, \alpha + A_3\epsilon)$ to show that Equation 49 satisfies the saddle point conditions Equation 47. To this end, we first focus on the derivatives with respect to A_1 and A_2,

$$\begin{aligned}
\frac{1}{\epsilon}\frac{\partial}{\partial A_1}\Sigma^{\text{red}}(4q^2 + A_1\epsilon, -\alpha + A_2\epsilon, \alpha + A_3\epsilon) &= -\frac{A_3\epsilon}{16q^2} + O(\epsilon^2),\\
\frac{1}{\epsilon}\frac{\partial}{\partial A_2}\Sigma^{\text{red}}(4q^2 + A_1\epsilon, -\alpha + A_2\epsilon, \alpha + A_3\epsilon) &= \left(\frac{1}{\sqrt{2\pi}q} - \frac{2A_2 + A_3}{2\alpha}\right)\epsilon + O(\epsilon^2).
\end{aligned} \tag{G4}$$

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

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

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

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