Fisher's Geometrical Model Emerges as a Property of Complex Integrated Phenotypic Networks

Guillaume Martin[1]
Institut des Sciences de l'Evolution, Unité Mixte de Recherche 5554–Centre National de la Recherche Scientifique–Université Montpellier 2, 34095 Montpellier Cedex 05, France

ABSTRACT Models relating phenotype space to fitness (phenotype–fitness landscapes) have seen important developments recently. They can roughly be divided into mechanistic models (*e.g.*, metabolic networks) and more heuristic models like Fisher's geometrical model. Each has its own drawbacks, but both yield testable predictions on how the context (genomic background or environment) affects the distribution of mutation effects on fitness and thus adaptation. Both have received some empirical validation. This article aims at bridging the gap between these approaches. A derivation of the Fisher model "from first principles" is proposed, where the basic assumptions emerge from a more general model, inspired by mechanistic networks. I start from a general phenotypic network relating unspecified phenotypic traits and fitness. A limited set of qualitative assumptions is then imposed, mostly corresponding to known features of phenotypic networks: a large set of traits is pleiotropically affected by mutations and determines a much smaller set of traits under optimizing selection. Otherwise, the model remains fairly general regarding the phenotypic processes involved or the distribution of mutation effects affecting the network. A statistical treatment and a local approximation close to a fitness optimum yield a landscape that is effectively the isotropic Fisher model or its extension with a single dominant phenotypic direction. The fit of the resulting alternative distributions is illustrated in an empirical data set. These results bear implications on the validity of Fisher's model's assumptions and on which features of mutation fitness effects may vary (or not) across genomic or environmental contexts.

THE distribution of the fitness effects (DFE) of random mutations is a central determinant of the evolutionary fate of a population, together with the rate of mutation. Obviously, it determines the rate of adaptation by *de novo* mutations, by setting the mutational input of fitness variance. Furthermore, by setting the distribution of fitness at mutation–selection balance, the DFE also determines the amount of standing variance in populations at equilibrium and their potential for future adaptation. The DFE is therefore central to evolutionary theory, for both adapting and equilibrium populations. There is, however, no widely accepted model that predicts the distribution of fitness effects of random mutations and how it is affected by various environmental or genetic contexts. Yet, predicting what happens under changed conditions is a minimum requirement for many applications of evolutionary theory. "Phenotype–fitness landscapes" provide a general tool for such inference: by defining changed conditions (genetic background or environment) as explicit alternative "positions" in the landscape, their effects can be handled quantitatively.

doi: 10.1534/genetics.113.160325
Manuscript received December 1, 2013; accepted for publication January 30, 2014; published Early Online February 28, 2014.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.160325/-/DC1.
[1]Address for correspondence: Institut des Sciences de l'Evolution–Montpellier, ISEM CNRS UMR 5554, Université Montpellier II, Pl. Eugène Bataillon, Bât. 22, 34090 Montpellier, France. E-mail: guillaume.martin@univ-montp2.fr

"Mechanistic" landscapes

One such approach has seen considerable development in the past decade: models that explicitly describe the "direct" molecular effect of a mutation (on RNA secondary structure, on metabolic reactions, etc.) and integrate its effect on cellular yield or growth rate, through a network of phenotypic interaction. This approach, which can take various forms, is often dubbed "systems biology" (reviewed in Papp *et al.* 2011). It relies on a phenotype–fitness landscape that is parameterized from some empirical knowledge of the system, to describe part of the complex functional effect of given mutations. Probably the most popular and most advanced example of this approach is flux balance analysis (FBA). FBA has proved accurate in predicting, from first principles, the fitness effect of a wide variety of gene deletions (alone or in combination) in several model microbial species, mostly the bacterium *Escherichia coli* (Ibarra *et al.* 2002) and the yeast *Saccharomyces*

cerevisiae (Papp *et al.* 2004; Segré *et al.* 2005). It relies on a description of the effect of the removal of a given gene on the full metabolic network of the cell and ultimately on cell yield and growth rate. These approaches are powerful in both predictivity and explanatory potential, as they provide hints on why a particular genetic change has a given fitness effect. Other landscape models focus on point mutations affecting particular metabolic pathways [*e.g.*, the lactose utilization pathway (Perfeito *et al.* 2011)]. These studies test whether given mechanistic models can be accurately *fitted* to observations. FBA, on the contrary, seeks to *predict*, from first principles and independent calibration data, the effect of a set of deletions. Finally, a mechanistic approach has recently been proposed at the scale of multicellular organisms, with a developmental model (based on tooth morphology) predicting how mutations affect morphology and subsequently fitness (Salazar-Ciudad and Jernvall 2010; Salazar-Ciudad and Marin-Riera 2013). However, this model is intended as illustrative rather than quantitatively predictive, and it has not been empirically tested.

All these mechanistic approaches come at a cost, almost by definition: they require more or less extensive empirical descriptions of the genotype–phenotype–fitness relationship, and they are bound to describe only the particular mechanism considered. Therefore, they are mostly applied in species/strains where this relationship has been characterized empirically or can be "guessed" (a minimum requirement for FBA is a full genome sequence plus good knowledge of the growth medium). Mechanistic models are designed to describe a given aspect of a mutation's effect (*e.g.*, metabolic effect, secondary structure and RNA stability, etc.), typically at a cellular level. It is challenging to extend these predictions to mutations of unknown type (indels, gene duplications, transposon inserts, or single-nucleotide substitutions) that affect various functions and that modify an unknown aspect of the organism's fitness (expression levels, behavior, etc.). The scale of the prediction also typically limits applications to multicellular organisms (where the model must be integrated over many differentiated cells) or viruses (where it is the host phenotype that must be modeled). Overall, the unprecedented refinement of these mechanistic models has clearly provided key information, some of which is used here. However, their very precision limits their ability to predict the effect of random mutations, in less well-characterized species and environments, and hence their potential application in medicine, agronomy, or ecology.

"Heuristic" landscapes

A different approach has also been used for decades to predict the DFE: more heuristic landscapes like Fisher's (1930) geometrical model (FGM) (reviewed in Orr 2005). In this model, which may take various forms according to the starting assumptions, adaptation is characterized by stabilizing selection (quadratic or Gaussian), on a set of unspecified traits. Pleiotropic mutations jointly modify these traits, forming smooth (typically normal) distributions. The most predictive version is the isotropic FGM, where all traits are equivalent with respect to selection or mutation. In principle, the FGM can be used to predict how the DFE is affected by any environmental or genotypic context (epistasis), with any type of nonsilent mutation. Empirical support of the model's predictions has recently accumulated (Martin and Lenormand 2006a,b; Martin *et al.* 2007; MacLean *et al.* 2010; Sousa *et al.* 2011; Weinreich and Knies 2013), although it was sometimes relatively indirect. The most quantitative tests (Martin *et al.* 2007; MacLean *et al.* 2010; Sousa *et al.* 2011), and hence those with most statistical power, used the model to predict how the DFE is affected by epistasis. More generally, the FGM indeed predicts both the pervasiveness and the diminishing-return form of epistasis, as documented repeatedly in experimental evolution (*e.g.*, MacLean *et al.* 2010; Chou *et al.* 2011; Khan *et al.* 2011; Sousa *et al.* 2011). A model of pleiotropic mutations affecting the distance to an optimum is also qualitatively consistent with the prevalence of antagonistic pleiotropic effects affecting unused functions during long-term adaptation (as observed in Cooper and Lenski 2000). Finally, note that the model has been applied to various types of mutations (random point mutations, transposon inserts, and antibiotic resistance mutations) and in several species, although most were model microbial species, for logistic reasons.

In spite of this potential, Fisher's model is typically considered merely heuristic and too simplified to quantitatively capture the complex processes relating mutations to fitness components (growth rate, viability, fertility, etc.). Indeed, to date, the model has proved to be predictive only in a small number of tests. Compared to mechanistic models, it also does not predict the effect of particular mutations or their functional underpinnings, but only *distributions* among sets of mutants (but see Weinreich and Knies 2013). In any case, it remains unclear why such a simple model should capture features of highly complex processes. Therefore, even if further tests confirmed its quantitative predictivity, we would still be unable to tell under what conditions it should break down, which would limit its usefulness in forecasting.

Aim of the article

This study is an attempt to bridge the gap between mechanistic and heuristic approaches to phenotype–fitness maps. The aim is to reduce, by a statistical treatment, the complexity of the process relating mutation to fitness in a mechanistic model, resulting in a simplified model akin to the FGM. Statistical physics provides a successful example of such an endeavor: countless interacting particles generate a group behavior that is captured by the simple laws of thermodynamics. Given some assumptions on the microscopic process, this group behavior is predictable from a few measurable macroscopic quantities like temperature, volume, pressure, etc. The accuracy of the prediction increases with the number of random particles, namely with the complexity of the process. Here, I hope to make it plausible that a very similar argument applies to Fisher's geometric model, under a few qualitative assumptions on the genotype–phenotype–fitness map. These assumptions mostly derive from general features identified by systems biology

Table 1 Glossary of notations

FGM: Fisher's geometrical model of adaptation.
DFE: Distribution of the fitness effects of a set of mutations, among a given set of genotypes (*de novo* random mutants, standing variants at equilibrium, etc.), in a given environment and/or genetic background.
LSD: Limit spectral distribution (distribution of eigenvalues) of a matrix as its dimensions get large.
i.i.d: identically and independently distributed (a set of values all drawn independently from the same distribution).
$\boldsymbol{\phi}(.)$: Developmental function relating mutable traits ($\mathbf{x}$) and optimized traits ($\mathbf{y}$), see Figure 1.
b_{ij} : Pathway coefficients relating mutable trait j to optimized trait i [the first derivatives of $\phi(.)$ at the parent phenotypic position], gathered into the $n \times p$ matrix $\mathbf{B}$.
M-P law (MP(β, ζ)): Marchenko–Pastur law with ratio index β and scale parameter ζ.
$\mathbf{I}_k$ with $k \in \mathbb{N}^+$: Identity matrix in k dimensions.
$\tilde{\rho}_{\mathbf{X}}(x), \tilde{\nu}_{\mathbf{X}}(x)$: pdf of the LSD of the matrix $\mathbf{X}$ and corresponding Shannon transform, respectively (same conventions for all other transforms).

regarding the structure of phenotypic networks (Barabasi and Oltvai 2004) and some observations from experimental evolution. Provided these assumptions are valid, we will see how some laws of large numbers yield the isotropic FGM.

I tried, as much as possible, to keep the details of the phenotypic network and its very nature unspecified, to retain the generality of heuristic models. The model is intended to describe the DFE among mutations in a single gene or set of genes in the same functional complex. I discuss its extension to mutations scattered across the genome. To obtain the key results, I used tools from random matrix theory (Bai and Silverstein 2010), which provides a statistical description of large matrices whose elements are drawn from random distributions. Derivations of the results are given in Supporting Information, File S1 and File S2 and a *Mathematica* (Wolfram Research 2012) notebook (File S3) (in freely readable [.cdf] format). The main text is reserved for assumptions, arguments, and key results, and a glossary of notations is given in Table 1.

The observed *vs.* predicted patterns are illustrated on a set of fitness measurements among random single-nucleotide substitutions in two ribosomal protein genes of the bacterium *Salmonella typhimurium* (Lind *et al.* 2010). This is more intended as an illustration than a test of the model, the latter being tackled in the *Discussion*.

Methods

Biological and mathematical assumptions

I first describe the key biological features behind the model and their justification and present a heuristic argument behind the main results. In all of the following, I use the shortcut "phenotype" to mean the genetic value of a lineage for the phenotype considered (averaged over microenvironmental variation). The model relies on eight key assumptions: the first five are basic "biological" assumptions about the relationship between genotype, phenotype, and fitness, and the last three are more technical "mathematical" requirements of the model:

1. There is a fitness optimum for a subset of key traits.
2. The parent phenotype is not too far from the optimum.
3. Mutations have mild effects on phenotype.
4. Each mutation pleiotropically affects many "mutable" traits (high level of pleiotropy).
5. The large set of mutable traits in turn affects a smaller subset of key "optimized" traits that determine the optimum (high developmental integration).

These "basic" assumptions are detailed and justified below. Three additional (milder) assumptions bear on the general class of distributions that are considered here:

6. The distributions of all random variables considered must have finite mean and variance and satisfy the "Lindeberg condition" (see, *e.g.*, Barton and Coe 2009).
7. The covariance between dependent random variables in the model must satisfy a "weak dependence condition" for dependent variables (Baxter *et al.* 2007).
8. The $n\,p$ coefficients relating mutable to optimized traits form a multivariate distribution that can be written as a linear combination of $n\,p$ independently distributed variables.

Bluntly, assumptions 6 and 7 require that, although unspecified, the distributions considered in the model behave "nicely" so that we may use central limit theorems on these random variables. Assumption 8 ensures that we can transform the distribution of the coefficients to a canonical form used in random matrix theory. It still allows for most distributions of possibly correlated coefficients, but they cannot be fully correlated (which is anyway ensured by assumption 7). We detail later the implied properties for the variables under study.

Definition of different trait types: It is a tricky exercise to characterize the traits that are considered in Fisher's model (discussed in Orr 2000; Martin and Lenormand 2006b). Indeed, the focus is on fitness, not on traits of particular interest, and no explicit mechanism relates traits and fitness. Therefore, one can use an infinite number of trait definitions (*i.e.*, of coordinate systems in phenotype space). It is not the case here. Below, three distinct trait types are defined in a top–down order.

Fitness m: The obvious first trait to define is fitness or any fitness component (growth rate per unit time, survival probability, competitive index over some period of time, etc.). This is the only quantity that is considered measurable

empirically, for a given set of genotypes. I refer to "fitness" to mean any such measurable fitness component and denote by m (for Malthusian fitness) the breeding value of a lineage for fitness. Malthusian fitness is our landmark fitness component.

Optimized traits $\mathbf{y}$: A second class of traits is dubbed "optimized traits," whose breeding values for a given genotype are given by the vector $\mathbf{y}$. These traits are characterized as follows: (i) there is an optimizing function relating them to fitness (with a maximal fitness at some value of $\mathbf{y}$) and (ii) they are not fully correlated by mutation. These traits can be thought of as the traits defined in the FGM: I denote n (for consistency with the FGM) the number of optimized traits. The dimension n counts the number of traits that are jointly modified by any single mutation, but not fully correlated. Technically, optimized traits have a mutational covariance matrix that is of full rank n. A unique and mathematically justified definition of these traits comes later.

Mutable traits $\mathbf{x}$: Finally, mutation defines another subset of the organism's phenotype: the set of traits that is pleiotropically affected by mutations in a given genomic target (gene or set of genes). I call "mutable traits" the traits in $\mathbf{x}$ and denote by p (for "pleiotropy") their number. As with optimized traits, mutable traits have a mutational covariance matrix that is of full rank p. We assume that mutation effects on mutable traits are unbiased (mean is zero).

A larger phenotypic set of $n' > n$ optimized traits (resp. $p' > p$ mutable traits) could always be defined, but then some $n' - n$ (resp. $p' - p$) traits would be linear combinations of the first n (resp. p) ones.

Developmental function: We must now define arbitrary functions relating these traits. The function $m(\mathbf{y})$ maps a given phenotype $\mathbf{y}$ to fitness: it is unspecified, but must define an optimum in $\mathbf{y}$ space, which we can set at the origin $\mathbf{y} = \mathbf{0}$, without loss of generality. There is also a mapping from mutable to optimized traits: this integration is mediated by an unspecified cascade of developmental, physiological, regulatory, etc., processes. Following Wagner (1984, 1989) and Rice (2002, 2004), we can define an arbitrary multivariate function $\boldsymbol{\phi}(.)$ relating these phenotypic sets: $\boldsymbol{\phi}(\mathbf{x}) = \mathbf{y}$. We denote this function "developmental function" in reference to Wagner's introduction of the concept: indeed we retrieve his particular (linear) function in the limit.

Now all the ingredients in the model are defined: Figure 1 illustrates the genotype–phenotype–fitness mapping that is used in this article. So far, we made no assumption on the particular properties of these functions or traits, except a causal relationship and the existence of a phenotypic optimum (assumption 1). We will see below the biological justification behind the five basic assumptions 1–5 and their implications for the model.

Practical illustrations: First, let us consider a practical (but limiting) example with a mutation affecting an enzyme involved in metabolism, in a unicellular organism. Mutations at the focal gene modify the concentrations of the products and substrates of the reaction catalyzed by the enzyme and in turn modify many other metabolite concentrations, via the metabolic network. This in turn alters a set of key metabolites (ATP, NADPH, etc.) that directly determine the cellular growth rate via an optimizing function. In the microbial metabolic network models used in flux balance analysis (see, *e.g.*, Price *et al.* 2004), this optimization function is empirically defined and calibrated. In this example, the concentrations of the p metabolites modified by mutations are the mutable traits ($\mathbf{x}$), and those of the n metabolites determining cellular growth are the optimized traits ($\mathbf{y}$). The metabolic network relating these metabolite concentrations determines the developmental function $\boldsymbol{\phi}(\mathbf{x})$, and the optimizing function relating the key metabolites to cellular growth is the fitness function $m(\mathbf{y})$.

In metabolic theory, the optimizing function is constructed to define an optimum, but not necessarily with respect to all the factors that enter the function. For example, some metabolite concentrations may enter the function in a linear fashion but be determined by other metabolite concentrations via quadratic functions. In that case, the corresponding optimized traits would be these lower-level metabolite concentrations that do not enter the optimizing function explicitly.

In a recent study, Le Nagard *et al.* (2011) simulated a phenotype–fitness landscape with underlying phenotypes encoded by a neural network (mimicking a set of interacting genes), which itself determines fitness. This is akin to the type of landscape considered here, and, indeed, the authors analyzed their results using complexity definitions from the FGM. However, their approach differs from ours in that their fitness is a Gaussian function of the distance between a genotype's reaction norm (to some environmental variable) and some optimal reaction norm. On the contrary, our fitness function depends on distances from a single phenotypic optimum, in fixed conditions.

Existence of a fitness optimum: To define the set of optimized traits, I assume that there is, at a local scale in phenotype space, a phenotypic state $\mathbf{y} = \mathbf{0}$ that maximizes fitness. This idea, initially introduced by Fisher (1930) in his geometrical model, is the common ground of many evolutionary models of course, but it is also supported by some observations from long-term evolution experiments (Elena and Lenski 2003). In such experiments, the fitness of replicate populations reaches a plateau [which may sometimes vary (Schoustra *et al.* 2009)] or at least shows a striking deceleration. Consistent with this finding, negative epistasis among beneficial mutations has been repeatedly reported: beneficial mutations tend to be less advantageous when arising from fitter parents (MacLean *et al.* 2010; Chou *et al.* 2011 ; Khan *et al.* 2011; Sousa *et al.* 2011). These observations are obviously suggestive of the existence of an optimum for fitness, at some local scale at least.

Local approximation: Assumptions 2 and 3, respectively, state that the parental phenotype is not too far from the

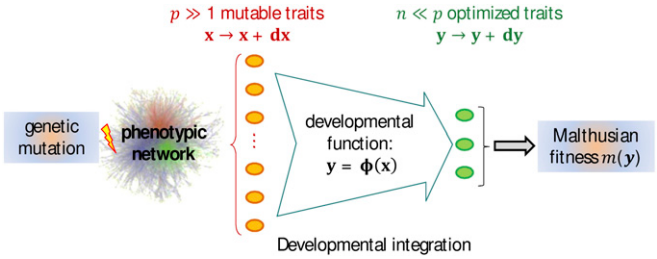

Figure 1 Model of genotype–phenotype–fitness map. This schematic representation shows the different levels of integration assumed in the model, from a single genetic change in the DNA (left) to its effect on the Malthusian fitness of the whole organism. Each mutation pleiotropically affects a large subset of p "mutable traits" (orange ovals), via a complex interaction network among proteins. The parent phenotype at all these traits is represented by the vector **x**. The effect of a mutation (on the offspring's phenotype) is a random small perturbation **dx**, with mean zero and arbitrary multivariate distribution (covariance **V**). These basic mutational changes "percolate" through the network of interactions to induce changes at a much smaller set of n key integrative traits ("optimized traits," green ovals), which are those under stabilizing selection, represented by the vector **y**. An arbitrary developmental function $\mathbf{y} = \boldsymbol{\phi}(\mathbf{x})$ relates mutable to optimized traits (developmental integration). The effect of mutations on **y** is a perturbation vector **dy** that is approximately linear in **dx** (to leading order), with linear coefficients $b_{ij} = \partial_{x_j}\phi(y_i)$, arbitrarily distributed. The optimized traits directly determine fitness via a locally quadratic function $m(\mathbf{y})$ around some optimum (set at $\mathbf{y} = 0$).

optimum and that mutation effects around this phenotype are mild. These are justified if we accept that, while there may be substantial adaptation going on, a genotype cannot be too far from a local optimum and remain viable. These ideas can be traced back to Fisher (1930) in the presentation of his geometrical model; we use them here to allow several key mathematical approximations.

Under these conditions, the entire population (parent plus mutants) lies in some vicinity of the optimum and will remain so over the course of the adaptive process. This allows a key simplification: we can derive the DFE from only the local behavior of the fitness function $m(\mathbf{y})$ about its optimum $\mathbf{y} = \mathbf{0}$. This local behavior is simply given by a Taylor-series approximation around the optimum. Assume that, in some local neighborhood, a real function $m : \mathbb{R}^n \to \mathbb{R}$ is continuous and defines a "nondegenerate" optimum (one whose second derivative is not vanishing, details below). Then, there always exists a unique set of (Cartesian) coordinates for the **y** space such that, in the vicinity of the optimum, the function can be written as $m(\mathbf{y}) = m(\mathbf{0}) - 1/2\|\mathbf{y}\|^2$, where $\|\mathbf{y}\|^2 = \sum_{i=1}^{n} y_i^2$ is the squared norm of **y**. The factor $1/2$ is merely for consistency with conventions in the FGM. This statement can be proved in two steps. First, a second-order Taylor-series approximation of any smooth fitness function about a (nondegenerate) optimum yields a quadratic function with positive-definite selective covariance matrix (Lande 1979). Second, one can use a linear transformation of this coordinate system so that the selective covariance matrix in the new system is equal to identity. The particular linear transformation is obtained by solving the generalized eigenvalue problem, (as, *e.g.*, in Martin and Lenormand 2006b; Chevin *et al.* 2010). Here, the resulting coordinate system has a locally quadratic and isotropic fitness function; namely, all resulting phenotypic directions are equivalently selected. The interested reader can relate this result to the more general "Morse lemma." Trait definitions are arbitrary in the FGM (Orr 2000; Martin and Lenormand 2006b), so we can choose to define **y** in those coordinates where, "locally," the function is isotropic and quadratic. This is possible because we are not focusing on particular traits, as typical quantitative genetics does, so the coordinate system can be arbitrarily "bent" and turned via this linear transformation. Once the coordinate system has been set in this manner, the traits in **y** are fully and uniquely characterized, contrary to the FGM in its original form. In the following, all other coefficients are defined in this particular set of coordinates.

The assumption that the optimum is "nondegenerate" deserves some development. It means that the function must be twice differentiable, with zero first derivatives at the optimum $[\partial_{y_i} m(\mathbf{0}) = 0]$ and nonzero second derivatives around this optimum $[\partial^2_{y_i} m(\mathbf{y}) < 0$, to define a maximum]. Most continuous functions have only nondegenerate critical points (maximum, minimum, or saddle point). However, the approximation cannot be invoked for some special optimization functions that happen to have been used in some previous versions of the FGM. Linear fitness functions of the form $m(\mathbf{y}) = m(\mathbf{0}) - \|\mathbf{y}\|$ (Poon and Otto 2000) or fitness functions of the form $m(\mathbf{y}) = \exp(-\|\mathbf{y}\|^Q)$ where $Q > 2$ (Tenaillon *et al.* 2007; Gros and Tenaillon 2009) have a "degenerate" optimum at $\mathbf{y} = \mathbf{0}$. The former function is not differentiable at $\mathbf{y} = \mathbf{0}$, while the latter has vanishing second derivatives at $\mathbf{y} = \mathbf{0}$. Note also that even discontinuous functions relating phenotype to growth rate can yield continuous Malthusian fitness functions once some nonheritable and continuous microenvironmental random component is accounted for. Figure 2 illustrates the types of fitness functions where the quadratic approximation does or does not apply. A wide variety of functions are allowed of course, like the Gaussian, asymmetric functions, etc.

The local approximation ("vicinity of the optimum") allows generalizations but imposes a limit: How close to the optimum must we remain to apply the approximation? As for any Taylor-series argument, the answer depends on the fitness function and cannot be general: the population must remain close enough to apply the approximation.

Pleiotropy within networks: Assumption 4 requires that the space of mutable traits has high dimension (high pleiotropy, $p \gg 1$): it allows the use of large-number arguments in the statistical treatment of the model. Some *a priori* and *a posteriori* arguments suggest that these assumptions are biologically reasonable. Note that I use a particular definition of pleiotropy relating to the whole set of possible mutants in a given genetic target; while some define pleiotropy at the scale of any particular mutation (Wagner and Zhang 2011), these two definitions should coincide if the distribution of mutation effects is continuous.

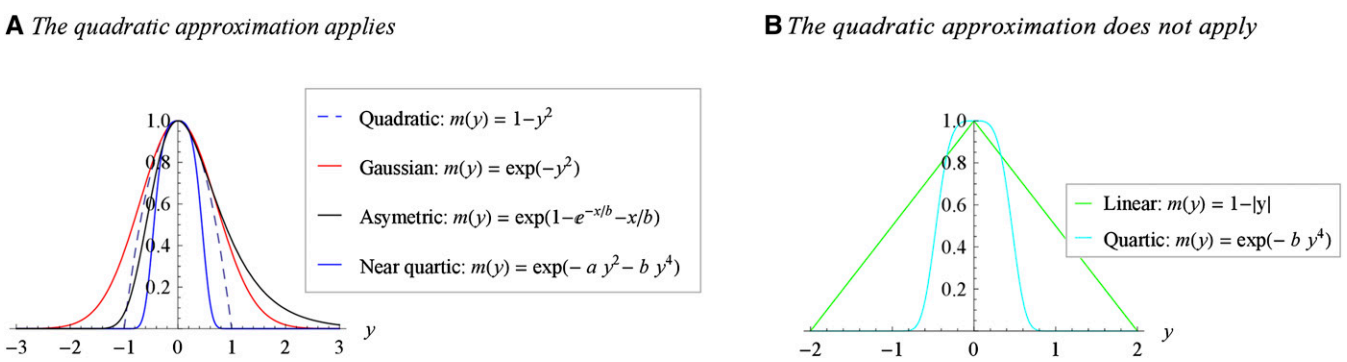

Figure 2 Applicability of the quadratic approximation around the optimum, for various fitness functions. Different fitness functions $m(y)$ (indicated in inset) are illustrated in one dimension ($y \in \mathbb{R}$). Conditional on the existence of a nonvanishing second derivative at the optimum ($m''(0) \neq 0$), the quadratic approximation (dashed line) applies (A) or does not apply (B).

First, a widely observed feature in most phenotypic networks [from gene regulation to protein interactions or metabolism (Barabasi and Oltvai 2004)] is the "small-world" property. This property implies that every node in the network is connected to most other nodes via a path of short length. We can thus expect that any mutation that affects a given node will in turn modify many other phenotypes through this set of short paths. This suggests that most mutations should be highly pleiotropic ($p \gg 1$, assumption 4). It does not, however, imply "universal pleiotropy": p can still be substantially smaller than the total number of phenotypes under genetic control (see the discussion in Wagner and Zhang 2011, 2012; Hill and Zhang 2012). Note that this small-world property is a *sufficient* condition to obtain high pleiotropy, but is not *necessary*.

Second, this *a priori* argument is reinforced by *a posteriori* empirical observations. The metabolome of the actinomycete *Nocardiopsis* has been shown to be widely modified by resistance mutations caused by single-nucleotide changes (Derewacz *et al.* 2013): more than 300 different metabolites expressed in the mutants are undetected in the wild type, and up to 80 metabolites in the wild type are undetected in the mutants. Note that mere changes in metabolite concentrations are not counted in this picture. RNA-chip studies have also revealed that the expression of many genes can be jointly altered by simple genetic changes, *e.g.*, gene deletions (Wagner 2002) or single fixed mutations in experimental evolution (reviewed in Hindré *et al.* 2012).

Developmental integration: Our last biological assumption 5 requires that the many mutable traits affect a much smaller subset of optimized traits ($n \ll p$). We refer to this assumption as "developmental integration" (although development may not be involved), because it refers to integration through the developmental function, in the sense of Wagner (1984, 1989) and Rice (2002, 2004). It might be the most difficult assumption to evaluate. It is known that oriented networks (like gene regulation networks, for example) are highly hierarchical over several levels, with a few central genes regulating large sets of other genes (Bhardwaj *et al.* 2010). This hints at integration within biological networks but remains not directly relevant to our model: what is assumed here is that the set of traits (not genes) that determine fitness is small and affected by many underlying traits. More directly, in flux balance analysis (our example above), the function relating metabolism to growth rate has typically a few tens of variables, much fewer than the whole set of metabolites in the system. Also, most of the metabolites in the system affect the key metabolites more or less directly. This assumption is also supported, although indirectly, by the observation that empirical DFEs are not normally distributed. Indeed, if there are many, not fully dependent, phenotypic traits that both mutate and affect fitness (via any function), then the resulting distribution should be close to Gaussian, by the central limit theorem. This is typically not observed: across several organisms, laboratories, and methods, empirical DFEs are skewed toward negative values (reviewed in Martin and Lenormand 2006b; Eyre-Walker and Keightley 2007). Empirical noise could not cause this observation because (i) it is fairly limited in several of these studies and (ii) noise should favor normally distributed observations, which are not observed.

Together, assumptions 4 and 5 are summarized in the notion of high developmental integration: from many traits connected by the network to fewer key traits selected for optimal intermediates ($p \gg n > 1$).

Emergence of the FGM

In all of the following, $\mathbf{X}.\mathbf{Y}$ denotes a matrix product, $\mathbf{X}^*$ denotes the matrix transposition of $\mathbf{X}$, and $\mathbf{X}^{1/2}$ denotes the Cholesky decomposition of $\mathbf{X}$ (matrix square root).

Distribution of mutation effects on optimized traits and fitness: Assumptions 1–4 suffice to obtain a landscape of the type described by an anisotropic Fisher model. Let us see how. From their very definition, mutations create a change in the mutable trait values $\mathbf{dx} = \{dx_j\}_{j\in[1,\ldots,p]}$, whose multivariate distribution is unspecified (continuous or discrete, etc.). Yet, we require that they have zero mean ($E(\mathbf{dx}) = \mathbf{0}$) and nonzero finite variance–covariance matrix $\mathbf{V} = E(\mathbf{dx}.\mathbf{dx}^*)$ and satisfy the Lindeberg and weak dependence conditions (our

assumptions 6 and 7). The random perturbation $\mathbf{dx}$ translates into $\mathbf{dy} = \{dy_i\}_{i\in[1,\ldots,n]}$, via the developmental function: $\mathbf{dy} = \boldsymbol{\phi}(\mathbf{x}+\mathbf{dx}) - \boldsymbol{\phi}(\mathbf{x})$. Assumption 3 (mild mutation effects) allows us to take a linear approximation of $\boldsymbol{\phi}(.)$ about the parent phenotype: $\mathbf{dy} \approx \mathbf{B}.\mathbf{dx}$, where $\mathbf{B} = \{b_{ij}\}_{i\in[1,n],j\in[1,p]}$ is an $n \times p$ matrix containing all the first derivatives of $\boldsymbol{\phi}(.)$ at the parent position $\mathbf{x}$. We thus retrieve exactly Wagner's (1984, 1989) linear developmental function, where the n p coefficients b_{ij} in $\mathbf{B}$ describe how mutable traits x_j integrate into optimized traits y_i. I denote them "pathway coefficients," as they relate to functional pathways connecting phenotypes.

Then, from assumptions 4–7, we can invoke a generalized central limit theorem (CLT). Each dy_i is a linear combination of many, not fully correlated, random variables (the dx_j): under certain conditions on the dependence between dx_j and on the nature of their distributions (which may vary across index j), these dy_i converge to a normal distribution as p gets large. Assumption 5, not just assumption 4, is required because it implies that the coefficients in $\mathbf{B}$ are not zero in a large proportion (p times the proportion of zero coefficients remains large). Assumption 6 applied to the dx_j guarantees that the distributions of the dx_j pertain to a class that does yield convergence to the CLT as $p \to \infty$. Technically, the Lindeberg condition is required when the dx_j are not all drawn from the same distribution. It is satisfied, *e.g.*, by any distribution whose fourth central moment scales with the squared variance (detailed in theorem 2.35 in Tulino and Verdù 2004) and more generally requires that higher moments be bounded (Baxter *et al.* 2007). This is an application of the narrower but simpler "Lyapunov condition." Assumption 7 states that the dependence between the dx_j is weak enough that the variance of their mean $V(d\bar{x})$ scales with $1/n$ as $n \to \infty$. It provides one among various sufficient conditions for the CLT to apply with dependent variables (Baxter *et al.* 2007). The conditions of convergence to the CLT are a vast and well-studied subject of probability theory that is obviously beyond the scope of this article. Our assumptions 6 and 7 thus simply require that we are under those conditions sufficient to apply the CLT, even to dependent and not identically distributed variables. This still encompasses a vast array of situations and distributions.

The CLT then implies that $\mathbf{dy}$ converges to a (multivariate) Gaussian as p gets large, with mean $E(\mathbf{dy}) = 0$ and covariance matrix $\mathbf{M} = E(\mathbf{dy}.\mathbf{dy}^*) = \mathbf{B}.\mathbf{V}.\mathbf{B}^*$, which is denoted $\mathbf{dy} \sim N(\mathbf{0}, \mathbf{M})$. Note that $\mathbf{M}$ may depend on the particular parent phenotype, because both $\mathbf{B}$ and $\mathbf{V}$ may depend on a position in $\mathbf{x}$ space, unless phenotype space for $\mathbf{x}$ is additive and the developmental function is linear. I drop the explicit reference to this fact for notational simplicity, but get back to it in the *Discussion*. Note also that this central limit theorem argument cannot be turned the other way around (from optimized to mutable traits). First, it is the mutable traits that are causally affected by mutation, by definition. Second, the developmental function $\boldsymbol{\phi}(.) : \mathbb{R}^p \to \mathbb{R}^n$ is not a bijection so we cannot define the inverse relationship $\mathbf{x} = \boldsymbol{\phi}^{-1}(\mathbf{y})$ that might yield a Gaussian distribution of the $\mathbf{dx}$.

Finally, as assumptions 1–3 yield a simple fitness function ($m(\mathbf{y}) \approx m(\mathbf{0}) - 1/2\,\|\mathbf{y}\|^2$), we can express the change in fitness induced by mutations as

$$s(\mathbf{dy} \mid \mathbf{y}) = m(\mathbf{y}+\mathbf{dy}) - m(\mathbf{y}) \approx -\left(\sum_{i=1}^{n} y_i\, dy_i + \frac{dy_i^2}{2}\right)$$
$$\mathbf{dy} \sim N(0, \mathbf{M}), \tag{1}$$

for a mutation with effect $\mathbf{dy}$ on optimized traits, arising in a parent with phenotype $\mathbf{y}$. In the special case where our fitness component is Malthusian fitness itself, s is exactly the selection coefficient of the mutation. It is also approximately so if $m = \log(W)$, where W is Darwinian fitness with discrete nonoverlapping generations. Otherwise, it describes only a linear change in the measured fitness component. Equation 1 corresponds to an anisotropic FGM (Martin and Lenormand 2006b): the distribution of the phenotypic effects of mutations is Gaussian, and the DFE is a quadratic form in Gaussian vectors, a well-characterized distribution (Mathai and Provost 1992). The difference is that the normality of phenotypic effects emerges from assumptions 3–7 and the central limit theorem, rather than being assumed from the start. Note also that the local approximation for m has reduced anisotropy to the mutational covariance: the selective covariance is $\mathbf{S} = \mathbf{I}_n$, the $n \times n$ identity matrix (because the fitness function is isotropic).

The DFE in Equation 1 can be expressed in simpler form (see, *e.g.*, Jaschke *et al.* 2004). Let λ_i be the n eigenvalues of $\mathbf{M} = \mathbf{B}.\mathbf{V}.\mathbf{B}^*$, all strictly positive by construction (since $\mathbf{V}$ and $\mathbf{B}.\mathbf{B}^*$ are positive definite). Let $\mathbf{Q}$ be the eigenbasis of $\mathbf{M}$ with column vectors given by the n eigenvectors of $\mathbf{M}$, such that $\mathbf{Q}^* = \mathbf{Q}^{-1}$ and $\mathbf{Q}^*.\mathbf{M}.\mathbf{Q} = \mathbf{diag}(\lambda_i)$ is diagonal. Let $\mathbf{z} = \{z_i\}_{i\in[1,n]} = \mathbf{Q}^*.\mathbf{y}$ be the projection of $\mathbf{y}$ in the eigenspace of $\mathbf{M}$ ($\mathbf{z}$ is simply $\mathbf{y}$ expressed in another basis for phenotype space). Finally, let $\chi_k^2[\nu]$ denote a noncentral chi-square deviate with k d.f. and noncentrality parameter ν. The DFE in Equation 1 can be written as a function of a set of independent known random variables (this is called a stochastic representation): we have

$$s(\mathbf{dy}|\mathbf{y}) \sim s_o - \sum_{i=1}^{n} \frac{\lambda_i}{2}\, \chi_1^2\left[\frac{z_i^2}{\lambda_i}\right], \tag{2}$$

where $s_o = \sum_{i=1}^{n} z_i^2/2 = \|\mathbf{z}\|^2/2 = \|\mathbf{y}\|^2/2$ is a constant with $\|\mathbf{z}\|^2 = \|\mathbf{y}\|^2$ the distance to the optimum from the parent position (in any orthonormal basis). The DFE is thus fully determined by the parental position in $\mathbf{y}$ space and the n eigenvalues λ_i of $\mathbf{M}$. We thus retrieve a known result from the FGM (*e.g.*, Martin and Lenormand 2006b), in the simpler case where $\mathbf{S}.\mathbf{M} = \mathbf{M}$.

A short comment on what follows is necessary at this point. Because the DFE in Equation 2 depends on the particular set of eigenvalues $\{\lambda_i\}_{i\in[1,n]}$, its parameters themselves are random: the λ_i's are inherently random in our

model, as the matrices **B** and **V** are random/unknown. The distribution of the eigenvalues λ_i is called the spectral distribution of **M** (it is random just as a sample distribution is in statistics). However, we will see that when $n,\ p \gg 1$, two simplifications occur. First, the DFE is approximately characterized by an *expectation* over the spectral distribution (the expectation is still random if the distribution is not fixed). Second, the spectral distribution itself proves to converge to a known limit distribution. Together these two points ensure that the DFE is well approximated by a deterministic limit distribution. The following section introduces tools to predict the spectral distribution of **M**, based only on assumptions 4 and 5 of developmental integration, plus mild mathematical conditions (assumptions 6–8) on the general class of distributions considered; most details are given in File S1.

Spectral distribution of M *and random matrix theory*

Rationale: The key principle behind our "statistical treatment" of the model is as follows. It seems impossible to make any *a priori* statement about the particular values of each pathway coefficient b_{ij} in **B** or each covariance among mutable traits in **V**. A way forward is to consider that these coefficients consist of a large set of draws from unknown distributions. Then, as we assume there are many such coefficients ($n\ p \gg 1$), some key properties of the landscape are approximately given by the expected outcome, averaged over coefficients. This is similar to the statistical physics approach and even more closely related to the infinite-allele approximation of quantitative genetics (Kimura 1965). From only distributional, not individual, properties of the microscopic variables, a resulting macroscopic quantity can still be predictable. In our case, the macroscopic variable is the DFE, and the microscopic variables are the underlying phenotypes, their interactions, and mutational properties.

The necessary tool is a description of the properties of the eigenvalues of large random matrices, whose entries are drawn at random following a given scheme. This is a field of probability theory of its own, known as random matrix theory (RMT); see Bai and Silverstein (2010) for a recent review. Our use of RMT here has its equivalent in wireless communication (Tulino and Verdù 2004) or the physics of large nuclei [from which it originates (Forrester *et al.* 2003)]; the aim is to model some macroscopic properties of a complex system with many elements interacting in a poorly known fashion. The central result from RMT is that the eigenvalues of many large random matrices, once properly scaled, are distributed according to simple predictable limits. As with the central limit theorem (to which RMT is related), these limits are largely independent of the very nature of the distribution of the entries, requiring only the same broad conditions to apply (*i.e.*, assumptions 6 and 7). In addition, it is well established that results from RMT converge quickly (Tulino and Verdù 2004): they already show reasonable accuracy with n, p of the order of 10.

Biologically, these statements imply that, even though the numerous parameters in the model are mostly unspecified (microscopic variables), their collective behavior results in a predictable spectral distribution for **M** (macroscopic variable); see File S1. This in turn results in a predictable DFE, as detailed in File S2. An obvious advantage is that this asymptotic approximation depends on much fewer parameters than the full model and gets more accurate as the system gets complex (as n and p get large).

Distribution of pathway coefficients: Let us first characterize the randomness in the entries of matrix $\mathbf{M} = \mathbf{B}.\mathbf{V}.\mathbf{B}^*$. Matrix $\mathbf{B} = \{b_{ij}\}_{i\in[1,n],j\in[1,p]}$ is an $n \times p$ matrix of pathway coefficients: for a given index $i \in [1, n]$, the line vectors of **B**, denoted $\mathbf{b}_i = \{b_{ij}\}_{j\in[1,p]} = \boldsymbol{\phi}_i'(\mathbf{x})$ given by the p derivatives, with respect to mutable phenotypes **x**, of the developmental function $\boldsymbol{\phi}_i(.)$ determining the optimized trait y_i. The n vectors $\mathbf{b}_i = \{b_{i1}, ..., b_{ip}\}$ can always be seen as n draws from an unspecified multivariate distribution in $\mathbb{R}^p$. The $\mathbf{b}_i$ have a given $1 \times p$ mean vector $\boldsymbol{\mu}_\mathbf{B} = E(\mathbf{b}_i)$ and a given positive definite $p \times p$ covariance matrix $\mathbf{C}_\mathrm{B} = \mathrm{E}((\mathbf{b}_i - \boldsymbol{\mu}_\mathbf{B}).(\mathbf{b}_i - \boldsymbol{\mu}_\mathbf{B})^*)$.

The nature of the distribution from which the vectors $\mathbf{b}_i$ are "drawn" remains fairly general; they must satisfy only the mild mathematical assumptions 6–8. This ensures that central limit arguments may apply (same as discussed for dx_i) and that we can express the b_{ij} as a linear combination of independently distributed variables h_{ij} (assumption 8). We additionally require that the entries of **B** do not include a large class at $b_{ij} = 0$ (this is implicit in assumption 5 of developmental integration): **B** is not too "sparse." The distribution of b_{ij} may vary across index j: the sets $\{b_{ij_1}\}_{i\in[1,n]}$ and $\{b_{ij_2}\}_{i\in[1,n]}$, with $j_1 \neq j_2$, can be drawn from distinct distributions, provided all distributions satisfy the Lindeberg condition (assumption 6, as for the dx_j above). As we have seen above for the dx_j, a narrower sufficient condition is that the fourth moment scale with the square of the variance.

By assuming that $\mathbf{C}_\mathbf{B}$ is positive definite, we ensure that assumptions 7 and 8 are satisfied: the b_{ij} are not fully correlated across columns j. However, I conjecture (see File S1B) that this entails a somewhat hidden limit: because $n \ll p$ (assumption 5), the pathway coefficients cannot be too correlated among rows i. There cannot be too much similarity between the pathways relating various x_j to each optimized trait y_i, because this would imply that $\mathbf{C}_\mathbf{B}$ becomes positive semidefinite. The effects of such correlations of the b_{ij} among rows i are tackled quickly in the *Discussion* and File S1B.

Under the broad conditions described above, the matrix $\mathbf{M} = \mathbf{B}.\mathbf{V}.\mathbf{B}^*$ has the structure of a "sample covariance matrix" (Bai and Silverstein 2010): it is of the form $\mathbf{M} = \mathbf{K}.\mathbf{K}^*$, where **K** has random entries (detailed in File S1B). Note, however, that there is no form of actual "sampling" going on here, of course.

Convergence to a limit spectral distribution: The key insight from RMT is that the spectral distribution of such a sample covariance matrix $\mathbf{M} = \mathbf{K}.\mathbf{K}^*$ is well approximated by a nonrandom (predictable) limit, when the dimensions (n, p) are large. This limit is called the limit spectral

distribution (LSD) of $\mathbf{M}$. We can define the spectral distribution of $\mathbf{M}$ by its unknown probability density function (pdf) denoted $\rho_{\mathbf{M}}(x), x \in \mathbb{R}^+$. The corresponding predictable limit when $n\,p \to \infty$ is the LSD of $\mathbf{M}$, with the pdf denoted $\tilde{\rho}_{\mathbf{M}}(x) : \rho_{\mathbf{M}} \Rightarrow \tilde{\rho}_{\mathbf{M}}$.

The Marchenko–Pastur law for standard sample covariance matrices: Our results rely heavily on a cornerstone of RMT, known as the Marchenko–Pastur (M-P) law. Mathematical details can be found in Tulino and Verdù (2004) and Bai and Silverstein (2010). Let us consider an $n \times p$ matrix $\mathbf{H}$ with "standardized" independent entries, namely with real entries independently drawn from (possibly distinct) distributions with zero mean and variance $1/n$, satisfying the Lindeberg condition (our assumption 6). As $n\,p \to \infty$ with $p/n = \beta$, the spectral distribution of $\mathbf{H}.\mathbf{H}^*$ converges to the M-P law, its limit spectral distribution. This result is independent of the nature of the parent distribution from which the entries h_{ij} are drawn (uniform, Gaussian, etc.). The M-P law has a simple analytic pdf, which depends only on $\beta = p/n$. The matrix $\mathbf{H}.\mathbf{H}^*$ may be scaled arbitrarily, by some constant ζ, yielding a "scaled M-P law," with two parameters β and ζ, which I refer to as $\mathrm{MP}(\beta, \zeta)$. A detailed presentation of this distribution is given in File S1A; I give only its pdf in the case $\beta > 1$, which is our focus here:

$$\rho_{\zeta\mathbf{H}.\mathbf{H}^*}(x) \Rightarrow \tilde{\rho}_{\zeta\mathbf{H}.\mathbf{H}^*}(x) = \frac{\sqrt{(b-x)(x-a)}}{2\,\pi\,\zeta\,x}, \qquad x \in [a,b]$$
$$a = \zeta(1-\sqrt{\beta})^2 \quad \text{and} \quad b = \zeta(1+\sqrt{\beta})^2. \tag{3}$$

The mean of this distribution is $E(\lambda) = \zeta\ \beta$, and its variance is $V(\lambda) = \zeta^2\beta$. A crucial point is that this distribution is bounded: $\zeta(1-\sqrt{\beta})^2 < \lambda < \zeta(1+\sqrt{\beta})^2$. This implies that, as β gets large (*i.e.*, when $p \gg n$), all the eigenvalues of $\zeta\,\mathbf{H}.\mathbf{H}^*$ converge to a constant value $\tilde{\lambda} = \zeta\,\beta$. This property is the basis of the convergence of the model to the isotropic FGM in the presence of developmental integration (assumptions 4 and 5).

Figure 3 shows the agreement between the spectral distribution of large simulated random matrices (see File S3) and the LSD in Equation 3. Note that these are not means over several simulations; *every* simulated matrix has spectral distribution approximately given by its LSD. The range of λ narrows as $\beta = p/n$ increases: this can be intuited by considering sampling covariance matrices. The M-P law describes the limit distribution, as n, p get large, of the eigenvalues of the covariance matrix from a sample of p random vectors, drawn from a multivariate distribution in n dimensions, with parent covariance proportional to identity. As p gets large (while n remains finite) the sample covariance converges to the actual covariance of the parent distribution, which is the identity in n dimensions: $\mathbf{H}.\mathbf{H}^* \to E(\lambda)\mathbf{I}_n$.

Note that, from theorem 3.6 on p. 47 in Bai and Silverstein (2010), Equation 3 applies to *any* distribution of the h_{ij} (with finite mean and variance) if they are drawn from the *same* distribution, without requiring the Lindeberg condition (our assumption 6). It extends to the case where they are drawn from *distinct* distributions, under the additional Lindeberg condition; see the details in theorem 3.10 on p. 51 in Bai and Silverstein (2010) and theorem 2.35 on p. 56 in Tulino and Verdù (2004).

File S1A summarizes a set of known results on the M-P law (details can be found in Tulino and Verdù 2004). File S1B applies it to the problem at hand: first, showing that the mutational covariance matrix $\mathbf{M} = \mathbf{B}.\mathbf{V}.\mathbf{B}^*$ is a sample covariance matrix and then deriving its LSD. In short, $\mathbf{M}$ can be equated to $\mathbf{M} = \mathbf{K}.\mathbf{K}^*$, where $\mathbf{K} = \mathbf{H}.\mathbf{W}^{1/2} + \mathbf{U}_{\mathbf{K}}$ is "linearly" related to a standardized random matrix $\mathbf{H}$. This is ensured by our assumption 8. Matrix $\mathbf{H}$ is as described above: an $n \times p$ matrix $\mathbf{H}$ with entries independently drawn from (possibly distinct) distributions with zero mean and variance $1/n$. Matrix $\mathbf{W} = n\mathbf{C}_{\mathbf{B}}^{1/2}.\mathbf{V}.\mathbf{C}_{\mathbf{B}}^{1/2}$ is a positive definite $p \times p$ matrix whose eigenvalues are the same as those of $n\mathbf{V}.\mathbf{C}_{\mathbf{B}}$ and are finite. Matrix $\mathbf{U}_{\mathbf{K}}$ is an $n \times p$ matrix of rank 1, with unique singular value equal to $\theta = \sqrt{\boldsymbol{\mu}_{\mathbf{B}}.\mathbf{V}.\boldsymbol{\mu}_{\mathbf{B}}}$. An approximation for the LSD of $\mathbf{M}$ is then obtained as a scaled M-P law.

Results

Approximation to the spectral distribution of M

This section gives an approximation for the LSD of $\mathbf{M}$ and then moves toward more simplification, looking for situations where all $\lambda_i \approx \tilde{\lambda}$ converge to a single positive constant (convergence to isotropy).

In the absence of developmental or mutational correlations: A first simple statement is that the M-P law is exactly the LSD of $\mathbf{M}$ when $\mathbf{K} = \mathbf{H}$. This happens whenever the pathway coefficients are unbiased ($\boldsymbol{\mu}_{\mathbf{B}} = \mathbf{0}$) and both mutation effects on $\mathbf{x}$ and pathway coefficients are independent with equal variance ($\mathbf{V}.\mathbf{C}_{\mathbf{B}} \propto \mathbf{I}_p$, the identity matrix in p dimensions). Then $\mathbf{M} = \zeta\,\mathbf{H}.\mathbf{H}^*$, where ζ is a scaling constant. Its LSD is the M-P law $\lambda \sim \mathrm{MP}(p/n, \zeta)$, whose pdf is given by Equation 3: $\rho_{\mathbf{M}}(x) \Rightarrow\ \tilde{\rho}_{\mathbf{M}}(x) = \tilde{\rho}_{\mathbf{H}.\mathbf{H}^*}(x)$, as $n, p \to \infty$.

General case with $\boldsymbol{\mu}_B = 0$: Let us now consider the more general model, with arbitrary covariance matrices $(\mathbf{V}, \mathbf{C}_{\mathbf{B}})$, but still assuming that the distribution of pathway coefficients is unbiased, so that $\boldsymbol{\mu}_{\mathrm{B}} = \mathbf{0}$. Then $\mathbf{M} = \mathbf{H}.\mathbf{W}.\mathbf{H}^*$ ($\mathbf{K} = \mathbf{H}.\mathbf{W}^{1/2}$) and its LSD are no longer an M-P law. Yet, it is shown in File S1B that, with high developmental integration ($n \ll p$), the LSD of $\mathbf{M}$ is still approximately an M-P law, with coefficients modified to account for mutational/developmental covariances ($\mathbf{V}.\mathbf{C}_{\mathbf{B}} \neq \mathbf{I}_p$). The impact of these covariances on $\tilde{\rho}(\lambda)$ can be predicted from the sole knowledge of the mean and coefficient of variation of the eigenvalues $\lambda_{\mathbf{W}}$ of $\mathbf{W}$. I denote the mean of these eigenvalues $E(\lambda_{\mathbf{W}}) = \zeta_{\mathbf{W}}$ and their coefficient of variation $cv_{\mathbf{W}}^2 = V(\lambda_{\mathbf{W}})/E(\lambda_{\mathbf{W}})^2$. The LSD of $\mathbf{M}$ is approximately

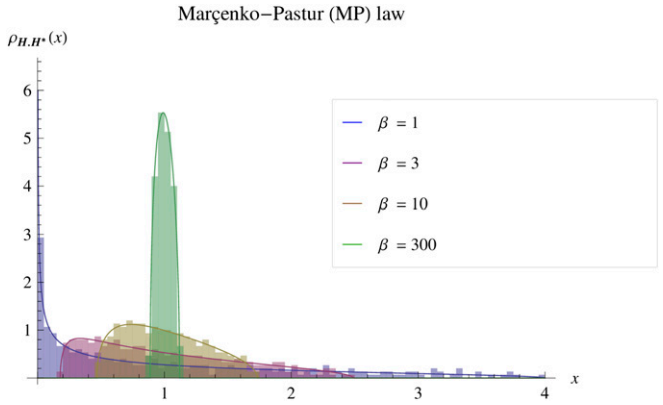

Figure 3 Convergence of the Marchenko–Pastur law toward isotropy. The pdf $\rho(x)$ of the Marchenko–Pastur (M-P) law is illustrated for a set of values of the shape parameter $\beta = p/n$ (see inset). The matrices are scaled to have $E(\lambda) = 1$ ($\zeta = 1/\beta$). The histograms show the spectral distribution of a simulated Wishart matrix ($n = 300, \beta = p/n$), and the lines show the limit spectral distribution given by the corresponding Marchenko–Pastur law $\lambda \sim \mathrm{MP}(\beta, 1/\beta)$ (Equation 3). The spectral distribution is well captured by the M-P law and becomes narrower as the ratio index becomes larger ($n \ll p$, high integration).

$$\lambda \sim \mathrm{MP}\left(\frac{p_e}{n}, \zeta_e\right), \quad \text{where} \quad \begin{cases} p_e = \dfrac{p}{1 + cv_{\mathbf{W}}^2} \\ \zeta_e = \zeta_{\mathbf{W}} \ \ p/p_e. \end{cases} \tag{4}$$

The effective parameters p_e and ζ_e account for the effect of multiplication by $\mathbf{W} = n\ \mathbf{C_B}^{1/2}.\mathbf{V}.\mathbf{C_B}^{1/2}$. The phenotypic covariances within the network, both among mutable traits ($\mathbf{V}$) and among pathway coefficients ($\mathbf{C_B}$), jointly affect the system but simply by reducing the effective ratio index of the M-P law, by a factor $1/(1 + cv_{\mathbf{W}}^2)$. The effect $cv_{\mathbf{W}}^2$ is the same as that seen in Figure 3 when β becomes smaller: a widening of the spectral distribution, namely an increase in anisotropy among dimensions. Both $\mathbf{V}$ and $\mathbf{C_B}$ contribute to increase this anisotropy through $cv_{\mathbf{W}}^2$.

Full model with arbitrary $\boldsymbol{\mu_B}$: It remains to state how the presence of a bias in the distribution of pathway coefficients b_{ij} affects the LSD of $\mathbf{M}$. This is detailed in File S1B. We have seen that $\mathbf{M} = \mathbf{K}.\mathbf{K}^*$, where $\mathbf{K} = \mathbf{H}.\mathbf{W}^{1/2} + \mathbf{U_K}$, so the only effect of the bias is to add the rank 1 matrix $\mathbf{U_K}$ to the model described by Equation 4. To gain an intuition for the effect of this bias, consider an extreme case. If the entries in $\mathbf{U_K}$ are much larger than those in $\mathbf{W}$, then $\mathbf{K} \approx \mathbf{U_K}$. In this situation, the mutational covariance is driven by the term due to bias ($\mathbf{M} = \mathbf{K}.\mathbf{K}^* \approx \mathbf{U_K}.\mathbf{U_K^*}$), yielding a covariance matrix of rank 1 (because $\mathbf{U_K}$ is of rank 1), namely with a single dominant direction. More precisely, this should happen when $\mathrm{Tr}(\mathbf{U_K}.\mathbf{U_K^*}) = \boldsymbol{\mu_B}.\mathbf{V}.\boldsymbol{\mu_B} \gg \mathrm{Tr}(\mathbf{H}.\mathbf{W}.\mathbf{H}^*) = \mathrm{Tr}(\mathbf{V}.\mathbf{C_B})$, where $\mathrm{Tr}(.)$ stands for matrix trace.

This effect of a small rank in the mutational covariance was described in detail in Chevin *et al.* (2010). Now the actual situation is less extreme, because $\mathbf{M}$ is positive definite, so that all eigenvalues are nonzero. The model must therefore behave in between one in n dimensions and one in one dimension.

A more explicit treatment of the effect of $\boldsymbol{\mu_B}$ on the LSD of $\mathbf{M}$ is given in File S1B, using a recent result by Benaych-Georges and Nadakuditi (2011). It can be summarized as follows. The bias affects only the leading eigenvalue λ_1, which shows a phase transition behavior determined by the ratio

$$cv^2 = \frac{\mathrm{Tr}(\mathbf{V}.\mathbf{C_B})}{\boldsymbol{\mu}_{\mathbf{B}}^*.\mathbf{V}.\boldsymbol{\mu_B}}. \tag{5}$$

This corresponds to our intuition: the effect of bias becomes effective whenever $\mathbf{U_K}$ affects $\mathbf{K}$ substantially. The parameter cv is akin to a coefficient of variation of the pathway coefficients: when the bias is small relative to the variance in pathway coefficients, $\boldsymbol{\mu}_{\mathbf{B}}^*.\boldsymbol{\mu_B} \ll \mathrm{Tr}(\mathbf{C_B})$ and $cv^2 \to \infty$. The phase transition occurs at $cv^2 = \sqrt{n\ p_e}$. When $cv^2 > \sqrt{n\ p_e}$, all eigenvalues fall into the M-P law, so that $\lambda_1 \to (1 + \sqrt{\beta_e})^2 \zeta_e$ (the upper bound of the M-P law in Equation 4). When $cv^2 < \sqrt{n\ p_e}$, λ_1 rises above the bulk of smaller eigenvalues, which remain under the M-P law. This can be summarized by the relative value of $\lambda_1 = \max(\lambda)$, over the mean of the bulk ($\tilde{\lambda} = E(\lambda_{i>1}) = \beta_e \zeta_e$). Beyond this phase transition, $\lambda_1 > \tilde{\lambda}$ and there is a favored direction in the mutant phenotype space (the direction associated with the first eigenvector). The factor of increase of the dominant eigenvalue, $\alpha = \lambda_1/\tilde{\lambda}$ satisfies

$$\alpha = \frac{\lambda_1}{\tilde{\lambda}} \approx \begin{cases} \left(1 + \dfrac{n}{cv^2}\right)\left(1 + \dfrac{cv^2}{p_e}\right) \underset{p_e \to \infty}{\to} \left(1 + \dfrac{n}{cv^2}\right), cv^2 < \sqrt{n\ p_e} \\ \left(1 + \sqrt{p_e/n}\right)^2 n \Big/ p_e \ \underset{p_e \to \infty}{\to} 1, cv^2 > \sqrt{n\ p_e}. \end{cases} \tag{6}$$

Putting all these results together yields the following spectral distribution for $\mathbf{M}$ in the general case:

$$\begin{aligned} \lambda_i &\sim \mathrm{MP}\left(\frac{p_e}{n}, \zeta_e\right) i \in [2, n] \\ \lambda_1 &= \frac{\alpha\, \zeta_e\, p_e}{n}, \end{aligned} \tag{7}$$

where the effective parameters p_e and ζ_e are given by Equation 4 and the factor α is given by Equation 6.

Equations 4–7 are fairly general as they rely on the general results of RMT. They apply for arbitrary distribution(s) of the pathway coefficients, provided they satisfy the (mild) assumptions, 6–8. They also make no assumption on the heterogeneity of the eigenvalues of $\mathbf{W}$, provided it remains finite, so that p_e remains $> n$ in Equation 4 (see File S1B). They do rely on the key assumptions that n and p are large enough to apply these asymptotic limits and that developmental integration is high ($\beta = p/n \gg 1$), namely on assumptions 4 and 5.

Figure 4 shows the accuracy of Equation 7 in approximating the actual spectral distribution of **M** in different situations. The example uses $n = 100$ for visual clarity of the histograms, but the convergence is much quicker, *e.g.*, with n of the order of 10 and p of the order of 100. The general spectral distribution of **M** is illustrated in Figure 4A, showing that the bulk of the $n-1$ lowest eigenvalues ($\lambda_{i>1}$) converges to the M-P law approximation, independently of the nature of the distribution of the b_{ij}'s and of its mean $\boldsymbol{\mu}_{\mathbf{B}}$. For this example, I used a bimodal discrete distribution ($b_{ij} = \pm 1/\sqrt{n}$), a normal ($b_{ij} \sim N(0, 1/\sqrt{n})$), a uniform ($b_{ij} \sim U(-3/\sqrt{n}, 3/\sqrt{n})$), or a mixture of the latter two (with roughly half of the coefficients drawn from the uniform and the other half into the normal). The corresponding behavior of the dominant eigenvalue λ_1, as a function of *cv*, is illustrated in Figure 4B. The simulations and figures were generated in File S3. Different distributions of the b_{ij} do not affect the spectral distribution of **M**, which is always well predicted by the M-P law approximation in Equation 7 (bulk by Equation 4 and λ_1 by Equation 6).

Distribution of fitness effects and isotropic approximations

Isotropic approximation below the phase transition: Let us assume that $cv^2 > \sqrt{n\ p_e}$ (Equation 6), namely that we are below the phase transition where λ_1 rises above the lower eigenvalues. Once the eigenvalue distribution of **M** is worked out, the DFE is obtained by a relatively straightforward argument, in the limit $\beta \to \infty$ (high developmental integration). In this case, the effective ratio index must also become large ($\beta_e = \beta/(1 + \mathrm{cv}_{\mathbf{W}}^2) \to \infty$, Equation 4). From the properties of the M-P law (and of the M-P law approximation in Equation 7), all the eigenvalues of **M** then converge to a single limit: $\lambda_i \to \tilde{\lambda} = \zeta_e \beta_e$ (see Figure 3). We denote this result the isotropic approximation because it boils down to isotropy in the FGM: all directions in **y** space become equivalent. This isotropic approximation is detailed in File S2. Note that, although framed as a simplistic and extreme limit here, this approximation involves more mathematical subtleties than meet the eye. The key quantity to describe the DFE as a function of the LSD of **M** happens to be robust to substantial variation across λ_i; this key technical point is illustrated in Figure S1. This is why the simplistic isotropic approximation ends up being accurate in situations where anisotropy is in fact substantial.

In the limit of $p/n \to \infty$, the isotropic approximation is equivalent to replacing $\lambda_i = \tilde{\lambda}$ for all i: recalling that $s_o = \|\mathbf{y}\|^2/2$, the stochastic representation in Equation 2 then reduces to

$$s \sim s_o - \frac{\tilde{\lambda}}{2}\,\chi_n^2\left[\frac{2s_o}{\tilde{\lambda}}\right], \qquad cv^2 > \sqrt{n\ p_e}, \tag{8}$$

namely a constant minus a noncentral chi-square deviate with n d.f. and noncentrality $2s_o/\tilde{\lambda}$. This distribution has an analytical pdf: letting $f_n(x, \nu)$ denote the pdf of the noncentral chi square with n d.f. and noncentrality ν, the pdf of the DFE in Equation 8 is

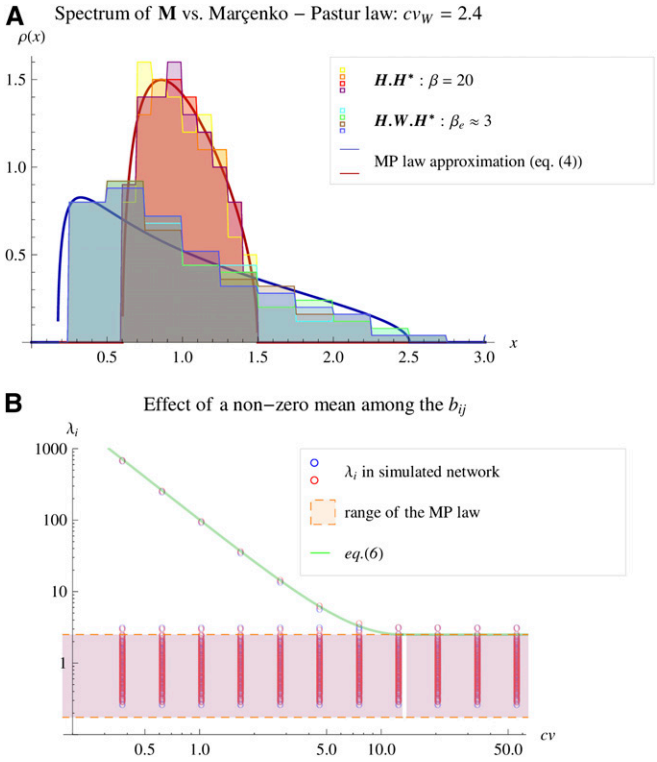

Figure 4 Spectral distributions under the random phenotypic network model. The observed distribution of eigenvalues of **M** (spectral distribution) is shown together with the predicted M-P law approximation. In each case, a single random phenotypic network (matrices **H**, **B**, and **V**) was drawn and the resulting spectral distribution of **H**.**W**.**H*** is shown. The $n \times p$ matrix of pathway coefficients $\mathbf{B} = \{b_{ij}\}_{i\in[1,n],j\in[1,p]}$ was set to have mean vector $\boldsymbol{\mu}_{\mathbf{B}} = \{\mu_j\}_{j\in[1,p]}$ drawn randomly with $\mu_j \sim N(0, \sigma)$ and $p \times p$ covariance matrix $\mathbf{C}_{\mathbf{B}}$ The pathway coefficients b_{ij} were drawn from various alternative distributions: discrete bimodal ($\pm 1/\sqrt{n}$), normal, uniform, or a mixture of the two. The covariance matrices $\mathbf{C}_{\mathbf{B}}$ and **V** were drawn independently as Wishart deviates: $\mathbf{V},\ \mathbf{C}_{\mathbf{B}} \sim W_{p+1}(\boldsymbol{\Lambda}_p)/(p+1)$, where $\boldsymbol{\Lambda}_p$ is a diagonal matrix. The matrix $\boldsymbol{\Lambda}_p$ has p gamma-distributed diagonal entries, thus allowing us to set a high coefficient of variation ($\mathrm{cv}_{\mathbf{W}}$) (A) of the eigenvalues of **W** (*i.e.*, of n $\mathbf{C}_{\mathbf{B}}$.**V**). The matrices were scaled so that $E(\lambda_{i\neq 1}) = 1$ in the bulk and the dimensions were $n = 100$ and $p = 2000$. A shows the resulting distribution of the bulk eigenvalues $\lambda_{i\neq 1}$ of either **H**.**H*** or **H**.**W**.**H*** (see inset) and the corresponding M-P law as solid lines. Each histogram color corresponds to a distinct distribution of the b_{ij}. B shows the behavior of the dominant eigenvalue λ_1. The coefficient *cv* (Equation 5) is set in the simulations via the scaling parameter σ. For each value of *cv* on the *x*-axis, the full set of eigenvalues is represented by the circles. Red and blue circles correspond to b_{ij} drawn from normal or uniform distributions (undistinguishable). The range of the bulk ($\lambda_{i\neq 1}$, orange area) and the dominant eigenvalue λ_1 (green line) are well predicted by the M-P law approximation.

$$f_s(s, n, \tilde{\lambda}, s_o) = \frac{2}{\tilde{\lambda}}\, f_n\left(\frac{2(s_o - s)}{\tilde{\lambda}}, \frac{2s_o}{\tilde{\lambda}}\right), cv^2 > \sqrt{n\ p_e},\ \ s \leq s_o. \tag{9}$$

This distribution depends on just three parameters $(n, \tilde{\lambda}, s_o)$, which constitutes a striking reduction in the parameterization of the model. Most importantly, there is no directionality effect on the DFE: the mutant selection coefficients have the same distribution, irrespective of the direction to the

optimum from the ancestor position. Only the overall distance to the optimum ($s_o = \mathbf{y}^2/2$) counts: this is the isotropic model in its exact form. At the optimum ($s_o = 0$), this DFE reduces to the negative gamma distribution $s_* \sim -\Gamma(n/2, \tilde{\lambda})$, where the "*" refers to the fact that this is the DFE at the optimum.

Beyond the phase transition: When $\mathrm{cv}^2 < \sqrt{n\,p_e}$ (Equation 6), anisotropy arises because, even with $\beta_e \gg 1$, the leading eigenvalue λ_1 can be strikingly higher than the others (see Figure 4B). This generates a favored direction in phenotype space where mutants tend to arise preferentially. However, this anisotropy is still of a relatively simple form, because the $n-1$ lower eigenvalues can still be equated to a constant whenever $\beta_e \gg 1$. We thus retain a form of isotropic approximation. We can then replace $\lambda_i = \tilde{\lambda}$ for all $i > 1$, while $\lambda_1 = \alpha\ \tilde{\lambda}$, and the stochastic representation in Equation 2 becomes

$$s \sim s_o - \frac{\tilde{\lambda}}{2}\left(\chi^2_{n-1}\left(\frac{2s_{n-1}}{\tilde{\lambda}}\right) + \alpha\ \chi^2_1\left(\frac{2s_1}{\alpha\,\tilde{\lambda}}\right)\right), \qquad (10)$$

where $s_{n-1} = \sum_{i=2}^{n} z_i^2/2$ and $s_1 = z_1^2/2$. It is easily checked that, below the phase transition ($\alpha = 1$), we retrieve our previous result (Equation 8), by properties of the noncentral chi-square distribution [$\chi^2_{n-1}(a) + \chi^2_1(b) \sim \chi^2_n(a+b)$ for any $a, b > 0$]. The two quantities s_1 and s_n depend on the parental position ($\mathbf{y}$ or $\mathbf{z}$) relative to the optimum, not just on its distance s_o. They sum up to the total maladaptation: $s_{n-1} + s_1 = \sum_{i=1}^{n} z_i^2/2 = s_o$. They are the contributions to maladaptation of the projections of the parental phenotype ($\mathbf{y}$) onto two orthogonal and complementary phenotypic subspaces: the eigenspace associated with the $n-1$ lowest eigenvalues for s_{n-1} and that associated with the dominant eigenvalue for s_1. This time, directionality is introduced: the same overall maladaptation (a given s_o) can result in a different DFE, according to the relative contributions of s_{n-1} and s_1. Note that, although more complex than Equation 9, this distribution still has only five parameters ($s_{n-1}, s_1, \alpha, n, \tilde{\lambda}$).

Unfortunately, I could not derive a closed-form pdf for this general model, so it would have to be computed as a convolution of two noncentral chi-square pdfs, by numerical integration. A pdf could be derived in the subcase where the ancestor is optimal, so that all mutations are deleterious ($s_o \to 0, s = s_*$). The DFE then becomes a sum of two independent negative gamma deviates $s_* \sim -\Gamma\left((n-1)/2, \tilde{\lambda}\right) - \Gamma(1/2,\ \alpha\ \tilde{\lambda})$: using a result from Di Salvo (2008), its pdf is

$$f^*_s(s, n, \tilde{\lambda}, \alpha) = \left(\frac{e^{s/\tilde{\lambda}}\left(-s/\tilde{\lambda}\right)^{n/2}}{-s\,\Gamma(n/2)}\right) {}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \frac{-s}{\tilde{\lambda}}\left(1 - \frac{1}{\alpha}\right)\right) \frac{1}{\sqrt{\alpha}}, \qquad (11)$$

where ${}_1F_1(.,.,.)$ is the Kummer confluent hypergeometric function. The left-hand factor in Equation 11 is the pdf of the negative gamma distribution $-\Gamma(n/2, \tilde{\lambda})$, which is the DFE when $\alpha = 1$ and $s_o = 0$. Consistently, the right-hand term in Equation 11 (hypergeometric function factor) simplifies to 1 when $\alpha = 1$.

Figure 5 shows the effect of the phase transition and the agreement between theory and exact simulations of the DFE with randomly drawn parameters (pathway coefficients b_{ij}, mutational covariance $\mathbf{V}$, and mutation effects $\mathbf{dx}$). The simulations and Figure 5 were generated in File S3, which can be used to generate similar figures and checks at will. The agreement between Equations 9–11 and the simulated DFEs is good, whether close to or far from the optimum, and both beyond and below the phase transition (Figure 5, B–D). As expected (Equation 4 and File S1), heterogeneity in the eigenvalues of $\mathbf{W}$, which is strong in this example ($\mathrm{cv}_\mathbf{W}^2 \approx 2$), does not affect the agreement with theory, as long as $p \gg 1$. Figure 5, B–D, also shows the fast convergence to the asymptotic result, as $n, p \to \infty$: here $n = 6$ and $p = 200$. Figure 5B illustrates visually the form of anisotropy that is generated by the phase transition for α: a single favored direction emerges, and the model converges more and more to that in one dimension as α gets larger, as suggested by our initial intuition. The consequence for the DFE is shown in Figure 5D.

File S2 details these approximations and why they prove accurate even though the actual model can be relatively far from isotropic. It also provides approximations for the first moments of the DFE.

Fitting empirical DFEs among random single-nucleotide substitutions

To illustrate how these results can be used, I fitted the observed DFE among random single-nucleotide substitutions introduced into two ribosomal genes of the bacterium *S. typhimurium* (Lind *et al.* 2010). In this study, a large set of mutants was created by site-directed mutagenesis, and their selection coefficient in competition (at 1:1 ratio) was estimated with high precision (detection limit $|s| < 10^{-3}$). I neglected the measurement error in this analysis, whose goal is merely to evaluate the qualitative agreement between theory and data in one example. I fitted the distribution of s among both synonymous and nonsynonymous mutations, because they showed no significant difference in DFE in this study (Lind *et al.* 2010). As no beneficial mutations were observed (suggesting $s_o = 0$), I fitted Equation 11 to the observed distribution of s by maximum likelihood, using R (R Development Core Team 2007). A first fitting procedure was performed on the scaled values [of $s/E(s)$], to find best-fitting values of n and α on scale-free data. Then a full maximization was performed starting with the best-fitting values thus obtained to jointly estimate $(\hat{\lambda}, \hat{n}, \hat{\alpha})$.

The results of the fits are given in Table 2, and the log-likelihood profiles for the parameters $\hat{n}$ and $\hat{\alpha}$ are shown in Figure S3. The resulting fitted distributions are illustrated in Figure 6, showing both the pdf and the cumulative distribution function (cdf). The gamma distribution (Equation 11 with $\alpha = 1$) already performs well, a pattern already shown by the original authors (Lind *et al.* 2010). The two ribosomal

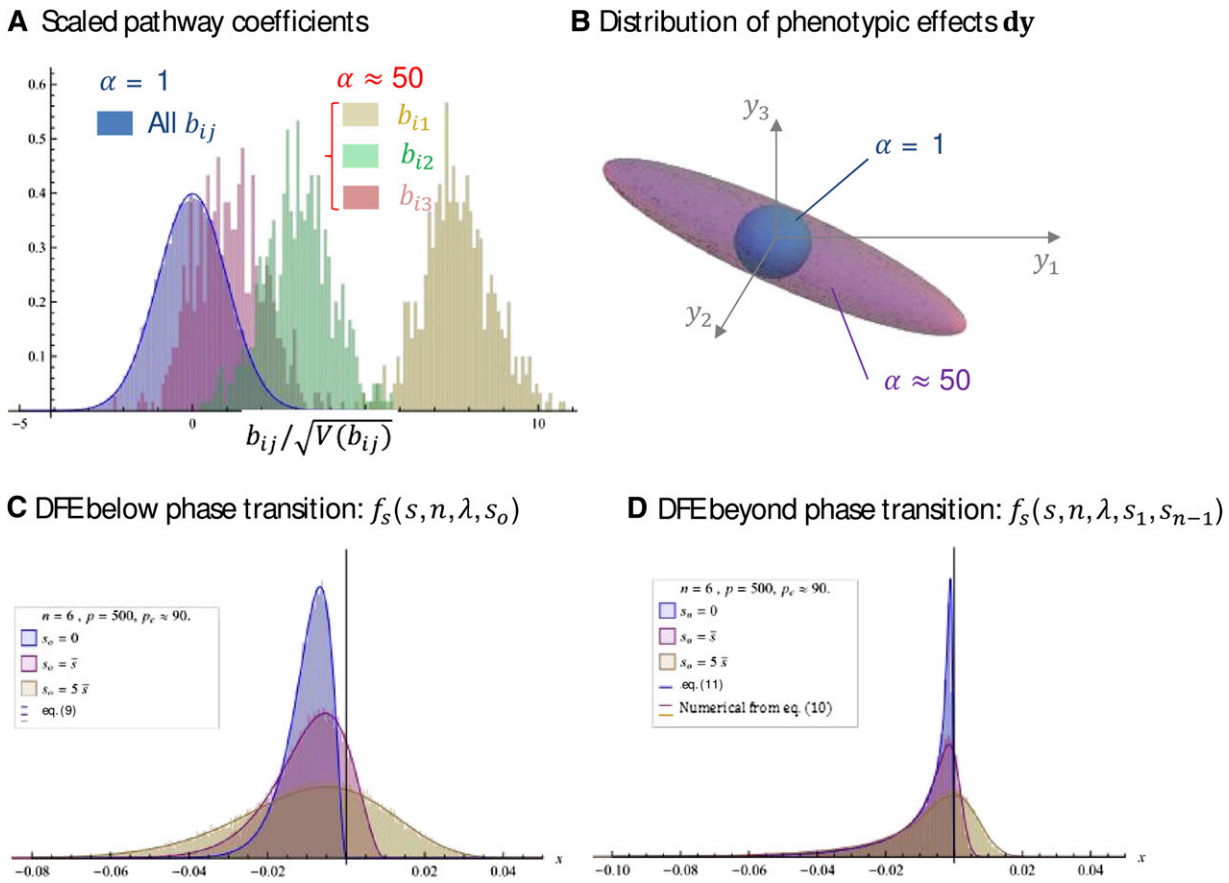

Figure 5 Simulated DFE *vs.* isotropic approximation(s). The distribution of selection coefficients among random mutations is represented for different situations (see File S3). The network was simulated as in Figure 4. The bias $\boldsymbol{\mu}_\mathbf{B}$ was scaled to obtain a given level of α: $\alpha = 1$ (C) or $\alpha = 50$ (D). The mutable traits **dx** were drawn as i.i.d. uniform deviates and scaled to obtain $V(\mathbf{dx}) = \mathbf{V}$. Then $\mathbf{dy} = \mathbf{B}.\mathbf{dx}$ was computed and scaled to enforce $E(s) = -0.01$. The n parental coordinates y_i were drawn as standard normal deviates and scaled to enforce a given $\|\mathbf{y}\|^2/2 = s_o$. The fitness effect of mutations was then computed from **y** and **dy** according to Equation 1. All other parameters are indicated in the insets: $n = 6$, $p = 200$. A illustrates the distribution of pathway coefficients, scaled by their variance ($b_{ij}/\sqrt{V(b_{ij})}$) for each situation: below the phase transition [$\alpha = 1$: all b_{ij} under the same blue histogram $N(0,1)$] or beyond the phase transition ($\alpha = 50$: each color gives the b_{ij} for a given index j for the three first mutable traits). The corresponding anisotropy in the spectral distribution of **M** is illustrated in B: ellipsoids are the 95% domain of the mutant phenotypes on the three first optimized traits y_i. Beyond the phase transition ($\alpha = 50$: purple cloud), there is a favored direction in **y** space. C and D show the corresponding DFE, below (C) or beyond (D) the phase transition, at the optimum ($s_o = 0$, blue), close to it ($s_o = \bar{s}$, red), or farther from it ($s_o = 5\ \bar{s}$, brown). The lines show the corresponding analytic predictions [Equation 9 (C) or Equations 10 and 11 (D)] (see insets).

genes yield similar estimates for $\hat{n}$ and $\hat{\lambda}$. However, in the *rpsT* gene, a value of $\hat{\alpha}$ significantly >1 ($\hat{\alpha} = 4.14$) was detected, while no such improvement in fit was obtained by letting $\alpha > 1$ for the *rplA* gene.

Overall, these results suggest that the prediction proposed here provides a good agreement with empirical DFEs. It is interesting to note that (i) allowing for an extra parameter (α) can sometimes improve the fit relative to the pure gamma and (ii) on the other hand, in the one case where it does so (*rpsT*), the estimated value was not very high. Telling whether the phase transition ($\alpha > 1$) can be ignored or not is of key importance for the predictions on adaptation away from the optimum. This could be done by the type of simple fitting procedure illustrated here.

Discussion

In this article, I sought to build a mathematically tractable phenotype–fitness landscape from a set of qualitative first principles. It relies on five key assumptions about the effect of genetic changes on phenotype and fitness: (1) existence of a fitness optimum, (2) parent not too suboptimal, (3) mutations of mild effects, (4) pleiotropy of mutations on a large set of p phenotypic traits, and (5) high integration from these traits to fewer ($n \ll p$) optimized traits (see Figure 1). A statistical model is then applied to describe a mostly unspecified network of integration from many phenotypic dimensions into fitness. Various laws of large numbers (central limit theorem and random matrix theory) then yield a simplified phenotype–fitness landscape, in the limit of large np and $n \ll p$. This provides explicit results in terms of a few summary parameters. The resulting landscape is well approximated by Fisher's geometrical model, either in its isotropic form (all directions equivalent, Equations 8 and 9) or with simple anisotropy (one dominant direction plus isotropy in the remaining phenotype space, Equations 10 and 11). The most general result is an approximate

Table 2 Fit of the pdf in Equation 11 to empirical distributions of selection coefficients among random single mutations

Data set	$\hat{n}$	$\hat{\alpha}$	$\hat{\lambda}$	nb par	AIC	$P(\alpha > 1)$
Salmonella rplA gene	4.0 [2.7–5.5]	1.20 [1–4.9]	0.0049	3	−408.0	1.0 (NS)
(56 mutants)	3.9	1	0.0052	2	−410.0	
Salmonella rpsT gene	5.3 [3.8–7.0]	4.14 [2.2–7.1]	0.0033	3	−473.4	0.025*
(70 mutants)	3.8	1	0.0073	2	−470.4	

For each data set the maximum-likelihood estimates (MLE) are given for the fit of a gamma sum model (Equation 11, parameters $n, \tilde{\lambda}, \alpha$) and for the pure gamma fit (setting $\alpha = 1$). The 95% confidence interval for the value of n is provided (based on a 1.92 point reduction in log-likelihood relative the maximum likelihood). The Akaike information criterion (AIC) and number of fitted parameters (nb par) are given for each model. The *P*-value [$P(\alpha > 1)$] for the significance of the parameter α is given based on a likelihood-ratio test between the pure gamma and the gamma sum models. A Kolmogorov–Smirnov goodness-of-fit test (not shown) does not reject any of the fitted models (but this test is notably too conservative for fitted distributions).

stochastic representation for the DFE in the general model (Equation 10), yielding explicit pdfs in two subcases (Equations 9 and 11), as well as the moments of the distribution (File S2). The accuracy of these approximate distributions was confirmed by extensive simulations of various situations (Figure 5 and Figure S2).

Whenever the isotropic approximation is valid, predictions can be made on how the DFE will change with context (environmental or genomic background), based only on fitness measurements. This is the case where empirical tests appear most feasible. Otherwise (with anisotropy), the DFE will be affected by the direction from the parent toward the optimum (not just its distance), which is much more difficult to infer from empirical measures. Assumptions 1–4 alone yield a Fisher model, but isotropy further requires that $n \ll p$ (assumption 5) and that $\alpha \approx 1$ (mild bias in the distribution of pathway coefficients, see Equation 6). In fact, provided $\alpha \approx 1$, isotropy always seems to be a reasonable approximation, even in situations where n is only slightly smaller than p, and, hence, the spectral distribution of $\mathbf{M}$ is quite spread (see Figure 4). We propose a tentative justification for this robustness in File S2 and Figure S1.

In what follows, I discuss limitations and several implications of the model for the distribution of mutation fitness effects across genetic or environmental contexts and for parallel evolution.

Model assumptions and limitations

The five central assumptions were presented and discussed in the *Methods* section, so I do not delve into this here. The two most obvious limitations of the model lie in its local approximations. The first approximation assumes that mutations have mild effects on phenotype (assumption 3), so that the developmental function $\boldsymbol{\phi}(\mathbf{x}) = \mathbf{y}$ can be approximated by its linear trend [slope $\mathbf{B} = \boldsymbol{\phi}'(\mathbf{x})$]. The second local approximation (assumptions 2 and 3) assumes that the parent and its mutants all lie close enough to the optimum that their phenotypes lie below the leading-order quadratic approximation to the fitness function. These assumptions may, of course, break down (strong maladaptation, critical genetic changes that induce large phenotypic changes). Whether the model is robust to strong deviations from these assumptions is a matter of simulating various such deviations, which is beyond the scope of this article (it might be done using File S3). Whether such deviations are actually important in real-life systems is, as for all tests of the model, a matter of generating empirical DFEs, possibly from mutants in particular genes and measured in various contexts (genetic background or environmental conditions). Provided the maladaptation s_o corresponding to these contexts is measured, the model should provide testable predictions. An extension of the model could also consider higher-order approximations to the developmental functions, using tools derived, *e.g.*, by Rice (2004, Chap. 8).

Finally, it was also assumed that mutation effects on mutable traits are unbiased [$E(\mathbf{dx}) = 0$]. This potentially limits the generality of the conclusions. The present model yields a gamma distribution (or a sum of two gammas) when the parent is close to optimal. This type of distribution has shown a good fit to empirical DFEs, both in Lind's data set (Figure 6 and Lind *et al.* 2010) and in several previous studies (reviewed in Bataillon 2000; Martin and Lenormand 2006b; Eyre-Walker and Keightley 2007). It may thus be reasonable to ignore bias on the dx_j in the first approximation. A detailed study of the effects of biased mutation in Fisher's model can be found in Waxman and Peck (2003, 2004).

Relationship to previous theory

The effect of the phase transition on the DFE (Equation 11) can be related to previous studies on anisotropy and its impact on effective dimensionality n_e (Martin and Lenormand 2006b; Chevin *et al.* 2010). Below the phase transition, the effective dimensionality (as defined in Martin and Lenormand 2006b) would be close to $n_e = n$, which is consistent with the model being roughly isotropic. Beyond the phase transition and with a large $\alpha \gg 1$, the model would approximately reduce to that in Chevin *et al.* (2010) with $n_e \approx m = 1 < n$ and $s_{max} = s_1$ in their notation. The model is then driven by the mutant effects on the leading direction (eigenspace associated with $\lambda_1 \gg \tilde{\lambda}$), although the mutational variance is not exactly zero in other directions (because I assumed $p > n$), as was the case in this previous study. The present model can, thus, provide a natural model for parallel evolution, in the same way as that modeled in Chevin *et al.* (2010), as detailed below. Overall, our statistical approximation of a complex network yields forms of the Fisher model that have already been studied and for which several results have been derived.

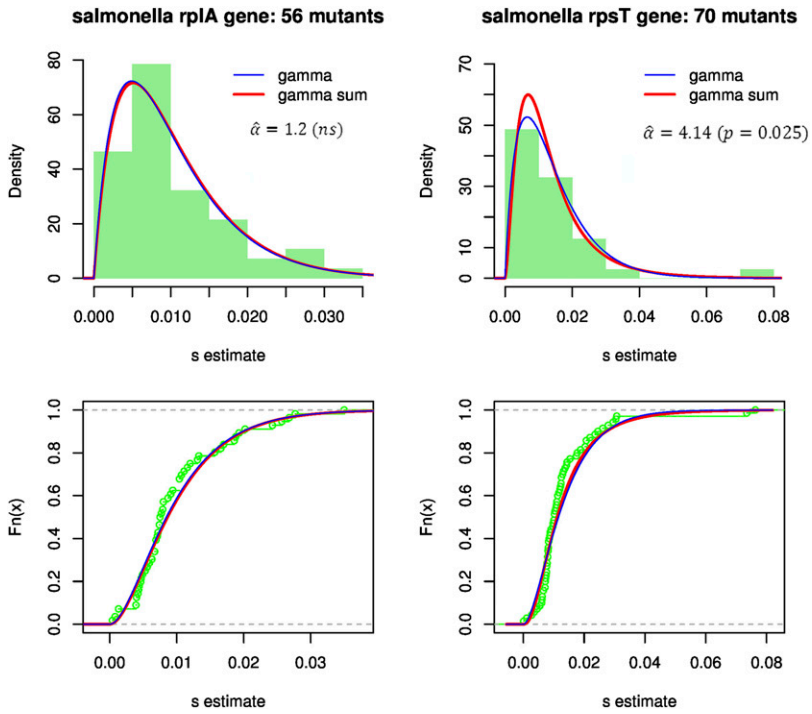

Figure 6 Fit of the gamma and gamma sum models to empirical DFEs in bacteria. Top, the distribution of selection coefficients among random nucleotide substitutions *vs.* alternative fitted models: Equation 11 with $\alpha = 1$ (a gamma distribution, blue curve) or α freely fitted (a convolution of two gammas, red curve). Bottom, the corresponding empirical cumulative distribution function *vs.* fitted models. The values of the best-fit parameters are given in Table 2. The significance of the test for $\alpha > 1$ *vs.* $\alpha = 1$ is given below the estimated value.

Another approximation of the DFE in Fisher's model, with a parent close to an optimum, has previously been proposed by Martin and Lenormand (2008), who used arguments from extreme-value theory. The present approach follows the same spirit in using a local approximation, but it differs from the extreme value approximation in three respects. First, the latter describes the distribution of only beneficial effects, not deleterious effects. Second, it applies in a potentially narrower range about the optimum, as it requires a low frequency of beneficial mutants, not just a local approximation in phenotype space. Finally, as for all previous work on the Fisher model, Martin and Lenormand (2008) had to assume the normality of mutation effects on optimized traits, whereas here it emerges from more general first principles. The present model can accurately capture fairly high proportions of beneficial mutants (Figure 5, Figure S2).

Finally, the present results might contribute to the debate over pleiotropy and its evolutionary cost (Orr 2000; Wagner and Zhang 2011; Hill and Zhang 2012). In the present model, two unrelated measures of dimensionality are defined at distinct levels of integration (see Figure 1). The dimension of pleiotropy p is the number of not fully dependent traits that are jointly affected by mutations in a given genetic target. It is the quantity often considered in empirical studies of pleiotropy. The second dimension is that of optimization, n, the number of not fully dependent traits that jointly define a local fitness optimum. The two are related by the developmental function $\boldsymbol{\phi}(.)$. Whenever $n < p$, it is clearly n and not p that affects the DFE and thus has an evolutionary impact: when pleiotropy (p) is higher than the dimension of optimization (n), it is the latter that drives the "cost of complexity." The recent study by Le Nagard *et al.* (2011) also appears consistent with our findings. They simulated evolution in a phenotype–fitness landscape where the underlying network of interacting genes could evolve. They showed that complexity, in "Fisherian" terms (roughly our n), could evolve in response to the complexity of the environmental challenge imposed (defined by the complexity of an optimal reaction norm). This evolution was rather independent of the size of the underlying network (which could be a good proxy for p), especially for larger networks (where $n < p$). In this study, too, the cost of complexity seems to be "paid" according to the dimensionality of optimization rather than that of pleiotropy itself.

Relationship to previous empirical findings

A central goal of this study was to propose an explanation for why the Fisher model, with all its simplifications, can still capture some real-life patterns with reasonable accuracy. Simple skewed distributions like the negative gamma have repeatedly shown good agreement with empirical DFEs (reviewed in Bataillon 2000; Martin and Lenormand 2006b; Eyre-Walker and Keightley 2007). This may not be so surprising: with high pleiotropy ($p \gg 1$) and mild mutation effects, we always expect to see such gamma-like DFEs, in permissive conditions ($s_o \approx 0$, see Figure 5). However, an extra parameter arises beyond the phase transition ($\alpha > 1$, Equation 11), and adding this parameter improved the fit of the DFE in the *rpsT* gene in Figure 6. This may also be the case for other data sets where the gamma alone did not capture the fat tail of observed distributions (*e.g.*, Elena *et al.* 1998), while the tail gets fatter here in Equation 11 as α increases. This result goes beyond optimal conditions: as long as the parent and mutants remain close enough to some local optimum, we expect the

isotropic FGM to accurately predict the DFE, with both beneficial and deleterious mutations (Figure 5). This may thus explain why the FGM also provided accurate predictions in several empirical studies of fitness epistasis and/or compensation, in the presence of beneficial mutations (Martin *et al.* 2007; MacLean *et al.* 2010; Sousa *et al.* 2011).

Impact of the genetic or environmental context?

By building a more mechanistic version of the Fisher model, we can evaluate which parameters of the DFE may or may not depend on the context (environment, parental lineage, mutational target). Of course, in the end, this issue can be settled only by empirical testing; in this respect, gene-specific measurements of the DFE (such as in Lind *et al.* 2010) offer the most promising approach. In its most general form (Equation 10), the model has five parameters $(n, \tilde{\lambda}, s_{n-1}, s_1, \alpha)$. I discuss each of them in turn, assuming a given species, so that variation in the "lineage" context is limited to genetic variation within a species.

Parameter n is the number of traits under optimizing selection (the dimension of optimization, in $\mathbf{y}$ space). It might be expected to be fairly stable across contexts within a given species, as it relates to the core internal processes determining fitness. The nature of the traits under selection can also change without affecting the parameter, as long as they remain in roughly the same number. In the empirical example on *Salmonella* (Figure 6, Table 2), no significant difference in $\hat{n}$ was detected between mutations within the two genes studied. However, the power to detect such differences was limited, and whether these genes truly pertain to distinct "targets" is difficult to assess: they are part of the same broad class of ribosomal protein genes.

Parameter $\tilde{\lambda}$ summarizes the contribution of mutational covariances on mutable traits ($\mathbf{V}$) and of covariances among pathway coefficients ($\mathbf{C_B}$). It may, in principle, be affected by any effect of the context on these covariances. A change in the nature of optimized traits (see above) would also affect $\tilde{\lambda}$, because a different developmental function (from $\mathbf{x}$ to $\mathbf{y}$) would then be defined, modifying $\mathbf{C_B}$. Yet, because it synthesizes the small contribution from many parameters, it may also prove stable across contexts. Even if $\mathbf{V}$ and/or $\mathbf{C_B}$ changes across contexts, the resulting effect on the DFE will be negligible if they still average out to the same $\tilde{\lambda}$. Settling this issue could be possible by empirical fitting of the sort described in Figure 6. In Table 2 the estimates of $\tilde{\lambda}$ were fairly similar in the two genes considered. The context here would be the mutational target, which may or may not differ between these genes. More such studies would obviously provide key insights.

Parameter α was the only parameter showing significantly different estimates in the two genes whose DFE was fitted in Figure 6 (see Table 2). This suggests that these two genes did pertain to distinct mutational targets and that α may vary among these targets (in the same environment and genetic background). A simple (maybe too simple) way to account for such a pattern is to consider that each gene "samples" a subset of pathways, with associated coefficients b_{ij}, within the larger set of all possible pathways. According to the bias in the b_{ij} sampled by a given gene, $cv^2 > \sqrt{n\ p_e}$ or not, and this gene will lie below or beyond the phase transition ($\alpha = 1$ or $\alpha > 1$, see Equations 5 and 6). This provides a null model to describe genomic variation in the DFE.

Finally, the fitness distance to the optimum [s_o or the pair (s_1, s_{n-1})] is typically expected to vary across contexts. It is the main application of Fisher's model to summarize the effect of various contexts into variation at a single parameter s_o. The genomic background and the environment should jointly determine s_o or (s_1, s_{n-1}) *a priori*. Whether the mutational target (gene) affects these parameters depends on the isotropy of the model. Below the phase transition (Equation 8), all genes "see" the same distance to the optimum s_o. Beyond the phase transition, however (Equation 10), the overall distance $s_o = s_1 + s_{n-1}$ may be divided into distinct pairs of (s_1, s_{n-1}) according to the subspace associated with the leading eigenvalue λ_1 for this gene. This is the basis for potential parallel evolution, by roughly the same process as described in Chevin *et al.* (2010).

To summarize, the DFE in Equations 8–10 is unaffected by (i) the nature of the distribution of the pathway coefficients or (ii) the nature of the mutation effects on mutable traits, by (iii) their number p (as long as $p \gg n$), or by (iv) their particular values or mutual covariance, as long as they retain the same overall trace $\mathrm{Tr}(\mathbf{V}.\mathbf{C_B})$. Overall, the argument above merely shows that several effects of the context should have negligible influence on the DFE, but it does not state that none has an impact. At least one advantage is that all these complex effects can be summarized into the effects of context on five parameters [$n, \lambda, \bar{s}, \alpha,$ and s_o or (s_1, s_{n-1})]. All these parameters are measurable empirically, except *a priori* for the pair (s_1, s_{n-1}), which arises when $\alpha > 1$ and the parent is suboptimal.

Essential genes

Let us now focus on a particular type of context dependence that has the advantage of having been well studied empirically. In recent years, studies of large sets of gene deletions and their impact on fitness have flourished, mostly in the yeast *S. cerevisae*. One salient feature that emerged is that genes fall into two distinct categories: "essential" genes (whose deletion is lethal or nearly lethal) and "nonessential" or "dispensable" genes (whose deletion has weak effects). One puzzling finding is that there are essential genes in comparable proportions in most functional gene complexes (Fudala and Korona 2009). Under the model proposed here, there are two simple and testable alternative interpretations for an essential gene. First, if a given gene samples the full set of pathway coefficients, it may by chance sample a set that leads to $\tilde{\lambda}$ larger than usual. This gene would then have higher average effect on fitness, and its deletion might thus be particularly harmful. Alternatively, the gene may sample a set of coefficients with a highly biased distribution ($\boldsymbol{\mu}_\mathbf{B}$ large), so that it falls beyond the phase transition [$\alpha = 1 + n/\mathrm{cv}^2 \gg 1$, Equations 5 and 6]. It

would then also be particularly harmful when deleted, as $\bar{s} = (n - 1 + \alpha)\tilde{\lambda}/2$. This simplistic view has the advantage of being testable, by extensions of the experiments in Lind *et al.* (2010) or using the more recent high-throughput method EMPIRIC (Hietpas *et al.* 2011). The test would consist of measuring the gene-specific DFE in a well-adapted parent ($s_o \approx 0$), in essential *vs.* nonessential genes. In the first case, the DFE at an essential gene should show a higher $\tilde{\lambda}$ than at nonessential genes (a gamma with large scale). In the second case, it should show a larger value of α (a heavier left tail than the gamma). In the future, this type of method might give valuable insights into gene essentiality. There are also more puzzles with lethal mutations in general than "simply" gene essentiality (discussed in Manna *et al.* 2011).

Parallel evolution and the genomic DFE

A natural follow-up of this argument is that some environments may lead to parallel evolution in those genes that lie beyond the phase transition (one of the mechanisms generating essential genes, see above). Because the model retrieved is close to the form described in Chevin *et al.* (2010), their framework applies to describe the DFE over the genome and how parallel evolution can arise. We propose that $\alpha \gg 1$ arises in those genes that sample a biased distribution of pathway coefficients. Those are also the ones likely to generate parallel evolution, in the particular environments where the direction to the optimum corresponds to their dominant direction. In these particular environments, selection will then fix mutations preferentially in those genes that (i) are beyond phase transitions and (ii) have the corresponding leading direction. Note that those genes with $\alpha \gg 1$ might not have noticeable influence on the genomic DFE, which is roughly the mean of the gene-specific DFEs (Chevin *et al.* 2010), unless there are many such genes across the genome. Yet these genes will be overrepresented among those undergoing parallel evolution.

Extensions of the model and further applications of RMT

This model illustrates the potential of RMT as a tool for multivariate evolution studies. This is not new, as several studies (Wagner 1984, 1989; Martin and Lenormand 2006b; Chevin *et al.* 2010) have already used random covariance matrices to model phenotypic covariances, but this choice was assumed from the start. Here, the random covariance matrix structure emerges from first principles. The approach could be extended, for example, using results on large "random sparse matrices" (Bai and Silverstein 2010, Chap. 7) to model less connected networks with isolated clusters (a large proportion of zeroes in matrix **B**). This would correspond to a more "modular" network where each mutable trait affects only a small portion of all optimized traits (contrary to our assumption 5).

Another interesting extension could be to follow the effect of correlations of pathway coefficients among optimized traits (b_{ij} correlated both across j *and* across i). As explained in the *Distribution of pathway coefficients* section, the model assumptions forbid, *de facto*, a substantial correlation of b_{ij} across optimized traits (across i) because $\mathbf{C_B}$ would become positive semidefinite. In File S1 (Equations A1.16–A1.20), I consider an extension, suggested by Dave McCandlish, to allow for weakly correlated b_{ij} among the rows i, not only the columns j. An approximate treatment can be found, provided that this correlation remains weak. This extension does not change the structure of the model: we retain an M-P law prediction, but with a further reduction of the ratio β_e. This ratio now takes a finite limit, even as $n/p \to 0$. Let cv_b be the coefficient of variation of the eigenvalues of the covariance matrix of the pathway coefficients (b_{ij}) among optimized traits (among rows i). The effective dimensionality n_e (Martin and Lenormand 2006b), which was roughly $n_e = n$ in the model without correlations among rows i, now becomes $<n$, more precisely $n_e \approx n/(1 + cv_b^2)$. This additional correlation can thus create anisotropy, but, *a priori*, of smaller order than what is obtained via the phase transition. My (clearly disputable) claim, here, is that these "developmental" correlations across optimized traits should be fairly weak, as the pathway coefficients b_{ij} result from "percolation," through many intermediate levels of organization, from the mutable traits j to the optimized traits i.

Without going into detailed extensions, these examples show that even more general models could be handled with tools from RMT.

Conclusion

This model of pleiotropic mutation over a phenotypic network (Figure 1) is less phenomenological than the original Fisher's model. Yet, it remains uninformative on the functional reason why any mutation has a given fitness effect. This is both a weakness and a strength: because the results prove independent of many details of the phenotypic network, they do not provide any information on it either. Focusing on evolutionary predictions, not on phenotypic networks, the fact that the isotropic Fisher model can emerge is good news. Indeed, it means that we may predict the DFE in various contexts, without much knowledge of its underlying functional determinants, a knowledge that is not available for most species of interest, in most environments.

Acknowledgments

I thank the editor, Michael Kopp and Dave McCandlish, whose involvement and constructive suggestions greatly improved the rigor and clarity of this article. I thank Dan Anderson for kindly providing data and insights on his random mutant experiment in *S. typhimurium* and Florent Benaych-Georges for a helpful discussion on his method to compute extreme eigenvalues. David Waxman, Rees Kassen, Thomas Lenormand, and Luis Miguel Chevin greatly helped this work by stimulating discussions; the latter also kindly provided valuable suggestions on an earlier version of the manuscript. This work was funded by Centre National de la

Recherche Scientifique (CNRS), in particular by Projets Exploratoires Pluridisciplinaires Inter-Instituts (PEPII 2011-2012), and by the Agence Nationale de la Recherche (EVORANGE ANR-09-PEXT-011).

Literature Cited

Bai, Z., and J. W. Silverstein, 2010 *Spectral Analysis of Large Dimensional Random Matrices*, Ed. 2. now Publishers Inc., Hanover, MA.

Barabasi, A. L., and Z. N. Oltvai, 2004 Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 5: 101–113.

Barton, N. H., and J. B. Coe, 2009 On the application of statistical physics to evolutionary biology. J. Theor. Biol. 259: 317–324.

Bataillon, T., 2000 Estimation of spontaneous genome-wide mutation rate parameters: Whither beneficial mutations? Heredity 84: 497–501.

Baxter, G. J., R. A. Blythe, and A. J. McKane, 2007 Exact solution of the multi-allelic diffusion model. Math. Biosci. 209: 124–170.

Benaych-Georges, F., and R. R. Nadakuditi, 2011 The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. Adv. Math. 227: 494–521.

Bhardwaj, N., K. K. Yan, and M. B. Gerstein, 2010 Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels. Proc. Natl. Acad. Sci. USA 107: 6841–6846.

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

Chou, H. H., H. C. Chiu, N. F. Delaney, D. Segre, and C. J. Marx, 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science 332: 1190–1192.

Cooper, V. S., and R. E. Lenski, 2000 The population genetics of ecological specialization in evolving Escherichia coli populations. Nature 407: 736–739.

Derewacz, D. K., C. R. Goodwin, C. R. McNees, J. A. McLean, and B. O. Bachmann, 2013 Antimicrobial drug resistance affects broad changes in metabolomic phenotype in addition to secondary metabolism. Proc. Natl. Acad. Sci. USA 110: 2336–2341.

Di Salvo, F., 2008 A characterization of the distribution of a weighted sum of gamma variables through multiple hypergeometric functions. Integr. Trans. Spec. Funct. 19: 563–575.

Elena, S. F., L. Ekunwe, N. Hajela, S. A. Oden, and R. E. Lenski, 1998 Distribution of fitness effects caused by random insertion mutations in Escherichia coli. Genetica 103: 349–358.

Elena, S. F., and R. E. Lenski, 2003 Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat. Rev. Genet. 4: 457–469.

Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8: 610–618.

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

Forrester, P. J., N. C. Snaith, and J. J. M. Verbaarschot, 2003 Developments in random matrix theory. J. Phys. A Math. Gen. 36: R1–R10.

Fudala, A., and R. Korona, 2009 Low frequency of mutations with strongly deleterious but nonlethal fitness effects. Evolution 63: 2164–2171.

Gros, P. A., and O. Tenaillon, 2009 Selection for chaperone-like mediated genetic robustness at low mutation rate: impact of drift, epistasis and complexity. Genetics 182: 555–564.

Hietpas, R. T., J. D. Jensen, and D. N. A. Bolon, 2011 Experimental illumination of a fitness landscape. Proc. Natl. Acad. Sci. USA 108: 7896–7901.

Hill, W. G., and X. S. Zhang, 2012 Assessing pleiotropy and its evolutionary consequences: pleiotropy is not necessarily limited, nor need it hinder the evolution of complexity. Nat. Rev. Genet. 13: 296.

Hindré, T., C. Knibbe, G. Beslon, and D. Schneider, 2012 New insights into bacterial adaptation through in vivo and in silico experimental evolution. Nat. Rev. Microbiol. 10: 352–365.

Ibarra, R. U., J. S. Edwards, and B. O. Palsson, 2002 Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420: 186–189.

Jaschke, S., C. Kluppelberg, and A. Lindner, 2004 Asymptotic behavior of tails and quantiles of quadratic forms of Gaussian vectors. J. Multivariate Anal. 88: 252–273.

Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper, 2011 Negative epistasis between beneficial mutations in an evolving bacterial population. Science 332: 1193–1196.

Kimura, M., 1965 A stochastic model concerning maintenance of genetic variability in quantitative characters. Proc. Natl. Acad. Sci. USA 54: 731–736.

Lande, R., 1979 Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33: 402–416.

Le Nagard, H., L. Chao, and O. Tenaillon, 2011 The emergence of complexity and restricted pleiotropy in adapting networks. BMC Evol. Biol. 11: 326.

Lind, P. A., O. G. Berg, and D. I. Andersson, 2010 Mutational robustness of ribosomal protein genes. Science 330: 825–827.

MacLean, R. C., G. G. Perron, and A. Gardner, 2010 Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. Genetics 186: 1345–1354.

Manna, F., G. Martin, and T. Lenormand, 2011 Fitness landscapes: an alternative theory for the dominance of mutation. Genetics 189: 923–937.

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

Martin, G., and L. Lenormand, 2008 The distribution of beneficial and fixed mutation effects close to an optimum. Genetics 179: 907–916.

Martin, G., and T. Lenormand, 2006a The fitness effect of mutations in stressful environments: a survey in the light of fitness landscape models. Evolution 60: 2413–2427.

Martin, G., and T. Lenormand, 2006b A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. Evolution 60: 893–907.

Mathai, A. M., and S. B. Provost, 1992 *Quadratic Forms in Random Variables*. Marcel Dekker, New York.

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

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

Papp, B., R. A. Notebaart, and C. Pal, 2011 Systems-biology approaches for predicting genomic evolution. Nat. Rev. Genet. 12: 591–602.

Papp, B., C. Pal, and L. D. Hurst, 2004 Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429: 661–664.

Perfeito, L., S. Ghozzi, J. Berg, K. Schnetz, and M. Lässig, 2011 Nonlinear fitness landscape of a molecular pathway. PLoS Genet. 7: e1002160.

Poon, A., and S. P. Otto, 2000 Compensating for our load of mutations: freezing the meltdown of small populations. Evolution 54: 1467–1479.

Price, N. D., J. L. Reed, and B. O. Palsson, 2004 Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat. Rev. Microbiol. 2: 886–897.

R Development Core Team, 2013 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rice, S. H., 2004 *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sinauer Associates, Sunderland, MA.

Rice, W. R., 2002 Experimental tests of the adaptive significance of sexual recombination. Nat. Rev. Genet. 3: 241–251.

Salazar-Ciudad, I., and J. Jernvall, 2010 A computational model of teeth and the developmental origins of morphological variation. Nature 464: 583–586.

Salazar-Ciudad, I., and M. Marin-Riera, 2013 Adaptive dynamics under development-based genotype-phenotype maps. Nature 497: 361–364.

Schoustra, S. E., T. Bataillon, D. R. Gifford, and R. Kassen, 2009 The properties of adaptive walks in evolving populations of fungus. PLoS Biol. 7: e1000250.

Segré, D., A. DeLuna, G. M. Church, and R. Kishony, 2005 Modular epistasis in yeast metabolism. Nat. Genet. 37: 77–83.

Sousa, A., S. Magalhaes, and I. Gordo, 2011 Cost of antibiotic resistance and the geometry of adaptation. Mol. Biol. Evol. 29: 1417–1428.

Tenaillon, O., O. K. Silander, J. Uzan, and L. Chao, 2007 Quantifying organismal complexity using a population genetic approach. PloS ONE 2: e217.

Tulino, A. M., and S. Verdù, 2004 *Random Matrix Theory and Wireless Communications*. now Publishers Inc., Delft, The Netherlands.

Wagner, A., 2002 Estimating coarse gene network structure from large-scale gene perturbation data. Genome Res. 12: 309–315.

Wagner, G. P., 1984 On the eigenvalue distribution of genetic and phenotypic dispersion matrices - evidence for a nonrandom organization of quantitative character variation. J. Math. Biol. 21: 77–95.

Wagner, G. P., 1989 Multivariate mutation-selection balance with constrained pleiotropic effects. Genetics 122: 223–234.

Wagner, G. P., and J. Z. Zhang, 2011 The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat. Rev. Genet. 12: 204–213.

Wagner, G. P., and J. Z. Zhang, 2012 Universal pleiotropy is not a valid null hypothesis: reply to Hill and Zhang. Nat. Rev. Genet. 13: 296.

Waxman, D., and J. R. Peck, 2003 The anomalous effects of biased mutation. Genetics 164: 1615–1626.

Waxman, D., and J. R. Peck, 2004 A one locus, biased mutation model and its equivalence to an unbiased model. Biosystems 78: 93–98.

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 insight into the shape and size of phenotype space. Evolution 67: 2957–2972.

Wolfram Research, 2012 *Mathematica Edition: Version 9.0*. Wolfram Research, Champaign, IL.

Communicating editor: J. Hermisson

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.160325/-/DC1

Fisher's Geometrical Model Emerges as a Property of Complex Integrated Phenotypic Networks

Guillaume Martin

DOI: 10.1534/genetics.113.160325

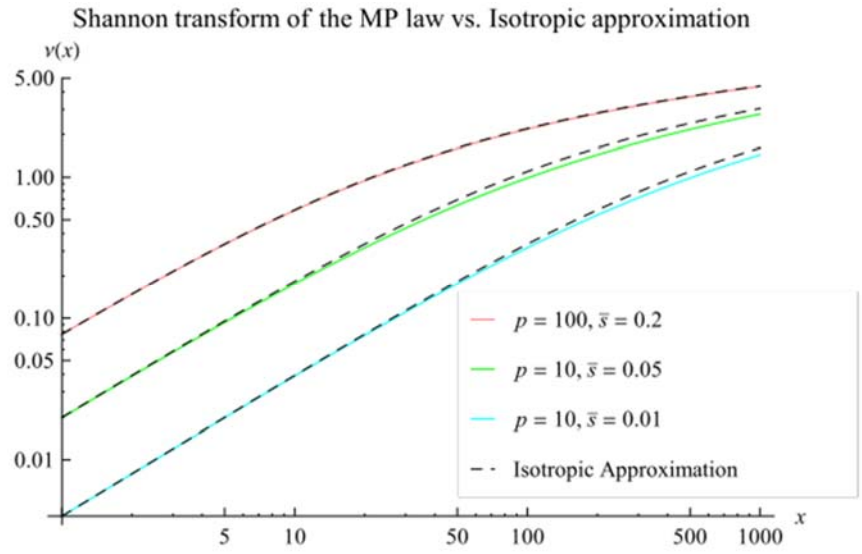

Figure S1 Convergence of the Shannon transform to the isotropic approximation

The Shannon transform of the M-P law ($v_{H.H^*}(x)$ with $\lambda_{\mathrm{H.H^*}} \sim MP(\beta,\zeta)$, see eq. A1.3) is compared to the isotropic approximation ($v(x) \approx \log(1+\tilde{\lambda}\,x)$) where $\tilde{\lambda} = \beta\zeta = 2\bar{s}/n$ where $\bar{s} = -E(s)$ is the mean (deleterious) effect of mutations. The parameters are indicated on the graph, with $n = 5$ and recalling that $\beta = p/n$ and $\zeta = 2\bar{s}/p$. The isotropic approximation proves accurate over a wide range of parameter values.

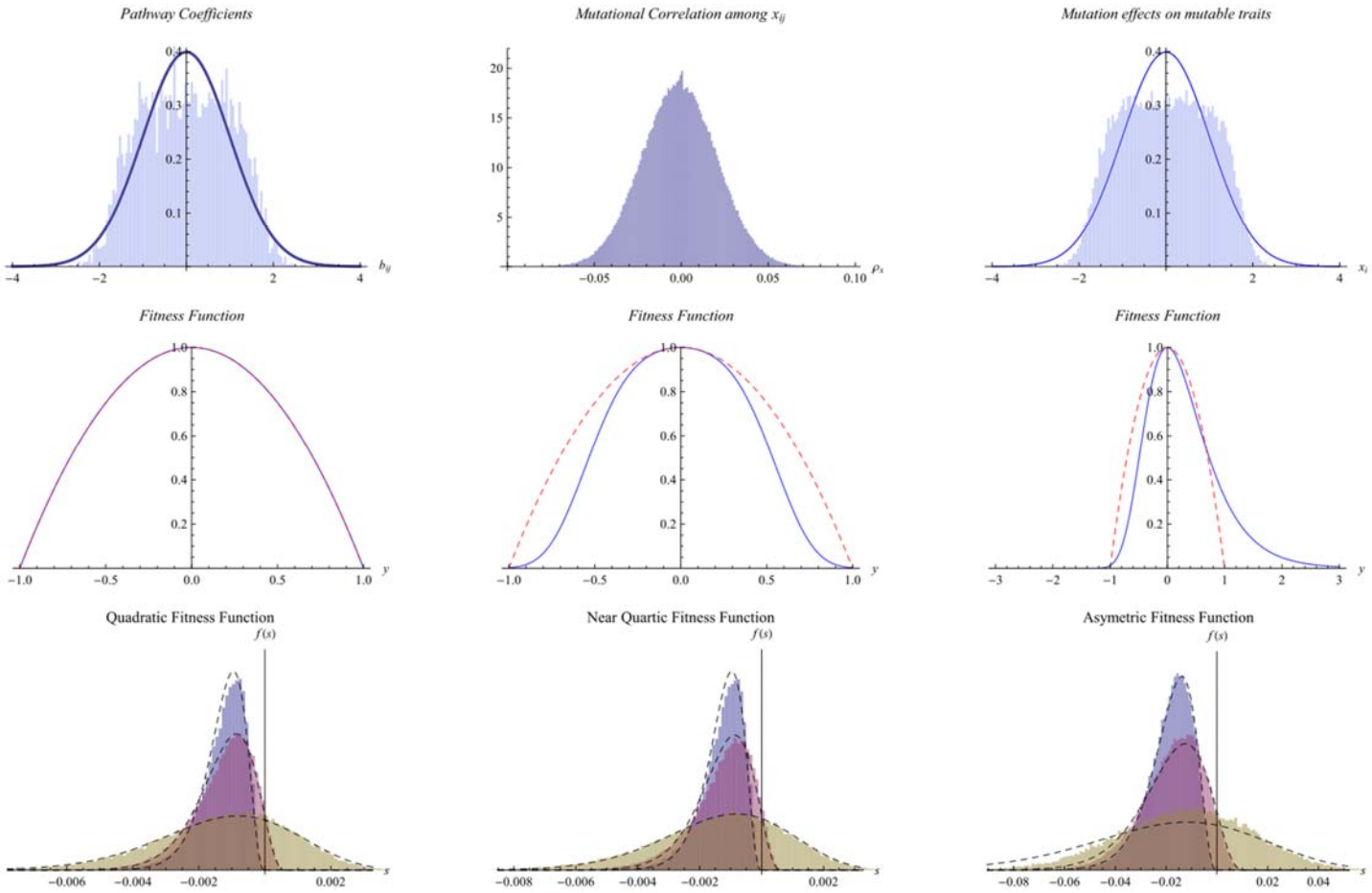

Figure S2 Distribution of mutation fitness effects with various fitness functions

The DFE in simulations (below phase transition, as in **Figure 5.c**) is compared to the analytic prediction (eq. 9) for three choices of fitness function for which the quadratic approximation applies (detailed in **Figure 2.a.**): quadratic, near quartic and asymmetric. The upper panels show various distributions common to all simulations: pathway coefficients, correlations between mutation effects on mutable traits $\mathbf{x}$ and mutation effects on $\mathbf{x}$ themselves. The middle panels show the three alternative fitness functions. The bottom panels show the corresponding DFE in simulations (histograms) and the corresponding prediction (dashed lines, eq. (9)) for various distances to the optimum (same as **Figure 5**: $s_o = 0$, blue, $s_o = \bar{s}$, red, $s_o = 5\bar{s}$, brown).

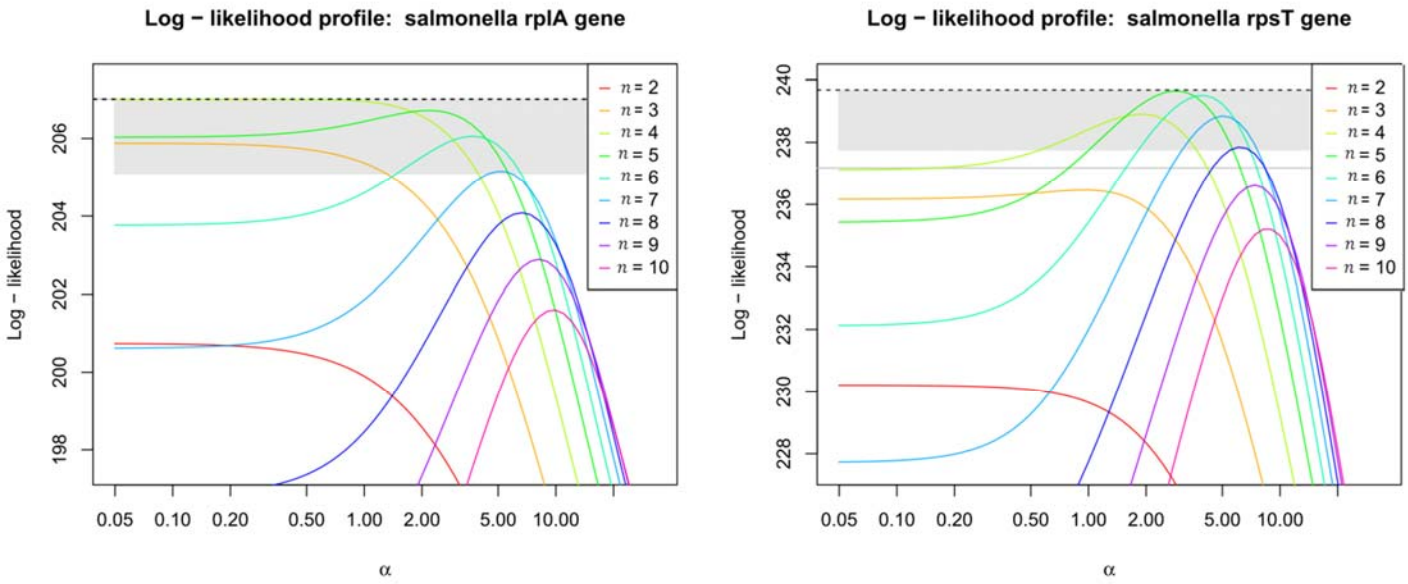

Figure S3 Log – likelihood profile for the parameter α from empirical DFEs

For each gene, the log-likelihood of the fitted model (**Figure 6** and **Table 2**) is given as a function of the parameter α, for various integer values of n (see legend) and with the scale λ set to match the mean of the distribution ($\lambda = 2\bar{s}/(n-1+\alpha)$). The dashed black line gives the log – likelihood of the best fitting gamma sum model (eq. (11) with n, α, λ jointly fitted) and the gray plain line gives the log – likelihood of the best fitting pure gamma model (i.e. setting $\alpha = 0$). The shaded area corresponds to log – likelihood lying within 1.92 points of the maximum log – likelihood for the best fitting model (with α fitted), implying no significant difference in the fit to the data. In the *rplA* gene, the pure gamma models with $n = 3,4,5$ or 6 are not statistically different from the best fitting model (**Table 3**): their maxima lie within the shaded area. For the *rpsT* gene, the 'gamma sum' model (eq. (11), with $\alpha > 0$) provides a significantly better fit than a pure gamma model with $\alpha = 0$ (the gray line lies below the gray shaded area). It further shows that the best fitting model (**Table 3**) and models with $n = 4,5,6$ or 7, with corresponding (increasing) values of $\alpha > 0$, cannot be distinguished in terms of goodness of fit (all their maxima lie within the shaded area).

File S1

Approximations to the LSD for some classes of random matrices

In this appendix, I first state several results on the Limit Spectral Distribution (LSD) of some large random covariance matrices. This LSD is the nonrandom limit of the distribution of the eigenvalues of a random matrix, as its dimensions get large (here as $n, p \to \infty$). The first section merely states various known results from Random matrix Theory: most are summarized in (TULINO and VERDÙ 2004). Then I derive an approximation for the LSD of $\mathbf{M}$ in the integrative phenotypic network model presented in the main text (**Figure 1**), and for key quantities related to the mutant fitness distribution. In what follows, I refer to the pdf of the spectral distribution of a given matrix $\mathbf{X}$ as $\rho_{\mathbf{X}}$ and to the corresponding Limit Spectral distribution (LSD) of this matrix as $\tilde{\rho}_{\mathbf{X}}$. I refer to any of the transforms of this LSD by an index referring to that matrix. All the derivations of this Appendix can be checked in a Mathematica® (WOLFRAM RESEARCH 2012) notebook (**file S3**, in freely readable [.cdf] format) available for download at **XXX**.

A) The Marchenko-Pastur law and its properties

Limiting Spectral Distribution (LSD): Central to our derivations is the Marchenko - Pastur (M-P) law, which provides the LSD of a covariance matrix with random entries. Consider a $n \times p$ matrix $\mathbf{H}$ whose entries h_{ij} are randomly and independently drawn from possibly different distributions. These distributions are arbitrary in nature but have mean zero, variance $1/n$ and fourth moments of order $O(1/n^2)$, i.e. they are 'standard' zero mean distributions, not too leptokurtic (normal, extreme value, uniform, etc.). Then, if $\zeta > 0$ is an arbitrary scale factor, the random $n \times n$ matrix $\zeta\, \mathbf{H}.\mathbf{H}^*$ is called a sample covariance matrix. Its name comes from the fact that $\zeta\, \mathbf{H}.\mathbf{H}^*$ can describe the covariance matrix in a sample (of size p) from a multivariate distribution with covariance matrix $\zeta \mathbf{I}_n$. When the entries of $\mathbf{H}$ are normally distributed, $n\, \mathbf{H}.\mathbf{H}^*$ follows the standard Wishart distribution: $n\, \mathbf{H}.\mathbf{H}^* \sim W_p(\mathbf{I}_n)$ with p degrees of freedom. As $(p, n) \to \infty$ with a fixed ratio ($p/n \to \beta$), the spectral distribution of $\mathbf{H}.\mathbf{H}^*$ converges to a nonrandom LSD whose pdf is given by eq. (1.12) in (TULINO and VERDÙ 2004)

$$\rho_{\mathbf{H}.\mathbf{H}^*} \Rightarrow \tilde{\rho}_{\mathbf{H}.\mathbf{H}^*}(x) = (1-\beta)^+ \delta(x) + \frac{\sqrt{(b-x)(x-a)}}{2\,\pi\,\zeta\,x}, \quad x \in [a, b] \qquad , \qquad \text{(A1.1)}$$
$$a = \zeta\left(1-\sqrt{\beta}\right)^2 \text{ and } b = \zeta\left(1+\sqrt{\beta}\right)^2$$

where $(x)^+ = \max(x, 0)$ and $\delta(.)$ is the Dirac delta function.

This pdf has a point mass at zero with weight $1-\beta$ whenever $0 < \beta < 1$, plus a bulk of positive eigenvalues with weight $\min(\beta, 1)$. When $\beta \geq 1$ (our case of interest), the left hand term vanishes and all eigenvalues are strictly positive: we then retrieve eq. (3) of the main text. The pdf of the bulk is $\rho_{+(x)} = \rho(\lambda|\lambda > 0) = \sqrt{(b-x)(x-a)}/(2\pi\,\zeta\,\lambda\,\min(\beta, 1))$. As stated in main text, I refer to this

distribution as the scaled M-P law with scale parameter ζ and ratio index β and denote the result by the stochastic representation $\lambda_{\zeta\,\mathbf{H}.\mathbf{H}^*} \sim MP(\beta,\zeta)$. The mean eigenvalue is $\mathrm{E}(\lambda) = \zeta\,\beta$, with variance $\mathrm{V}(\lambda) = \zeta^2\beta$. Note that the original (and more standard) statement of the M-P law (TULINO and VERDÙ 2004; BAI and SILVERSTEIN 2010) is in terms of the LSD of $\mathbf{H}^*.\mathbf{H}$, which has ratio index $1/\beta$ and corresponding scale $\zeta\beta$: $\lambda_{\mathbf{H}^*.\mathbf{H}} \sim MP(1/\beta,\zeta\beta)$ in our notation. Note also that, for notational simplicity, I drop the reference to the scaling factor ζ in the indexing when referring to the LSD of sample covariance matrices ($\tilde{\rho}_{\mathbf{H}.\mathbf{H}^*}(x), v_{\mathbf{H}.\mathbf{H}^*}(x)$ etc.).

One powerful property of the M-P law (and of random matrix theory in general) is that the actual spectral distributions of finite random covariance matrices converge quickly to their limit: for any single draw of a matrix $\mathbf{H}$ with dimensions, say, $n = 10$ and $p = 100$ the spectral distribution of $\mathbf{H}.\mathbf{H}^*$ is already well described by the M-P law given above, in that it is bounded within the predicted domain $[a,b]$ in eq. (A1.1), and that the pdf $\rho(x)$ is close to the M-P law. However, it is difficult to represent this pdf with small $\min(n,p)$ as there are then only few eigenvalues to show. This is why I show examples with larger dimensions (e.g. $n = 100, p = 500$). However, when simulating the DFE I will use smaller parameter values and show the convergence of the DFE to the predicted distribution based on Random Matrix Theory.

Transforms of the LSD: Various transforms of the LSD of random matrices have been defined in Random Matrix Theory, I propose here a quick overview, drawn from section 2.2 of (TULINO and VERDÙ 2004). The purpose of these transforms is akin to that of generating functions in standard probability theory. These transforms can be derived from one another and allow to derive various properties of the LSD, or to compute the LSD of the sum or product of random matrices, which will prove useful in our case. Each transform fully characterizes a given LSD, just as the pdf $\tilde{\rho}(x)$ does. The mutual relationships between these transforms are illustrated in the notebook **file S3**.

Let $\tilde{\rho}(x)$ be the pdf of the LSD of a random matrix, defined on some finite or infinite range $[\lambda_{min},\lambda_{max}]$. The first important transform is the Stieltjes transform: $S(z) = \int_{\lambda_{min}}^{\lambda_{max}} 1/(x-z)\rho(x)dx$ with $z \in \mathbb{C}$ the set of complex numbers. It provides the range and pdf of the eigenvalues of the matrix, in particular $\tilde{\rho}(x) = \lim_{\omega\to 0+} S(x + \mathbf{i}\,\omega)$ where $\mathbf{i}$ is the complex unit number. The η transform: $\eta(z) = \int_{\lambda_{min}}^{\lambda_{max}} 1/(1+z\,x)\tilde{\rho}(x)dx$, with $z \in \mathbb{R}^+$(positive real numbers) is defined for positive semi-definite matrices only (with eigenvalues all positive or zero, $[\lambda_{min},\lambda_{max}] \subset \mathbb{R}^+$). This function is a generating function for the raw moments of the LSD: we have $\eta(0) = 1$ and $\eta[-z] = \sum_{k=0}^{\infty} z^k \mathrm{E}(\lambda^k)$. The η transform is related to the Stieltjes transform via $\eta(z) = S(-1/z)/z$. For the scaled M-P law in eq. (A1.1), the η transform is given by

$$\begin{aligned} \eta_{\mathbf{H}.\mathbf{H}^*}(z) &= 1 - \phi(z)/(4\,z\,\zeta) \\ \phi(z) &= \left(\sqrt{b\,z+1} - \sqrt{a\,z+1}\right)^2 \end{aligned} \quad , \qquad \text{(A1.2)}$$

where (a, b) are given in eq. (A1.1). Taking the limit at infinity provides the weight of the point mass at zero (the proportion of zero eigenvalues) $\lim_{z\to\infty} \eta_{\mathbf{H}.\mathbf{H}^*}(z) = \max(1-\beta, 0)$.

Central to our applications is the Shannon transform $v(z) = \int_{\lambda_{min}}^{\lambda_{max}} \log(1 + x\, z)\, \tilde{\rho}(x) dx$, also defined only for positive semi-definite matrices ($[\lambda_{min}, \lambda_{max}] \subset \mathbb{R}^+$), with $z \in \mathbb{R}^+$. It is related to the η transform via $v(z) = \int \big(1 - \eta(z)\big)/z\, dz$. For the scaled M-P law the Shannon transform is

$$v_{\mathbf{H}.\mathbf{H}^*}(z) = \beta \log\left(1 + \zeta\, z - \frac{\phi(z)}{4}\right) + \log\left(1 + \zeta\, z\, \beta - \frac{\phi(z)}{4}\right) - \frac{\phi(z)}{4\,\zeta\, z} \quad , \qquad \text{(A1.3)}$$

Where $\phi(.)$ is given in eq.(A1.2).

The two next transforms are key to compute the LSD of sums or products of random matrices whose entries are drawn independently, more specifically matrices that are asymptotically free, meaning that their LSD are independent (see details on free probability in section 2.4 p. 77 of TULINO and VERDÙ 2004). In the next section, we will use these transforms to compute approximations of the spectral distribution of various covariance matrices in terms of an M-P law with modified parameters. The R transform is related to the Stieltjes transform via $R(z) = S^{-1}(z) - 1/z$ where $S^{-1}(.)$ is the functional inverse of S such that $S^{-1}\big(S(z)\big) = z$. The R transform of the LSD of the sum of two matrices ($\mathbf{A}$ and $\mathbf{B}$) with entries independently drawn is the sum of the components' R transforms: $R_{\mathbf{A}+\mathbf{B}}(z) = R_{\mathbf{A}}(z) + R_{\mathbf{B}}(z)$. For the scaled M-P law, the R transform is

$$R_{\mathbf{H}.\mathbf{H}^*}(z) = \frac{\beta\, \zeta}{1 - z\, \zeta} \quad . \qquad \text{(A1.4)}$$

The S transform plays the exact same role as the R transform for products of semi-definite matrices. It is related to the η transform via $\Sigma(z) = -(z+1)/z\, \eta^{-1}(z+1)$ where $\eta^{-1}(.)$ is the functional inverse of η. The S transform of the LSD of the product of two positive semi-definite matrices ($\mathbf{A}$ and $\mathbf{B}$), with entries independently drawn, is $\Sigma_{\mathbf{A}.\mathbf{B}}(x) = \Sigma_{\mathbf{A}}(x)\Sigma_{\mathbf{B}}(x)$. For the scaled M-P law, the S transform is

$$\Sigma_{\mathbf{H}.\mathbf{H}^*}(z) = \frac{1}{(\beta + z)\, \zeta} \quad . \qquad \text{(A1.5)}$$

Another formula will prove particularly useful in what follows: for any $n \times n$ positive semi-definite matrix $\mathbf{T}$ whose LSD exists and has S transform $\Sigma_{\mathbf{T}}(z)$, we have (adapted from 2.216 p. 91 of TULINO and VERDÙ 2004):

$$\Sigma_{\mathbf{H}.\mathbf{T}.\mathbf{H}^*}(z) = \frac{z+1}{z+\beta} \Sigma_{\mathbf{H}^*.\mathbf{H}}\left(\frac{z}{\beta}\right) \Sigma_{\mathbf{T}}\left(\frac{z}{\beta}\right) = \frac{1}{\zeta(z+\beta)} \Sigma_{\mathbf{T}}\left(\frac{z}{\beta}\right) \quad . \qquad \text{(A1.6)}$$

The last transform used here is the D transform, introduced in (BENAYCH-GEORGES and NADAKUDITI 2011). It will prove useful to predict the behavior of the maximal eigenvalue when the entries h_{ij} have non zero mean. In this article, it is defined in terms of the distribution of singular values of $\mathbf{H}$ (when $\mathbf{H} = \mathbf{H}_0$ has entries with a zero mean). As the singular values of $\mathbf{H}$ are simply the square roots of the non zero eigenvalues of $\mathbf{H}.\mathbf{H}^*$, the D transform in (BENAYCH-GEORGES and NADAKUDITI 2011) can also be expressed in terms of the spectral distribution of $\mathbf{H}.\mathbf{H}^*$. More precisely, let $\varphi(z) = \int_a^b z/(z^2 - x)\rho(x)dx = \eta(-1/z^2)/z$, the D transform is $D(z) = \varphi(z)(\varphi(z)/\beta + (1 - 1/\beta)/z)$ which, for the M-P law, yields

$$D_{\mathbf{H}.\mathbf{H}^*}(z) = \frac{z^2}{2\beta\zeta^2}\left(\left(1 - \sqrt{1 + \frac{\zeta\left((\beta-1)^2\zeta - 2z^2(1+\beta)\right)}{z^4}}\right) - (1+\beta)\zeta\right) \quad . \qquad \text{(A1.7)}$$

In what follows we use these transforms and their approximations to derive an M-P law approximation for the spectral distribution of $\mathbf{M}$ when $1 \ll n \ll p$.

B) Spectral distribution of $\mathbf{M}$ with high phenotypic integration

In this section, I study the matrix $\mathbf{M}$ of mutational covariance among optimized traits, under the model of integrative phenotypic network described in **Figure 1**. I first describe its structure in detail, then derive an approximation of its LSD in terms of an M-P law.

***Structure of the mutational covariance matrix M*:** Matrix $\mathbf{B} = \{b_{ij}\}_{i\in[1,n],j\in[1,p]}$ is an $n \times p$ matrix of pathway coefficients given by the first derivatives of the developmental function about the parent phenotype. As explained in the main text, the $1 \times p$ vector $\mathbf{b}_i = \{b_{ij}\}_{j\in[1,p]}$ of the p pathway coefficients determining a given trait y_i is a single draw from a multivariate distribution. This distribution has mean vector $\boldsymbol{\mu}_\mathbf{B}$ and positive-definite $p \times p$ covariance matrix $\mathbf{C}_\mathbf{B}$. Otherwise the nature of these distributions is unspecified. As its entries are randomly distributed, $\mathbf{M} = \mathbf{B}.\mathbf{V}.\mathbf{B}^*$ has the structure of a sample covariance matrix (BAI and SILVERSTEIN 2010).

In order to go any further, we must characterize the structure of $\mathbf{M}$ in more mathematical detail. Let us first ignore any potential bias in the b_{ij} ($\boldsymbol{\mu}_\mathbf{B} = \mathbf{0}$) and denote by $\mathbf{B}_\mathbf{0}$ the matrix of pathway coefficients in this case. By assumption (8), the matrix $\mathbf{B}_0$ can be decomposed into the product $\mathbf{B}_\mathbf{0} = \mathbf{H}.\mathbf{A}$ where $\mathbf{H}$ is an $n \times p$ matrix with independent entries h_{ij} with mean 0 and variance $V(h_{ij}) = 1/n$, and $\mathbf{A}$ is a $p \times p$ matrix introduced to generate the suitable covariance among b_{ij}'s (matrix bending). By this definition, $\mathbf{C}_\mathbf{B} = E(\mathbf{B}.\mathbf{B}^*) = \mathbf{A}^*.\mathbf{A}/n$, which is positive-definite as required. The matrix $\mathbf{A}$ can thus be set as $\mathbf{A} = \sqrt{n}\,\mathbf{C}_\mathbf{B}^{1/2}$, the Cholesky decomposition of $n\,\mathbf{C}_\mathbf{B}$, but there will typically be many other possibilities. By definition, a given draw of the matrix $\mathbf{B}_0$ corresponds to a given draw of the matrix $\mathbf{H}$. This matrix $\mathbf{H}$ is

the building block of most models of Random Matrix Theory, and we derive the structure of $\mathbf{B}.\mathbf{V}.\mathbf{B}^*$ in terms of this building block.

From now on, we take the expectation $E(.)$ to mean the expected outcome of a given draw of the random coefficients b_{ij} (and the corresponding h_{ij}). Importantly, the random entries in matrix $\mathbf{H}.\mathbf{H}^*$ are independent of $\mathbf{C_B}$. If we now consider the general model where $\boldsymbol{\mu_B} \neq \mathbf{0}$, the bias in the distribution of the pathway coefficients b_{ij} boils down to adding an $n \times p$ matrix $\mathbf{U_B}$ to $\mathbf{B_o}$. We can write $\mathbf{B} = \mathbf{B_o} + \mathbf{U_B}$ where $\mathbf{B_o} = \mathbf{H}.\mathbf{A}$ and matrix $\mathbf{U_B}$ has all its n line vectors equal to $\boldsymbol{\mu_B}$, so by construction $\mathbf{U_B}$ has rank 1.

We can now derive the structure of the random matrix $\mathbf{B}.\mathbf{V}.\mathbf{B}^*$: it can be decomposed into $\mathbf{B}.\mathbf{V}.\mathbf{B}^* = \mathbf{K}.\mathbf{K}^*$ where $\mathbf{K} = \mathbf{K}_\mathrm{o} + \mathbf{U}_\mathrm{K}$ with $\mathbf{K}_\mathrm{o} = \mathbf{H}.\mathbf{A}.\mathbf{V}^{1/2}$ and $\mathbf{U_K} = \mathbf{U_B}.\mathbf{V}^{1/2}$. The singular values of $\mathbf{K}_\mathrm{o}$ are the square roots of the eigenvalues of $\mathbf{K}_\mathrm{o}.\mathbf{K}_\mathrm{o}^* = \mathbf{H}.\mathbf{W}.\mathbf{H}^*$, where $\mathbf{W} = \mathbf{A}.\mathbf{V}.\mathbf{A}^*$ is a positive-definite $p \times p$ matrix. The spectral distribution of $\mathbf{W}$ is equal to that of $n\,\mathbf{V}.\mathbf{C}_\mathrm{B}$. Like $\mathbf{U_B}$, matrix $\mathbf{U_K}$ is a rank 1 matrix by construction. Indeed, basic properties of the rank of matrix products imply that $0 < rank(\mathbf{U_K}) \leq \min\big(rank(\mathbf{U_B}), rank(\mathbf{V})\big) = \min(1, p) = 1$. Therefore, $\mathbf{U_K}$ has a single non-zero singular value θ which can be computed by using the fact that $\mathbf{U_K}.\mathbf{U_K^*}$ then has a single non-zero eigenvalue equal to θ^2 by definition. Therefore, its matrix trace $Tr(.)$ must satisfy $\theta^2 = Tr(\mathbf{U_K}.\mathbf{U_K^*})$. By construction of matrix $\mathbf{U_K} = \mathbf{U_B}.\mathbf{V}^{1/2}$, we also have $Tr(\mathbf{U_K}.\mathbf{U_K^*}) = \boldsymbol{\mu}_\mathbf{B}^*.\mathbf{V}.\boldsymbol{\mu_B}$ so the unique singular value of $\mathbf{U_K}$ is $\theta = \sqrt{\boldsymbol{\mu}_\mathbf{B}^*.\mathbf{V}.\boldsymbol{\mu_B}}$.

To summarize, matrix $\mathbf{M}$ has the same eigenvalues as a (non-standard) sample covariance matrix $\mathbf{K}.\mathbf{K}^*$ where $\mathbf{K} = \mathbf{K}_\mathrm{o} + \mathbf{U_K}$ with $\mathbf{K}_\mathrm{o} = \mathbf{H}.\mathbf{W}^{1/2}$ a 'standard' random matrix $\mathbf{H}$ multiplied by a positive-definite 'constant' matrix $\mathbf{W} = \mathbf{A}.\mathbf{V}.\mathbf{A}^*$. Matrix $\mathbf{U_K}$ is of rank 1 and its unique singular value is $\theta = \sqrt{\boldsymbol{\mu}_\mathbf{B}^*.\mathbf{V}.\boldsymbol{\mu_B}}$. Both components are $n \times p$ matrices. This whole decomposition argument is exemplified in the notebook **file S3**.

Existence of the LSD of M: A cornerstone result of RMT is that the spectral distribution of $\mathbf{H}.\mathbf{H}^*$ admits a limit when $n\,p \to \infty$, given by the M-P law: $\lambda_{H.H^*} \sim MP(p/n, 1)$. In the absence of a bias among pathway coefficients ($\boldsymbol{\mu_B} = \mathbf{0}$), $\mathbf{M} = \mathbf{B_o}.\mathbf{B_o^*} = \mathbf{H}.\mathbf{W}.\mathbf{H}^*$ is a non-standard sample covariance matrix: as the entries of $\mathbf{H}$ are independent of $\mathbf{W}$ **,** the existence of an LSD for this matrix is also certain (chapter 4 of BAI and SILVERSTEIN 2010). This LSD is fully determined by that of $\mathbf{W}$ and $\mathbf{H}.\mathbf{H^T}$ taken separately, and is still independent of the nature of the distributions of the entries in $\mathbf{H}$. Tools from RMT can then be used to compute explicit approximations for the LSD of such a matrix product. The effect of bias, as we will see, does not affect the existence of an LSD, as it only modifies the leading eigenvalue of $\mathbf{M}$.

Approximation to the LSD of matrix $\mathbf{M}$ when $\boldsymbol{\mu_B} = \mathbf{0}$: Here, I describe how the LSD of the sample covariance matrix $\zeta\,\mathbf{H}.\mathbf{H}^*$ is modified by inner multiplication by the matrix $\mathbf{W}$, yielding the LSD of $\mathbf{M} = \mathbf{H}.\mathbf{W}.\mathbf{H}^*$.

As shown above, the matrix $\mathbf{W} = \mathbf{A}.\mathbf{V}.\mathbf{A}^*$ is an unspecified positive-definite matrix, whose eigenvalues are the same as those of $n\,\mathbf{V}.\mathbf{C_B}$. Overall, $\mathbf{W}$ gathers all the correlation among mutation effects on mutable traits and among pathway coefficients within $\mathbf{B}$. It is impossible to have a general *a priori* knowledge on the structure of $\mathbf{W}$, so we must rely on an approximate treatment under less general conditions, in order to keep as much generality as possible regarding $\mathbf{W}$. This is possible if we assume that the phenotypic integration from mutable to optimized traits is high (our assumption (4) : $\beta = p/n \gg 1$).

First, let us note that the LSD of $\mathbf{H}.\mathbf{W}.\mathbf{H}^*$ always exists as long as $\mathbf{W}$ is positive-definite and has a bounded spectral distribution (Chapter 4 in BAI and SILVERSTEIN 2010). Under the additional assumption that $n/p \to 0$ (or $\beta \to \infty$), we can approximate eq. (A1.6) to obtain a simple limit for this LSD. Let $\eta_{\mathbf{W}}(z)$ be the (unspecified) η transform of the spectral distribution of $\mathbf{W}$, and let $\eta_{\mathbf{W}}^{-1}(x)$ be the functional inverse of this transform. From the relationship between η and S transforms ($\Sigma(z) = -(z+1)/z\,\eta^{-1}(z+1)$) we can express eq. (A1.6) as $\Sigma_{\mathbf{H}.\mathbf{W}.\mathbf{H}^*}(x) = -\eta_{\mathbf{W}}^{-1}(1+x/\beta)/x$. Assuming that β is large, we can then take a series expansion of $\eta_{\mathbf{W}}^{-1}(1+x/\beta)$ for small x/β which gives

$$\eta_{\mathbf{W}}^{-1}\left(1+\frac{x}{\beta}\right) = \eta_{\mathbf{W}}^{-1}(1) + \frac{x}{\beta}\eta_{\mathbf{W}}^{-1\prime}(1) + \frac{x}{2\beta^2}\,\eta_{\mathbf{W}}^{-1\prime\prime}(1) + o\left(\frac{x}{\beta^2}\right) \quad . \qquad \text{(A1.8)}$$

This Taylor series can be expressed in terms of the derivatives of $\eta_{\mathbf{W}}(z)$ taken at $z = \eta_{\mathbf{W}}^{-1}(1) = 0$ (e.g. applying the method in KOEPF 1994), and we find that $\eta_{\mathbf{W}}^{-1\prime}(1) = 1/\eta_{\mathbf{W}}{}'(0)$ and $\eta_{\mathbf{W}}^{-1\prime\prime}(1) = -\eta_{\mathbf{W}}{}''(0)/(\eta_{\mathbf{W}}{}'(0)^3)$. Recalling that the η transform is a moment generating function for the LSD ($\eta[-z] = \sum_{k=0}^{\infty} z^k \mathrm{E}(\lambda^k)$), we have $\eta_{\mathbf{W}}^{-1\prime}(1) = -1/\zeta_{\mathbf{W}}$ and $\eta_{\mathbf{W}}^{-1\prime\prime}(1) = 2(1+cv_{\mathbf{W}}^2)/\zeta_{\mathbf{W}}$, where $\zeta_{\mathbf{W}}$ is the mean of the eigenvalues of $\mathbf{W}$ and $cv_{\mathbf{W}}$ is their coefficient of variation. Plugging these expressions into eq. (A1.8), we obtain a simple expression for the inverse η transform, which depends only on the mean and coefficient of variation of the eigenvalues of $\mathbf{W}$:

$$\eta_{\mathbf{W}}^{-1}\left(1+\frac{x}{\beta}\right) \approx \frac{x^2\left(1+cv_{\mathbf{W}}^2\right) - x\,\beta}{\zeta_{\mathbf{W}}\,\beta^2} + o\left(\frac{x}{\beta}\right)^2 \quad . \qquad \text{(A1.9)}$$

Plugging eq. (A1.9) into the formula $\Sigma_{\mathbf{H}.\mathbf{W}.\mathbf{H}^*}(x) = -\eta_{\mathbf{W}}^{-1}(1+x/\beta)/x$, we obtain a linear function of x: introducing the parameters $\zeta_e = \zeta_{\mathbf{W}}\left(1+cv_{\mathbf{W}}^2\right)$ and $\beta_e = \beta/\left(1+cv_{\mathbf{W}}^2\right)$:

$$\Sigma_{\mathbf{M}}(x) = \Sigma_{\mathbf{H}.\mathbf{W}.\mathbf{H}^*}(x) = -\frac{\eta_{\mathbf{W}}^{-1}\left(1+\frac{x}{\beta}\right)}{x} = \frac{1}{\zeta_e\beta_e}\left(1-\frac{x}{\beta_e}\right) + o\left(\frac{x}{\beta^2}\right) \quad , \qquad \text{(A1.10)}$$

which can be rearranged into $\Sigma_{\mathbf{M}}(x) = \zeta_e^{-1}\left(x+\beta_e\right)^{-1} + \mathrm{o}(x\beta^{-2})$. To first order in x/β^2, we retrieve the S transform of an M-P law (eq.(A1.5)) with parameters ζ_e and β_e.

Therefore, we obtain a simple approximation to the LSD of the mutational covariance, accounting for arbitrary covariance matrices among mutation effects on metabolic traits or among

pathway coefficients. This approximation is to the second order in $1/\beta$ and is thus relatively robust: it remains valid even when p is not much larger than n. All along, we made no assumption on the mean or variance of the spectral distribution of $\mathbf{W}$ (on $\zeta_{\mathbf{W}}$ or $cv^2_{\mathbf{W}}$). Therefore, eq. (A1.10) remains valid even with high correlations and heterogeneity within $\mathbf{V}$ or $\mathbf{B}$. As expected, we retrieve the original M-P law when $\mathbf{W} = \zeta_{\mathbf{W}}\mathbf{I}_p$: in this case $cv^2_{\mathbf{W}} = 0$ (all the eigenvalues of $\mathbf{W}$ are equal to their mean $\zeta_{\mathbf{W}}$) so that $\beta_e = \beta$ and $\zeta_e = \zeta_{\mathbf{W}}$, and multiplication by $\mathbf{W}$ results in a mere scaling.

Obviously, this derivation cannot claim full mathematical rigor: caution might be necessary on how the approximation of the S transform implies convergence in distribution to the corresponding LSD. Yet, **Figure 4.a** and other extensive simulations (not shown) do suggest that this approximation is valid: the convergence of the spectral distribution of $\mathbf{M}$ to this modified M-P law is good, even with very high heterogeneity in the eigenvalues of $\mathbf{W}$ (here $cv^2_{\mathbf{W}} = 2.1$).

Effect of a non-zero mean in b_{ij}: The above treatment describes the LSD of $\mathbf{K_o}.\mathbf{K_o^*}$ that equals that of $\mathbf{M}$ whenever the $\mathbf{b}_i$'s are unbiased ($\boldsymbol{\mu}_{\mathrm{B}} = \mathbf{0}$), see main text. However, in general we want to allow for arbitrary $\boldsymbol{\mu}_{\mathrm{B}} \neq \mathbf{0}$. To do so, we rely on Theorem 2.9 of Benaych-Georges & Nadakuditi (2011) on the effect of small rank perturbations on the singular value distribution of random rectangular matrices (such as $\mathbf{K_o}$). As the singular values of $\mathbf{K}$ are simply the square roots of the eigenvalues of $\mathbf{K}.\mathbf{K}^*$ and thus of $\mathbf{M}$, this provides the required result.

Let us order the eigenvalues of $\mathbf{M}$ in decreasing order ($\lambda_1 \geq \lambda_2 .. \geq \lambda_n > 0$). To generate $\mathbf{K}$, the random matrix $\mathbf{K}_{\mathrm{o}}$ (the standard one with zero mean entries) is "additively perturbed" by a small rank matrix $\mathbf{U}_{\mathrm{K}}$: $\mathbf{K} = \mathbf{K}_{\mathrm{o}} + \mathbf{U}_{\mathrm{K}}$ has a number of singular values that differ from those of $\mathbf{K}_{\mathrm{o}}$. When $\mathbf{K}$ is a large random matrix, these differing eigenvalues are the r largest in general, where r is the rank of the perturbation matrix (BENAYCH-GEORGES and NADAKUDITI 2011). The perturbation $\mathbf{U_K}$ is of rank 1 with a unique singular value $\theta = \sqrt{\boldsymbol{\mu}_{\mathbf{B}}^*.\mathbf{V}.\boldsymbol{\mu}_{\mathbf{B}}}$ (see above). Therefore, only the leading singular value $\sqrt{\lambda_1}$ is affected by the perturbation, while all the lower singular values $\sqrt{\lambda_i}$ of $\mathbf{K}$ retain the same distribution as those of $\mathbf{K}_{\mathrm{o}}$. The leading singular value $\sqrt{\lambda_1}$ shows a phase transition behavior that is determined by what the Benaych-Georges and Nadakuditi dubbed the D-transform of the singular value distribution, which is the expectation $D(z) = \mathrm{E}_{\sigma}(z/(z^2 - \sigma_o))$ over the distribution of the singular values σ_o of the unperturbed matrix, in their notation. In our context, the unperturbed matrix is $\mathbf{K_o}$, and it is useful to express $D(z)$ in terms of the eigenvalues $\lambda_o = \sigma_o^2$ of $\mathbf{K_o}$. This gives: $D(z) = \mathrm{E}_{\lambda_o}(z/(z^2 - \lambda_o))$ over the spectral distribution of $\mathbf{K_o}.\mathbf{K_o^*}$. Now as we have seen above, the spectral distribution of $\mathbf{K_o}.\mathbf{K_o^*}$ can be approximated by an M-P law: $\lambda_o \sim MP(\beta_e, \zeta_e)$, so the corresponding D transform is $D_{\mathbf{H}.\mathbf{H}^*}(z)$ given above (eq.(A1.7)) for the M-P law.

To summarize: the effect of the bias in the b_{ij} is (i) no effect on the bulk of lower eigenvalues ($\lambda_{i\neq 1} \sim MP(\beta_e, \zeta_e)$) and (ii) a "phase transition" behavior for the leading eigenvalue λ_1. Denote $\theta_{th} = 1/\sqrt{D_{\mathbf{H}.\mathbf{H}^*}(\max \sigma_o)}$ the functional inverse of the D transform, taken at $\max \sigma_o = \max \sqrt{\lambda_o}$. Whenever $0 \leq \theta^2 < \theta_{th}^2$, all eigenvalues (including λ_1) pertain to the M-P law so that $\lambda_1 \to \max \lambda_o$

which is the expected maximum of the spectral distribution of the unperturbed matrix. However, whenever $\theta^2 > \theta_{th}^2$, λ_1 rises above the bulk of smaller eigenvalues, to a higher value $\sqrt{\lambda_1} \rightarrow D_{\mathbf{H}.\mathbf{H}^*}^{-1}(1/\theta^2)$.

Using the M-P law approximation to the LSD of $\mathbf{K_o}.\mathbf{K_o^*}$ ($\lambda_o \sim MP(\beta_e, \zeta_e)$), the maximum of the M-P law is $\max \sigma_o^2 = \max \lambda_o = \left(1 + \sqrt{\beta_e}\right)^2 \zeta_e$, so the threshold for the phase transition is

$$\theta_{th} = \frac{1}{\sqrt{D_{\mathbf{H}.\mathbf{H}^*}\left(\max \sqrt{\lambda_o}\right)}} = \beta_e^{1/4} \sqrt{\zeta_e} \quad . \qquad \text{(A1.11)}$$

The limit reached by the leading eigenvalue beyond the phase transition depends on the functional inverse of the D transform of λ_o, which, for the M-P law, is (from eq. (A1.7)):

$$D_{\mathbf{H}.\mathbf{H}^*}^{-1}(z) = \sqrt{\frac{(1 + z\,\zeta_e)(1 + z\,\zeta_e\,\beta_e)}{z}} \quad . \qquad \text{(A1.12)}$$

This yields the limit:

$$\lambda_1 \underset{\theta > \theta_{th}}{\rightarrow} \left(D_{\mathbf{H}.\mathbf{H}^*}^{-1}\left(\frac{1}{\theta^2}\right)\right)^2 = \frac{(\zeta_e + \theta^2)(\beta_e \zeta_e + \theta^2)}{\theta^2} \quad . \qquad \text{(A1.13)}$$

We can express this result in more intuitively amenable terms. Define $cv = \sqrt{Tr(\mathbf{C}_\mathrm{B}.\mathbf{V})/(\boldsymbol{\mu}_\mathrm{B}^*.\mathbf{V}.\boldsymbol{\mu}_\mathrm{B})}$, which is analogous to a coefficient of variation of the means μ_j but modified by the mutational covariance $\mathbf{V}$. When $\mathbf{V} \propto \mathbf{I}_p$, this is exactly the mean coefficient of variation of the linear coefficients b_{ij} across mutable traits x_j. Recall that $\zeta_e = \zeta_\mathbf{W}\, p/p_e$ and that the $p \times p$ matrix $\mathbf{W}$ has the same spectrum as $n\,\mathbf{V}.\mathbf{C}_\mathrm{B}$: its mean eigenvalue is therefore $\zeta_\mathbf{W} = Tr(n\,\mathbf{V}.\mathbf{C}_\mathrm{B})/p = n/p\; Tr(\mathbf{C}_\mathrm{B}.\mathbf{V})$. Putting this together, we get

$$\frac{\theta^2}{\theta_{th}^2} = \frac{\theta^2}{\sqrt{\beta_e}\zeta_e} = \frac{\sqrt{n\,p_e}}{cv^2} \quad . \qquad \text{(A1.14)}$$

We see that when the bias in the coefficients b_{ij} is small enough, cv^2 is large enough to outweigh $\sqrt{n\,p_e}$. By a "small enough bias", we mean specifically that the cumulated variance of mutational effects on mutable traits, and of the linear coefficients is larger than $\boldsymbol{\mu}_\mathrm{B}^*.\mathbf{V}.\boldsymbol{\mu}_\mathrm{B}$. Otherwise, a phase transition appears and the leading eigenvalue rises above the bulk of lower eigenvalues, by a factor

$$\frac{\lambda_1}{\mathrm{E}(\lambda_o)} \underset{\theta > \theta_{th}}{\rightarrow} \left(1 + \frac{n}{cv^2}\right)\left(1 + \frac{cv^2}{p_e}\right) \underset{p_e \rightarrow \infty}{\rightarrow} 1 + \frac{n}{cv^2} \quad . \qquad \text{(A1.15)}$$

This is our eq. (6), except that we replaced λ_o (all eigenvalues of $\mathbf{K}.\mathbf{K}^*$ when $\boldsymbol{\mu}_\mathbf{B} = \mathbf{0}$) by $\lambda_{i\neq 1}$ (the set of $n-1$ smallest eigenvalues, namely the "bulk" eigenvalues in the general situation (arbitrary $\boldsymbol{\mu}_\mathbf{B}$). The latter is distributed as λ_o, except that it is depleted of its maximal value λ_1, so it is very slightly (unnoticeably if $n \gg 1$) biased downwards.

The accuracy of this result is checked **Figure 4b**. In this figure, the actual LSD of $\mathbf{M}$ is not exactly the M-P law because of the covariances in mutable traits ($\mathbf{V} \neq \mathbf{I}_p$) and pathway coefficients ($\mathbf{C_B} \neq \mathbf{I}_p$), so that $\mathbf{H}.\mathbf{W}.\mathbf{H}^* \neq \zeta_W \mathbf{H}.\mathbf{H}^*$. However, the prediction for the phase transition behavior, which is based on the M-P law approximation (derivation above), is accurate. This is simply because the behavior of λ_1 is entirely determined by the spectral distribution of the unperturbed matrix (λ_o), which was shown to be accurately captured by the M-P law approximation as long as $n \ll p$ (e.g. **Figure 4a**).

Extensions to allow for correlations among the rows of B: In the end of the Discussion, I stress the fact that in the model as formulated, it is impossible to have substantial correlations among the rows of $\mathbf{B}$, namely among the b_{ij}, across i, for a given j. This is because the covariance $\mathbf{C_B}$ is of full rank p. A possible way to incorporate such correlations was suggested by one of the reviewers, by letting $\mathbf{B} = \mathbf{A_1}.\mathbf{H}.\mathbf{A_2}$ still $n \times p$, with $\mathbf{A_1}$ an $n \times n$ matrix and $\mathbf{A_2}$ a $p \times p$ matrix, both invertible (our model so far corresponds to $\mathbf{A_1} = \mathbf{I}_n$ and $\mathbf{A_2} = \mathbf{A}$). I provide a quick analysis of this case, but merely to illustrate how extensions can be made, and mostly in the form of conjectures rather than proofs. A full treatment of this more general model is beyond the scope of this article.

The covariance of b_{ij} among the rows i is now given by the $p \times p$ matrix $\mathbf{C_2} = E(\mathbf{B}^*.\mathbf{B}) = E\big((\mathbf{A_2^*}.\mathbf{H}^*.\mathbf{A_1^*}).(\mathbf{A_1}.\mathbf{H}.\mathbf{A_2})\big)$. When $\mathbf{A_1} = \mathbf{I}_n$ and $\mathbf{A_2} = \mathbf{A}$ (our former model), this gives $\mathbf{C_2} = \mathbf{C_B} = \mathbf{A}^*.E(\mathbf{H}^*.\mathbf{H}).\mathbf{A} = \mathbf{A}^*.\mathbf{A}/n$, as $E(\mathbf{H}^*.\mathbf{H}) = 1/n$. Even in the general case ($\mathbf{A_1} \neq \mathbf{I}_n$), with $n \ll p$ and $\mathbf{A_1}.\mathbf{A_1^*}$ of full rank n, we have convergence to $\mathbf{C_2} \propto \mathbf{A_2^*}.\mathbf{A_2}/n$. Conversely, the covariance of the b_{ij} among the columns j is now given by the $n \times n$ matrix $\mathbf{C_1} = E(\mathbf{B}.\mathbf{B}^*) = E\big((\mathbf{A_1}.\mathbf{H}.\mathbf{A_2}).(\mathbf{A_2^*}.\mathbf{H}^*.\mathbf{A_1^*})\big)$. In our former model $\mathbf{C_1} = E(\mathbf{H}.\mathbf{A_2}.\mathbf{A_2^*}.\mathbf{H}^*)$: as $n/p \to 0$ with $\mathbf{A_2}.\mathbf{A_2^*}$ of full rank p, we get $\mathbf{C_1} \propto \mathbf{I}_n$, and in the general case, we get $\mathbf{C_1} \propto \mathbf{A_1}.\mathbf{A_1^*}$. Therefore the introduction of matrix $\mathbf{A_1}$ does introduce a potential for correlations among the rows i, which is effectively negligible otherwise (provided $n \ll p$ and $\mathbf{C_B}$ is of full rank p), as conjectured.

Let us now see how the mutational covariance $\mathbf{M}$ is affected by multiplication of $\mathbf{B}$ on the left by $\mathbf{A_1}$. In our extension, if we let $\mathbf{M} = \mathbf{H}.\mathbf{A_2}.\mathbf{V}.\mathbf{A_2^*}.\mathbf{H}^*$ be the former form of the mutational covariance, we now have $\mathbf{M'} = \mathbf{A_1}.\mathbf{M}.\mathbf{A_1^*}$, as the new mutational covariance. The LSD of $\mathbf{M'}$ is the same as that of $\mathbf{C_1}.\mathbf{M}$ where we define $\mathbf{C_1} = \mathbf{A_1}.\mathbf{A_1^*}$ which is approximately the covariance matrix of b_{ij} among rows i. As the elements in $\mathbf{C_1}$ are independent of those in $\mathbf{H}$, $\mathbf{M'}$ has S transform given by $\Sigma_{\mathbf{M'}}(x) = \Sigma_{\mathbf{M}}(x)\Sigma_1(x)$ where Σ_1 is the S transform of the LSD of $\mathbf{C_1}$.

We must now approximate each S transform. We have seen above that the LSD of $\mathbf{M}$ is approximately the M-P law so that $\Sigma_{\mathbf{M}}(x) = 1/(\zeta_e(x+\beta_e))$ with ζ_e, β_e given above. I could not derive a general result for $\Sigma_1(.)$ without making an additional assumption. I assume that the spectral distribution

of $\mathbf{C_1}$ is not widely spread. More precisely, define$\lambda_{\mathbf{C_1}}$ and set, without loss of generality, that $E(\lambda_{\mathbf{C_1}}) = 1$ (all the scaling can be absorbed into ζ_e). I assume that (i) $v_1 = V(\lambda_{\mathbf{C_1}}) \ll 1$ and that (ii) the higher raw moments of the LSD of $\mathbf{C_1}$ scale with this variance $E(\lambda_{\mathbf{C_1}}^k) = 0(v_1^k)$ for all $k \geq 2$. This simply implies that the eigenvalues of $\mathbf{C_1}$ are not too spread apart, corresponding to mild correlations among the rows of $\mathbf{B}$. In this case, we can find a first order approximation for the S transform of the LSD of $\mathbf{C_1}$ and a corresponding approximation for the LSD of $\mathbf{M'}$ again in terms of an M-P law with modified parameters. As for the rest of the Appendix, details of the computations can be found in the notebook **file S3**.

Define $\eta_1(x) = E(1/(1 + \lambda_{C_1} x))$ the η transform of the LSD of $\mathbf{C_1}$ and $\eta_1^{-1}(.)$ the functional inverse such that $z = \eta_1\left(\eta_1^{-1}(z)\right)$. First, take the Taylor series expansion for ratio: $1/(1 + \eta_1^{-1}(z)\,\lambda_{\mathbf{C_1}})$ to order $\lambda_{\mathbf{C_1}}^2$; then take expectations with respect to the distribution of $\lambda_{\mathbf{C_1}}$: this yields

$$z = \eta_1\left(\eta_1^{-1}(z)\right) = E\left(\frac{1}{1 + \eta_1^{-1}(z)\,\lambda_{\mathbf{C_1}}}\right) = \frac{\left(v_1\,\eta_1^{-1}(z)\right)^2}{\left(1 + \eta_1^{-1}(z)\right)^3} + \frac{1}{1 + \eta_1^{-1}(z)} + O\left(v_1^{3/2}\right) \quad , \quad \text{(A1.16)}$$

Solving for $\eta_1^{-1}(z)$, we retrieve four solutions, but only one has the correct behavior when $v_1 \to 0$: namely, it converges to the solution when all $\lambda_{\mathbf{C_1}} = E(\lambda_{\mathbf{C_1}}) = 1$, which is $\eta_1^{-1}(z) = (1 - z)/z$. Taking this solution, we can compute the S transform $\Sigma_1(x)$ of $\mathbf{C_1}$ under the approximation. The expression is analytic but unpractical. Yet, after taking a series of $1/\Sigma_1(x)$ to leading order in v_1 , we obtain an approximate expression:

$$\Sigma_1(x) = -\frac{x+1}{x}\eta_1^{-1}(x+1) \underset{v_1 \ll 1}{\approx} \frac{1}{1 + v_1\,x} \quad . \quad \text{(A1.17)}$$

Interestingly, this approximation is exact when $\mathbf{C_1}$ is a sample covariance matrix, with LSD given by the M-P law ($\lambda_{\mathbf{C_1}} \sim MP(\beta_1, 1/\beta_1)$ and arbitrary $\beta_1 > 0$). Note that, in this particular case, this result should thus be valid for an arbitrary level of variation in the eigenvalues of $\mathbf{C_1}$ (arbitrary $\beta_1 > 0$).

Using eq. (A1.17), we can compute the S transform of the LSD of $\mathbf{M'}$: $\Sigma_{\mathbf{M'}}(x) = \Sigma_{\mathbf{M}}(x)\Sigma_1(x) \approx 1/\left((x + \beta_e)\,\zeta_e\,(1 + v_1 x)\right)$. In general, this S transform is not related to any specific form of random matrix. However, when the ratio index of $\mathbf{M} = \mathbf{H}_e.\mathbf{H}_e^*$ is large ($\beta_e \gg 1$), we retrieve a simple approximation in terms of an MP law. Taking the leading order when $v_1 = \mathrm{o}(1)$ while $\beta_e v_1 = 0(1)$ ($v_1 \ll 1$ and $\beta_e \gg 1$), we get the S transform of an M-P law with modified parameters $\zeta_e' = \zeta_e\,(1 + \beta_e v_1)$ and $\beta_e' = \beta_e/(1 + \beta_e v_1)$:

$$\Sigma_{\mathbf{M'}}(x) = \Sigma_{\mathbf{M}}(x)\Sigma_1(x) \underset{v_1 \ll 1}{\approx} \frac{1}{\zeta_e(x + \beta_e)\,(1 + v_1 x)} \underset{\beta_e \gg 1}{\approx} \frac{1}{\zeta_e'(1 + \beta_e'\,x)} \quad . \quad \text{(A1.18)}$$

As expected, we retrieve the original MP law when $v_1 \to 0$ (our former model: $\mathbf{M'} = \mathbf{M}$, with $\beta_e' = \beta_e$). The approximation in eq. (A1.18) must break whenever v_1 is large enough or β_e small enough that $0 <$

$\beta'_e < 1$. Indeed, in this case, the approximation would predict that a portion $(1 - \beta'_e)$ of eigenvalues should be zero whereas in fact there are none ($\beta_e > 1$ and $\mathbf{C_1}$ is positive-definite). This sets a limit for the validity of eq. (A1.18): we must have $0 < v_1 < (\beta_e - 1)/\beta_e$, i.e. $v_1 \leq 1$ in the best case scenario ($\beta_e \to \infty$).

To conclude, the effect of correlations in the rows of $\mathbf{B}$ is thus to further reduce the shape parameter of the LSD of the mutational covariance, while retaining the M-P law structure. When $n/p \to \infty$, so that $\beta_e \propto \beta \to \infty$, we retrieve a finite shape parameter this time:

$$\beta'_e = \frac{\beta_e}{(1 + \beta_e v_1)} \underset{\beta_e \to \infty}{\approx} \frac{1}{v_1} \quad . \qquad \text{(A1.19)}$$

Then, there is mild anisotropy, all the more as v_1 gets larger, namely as the rows of $\mathbf{B}$ get more correlated. The effective dimensionality in the sense of the matching moment approximation in (MARTIN and LENORMAND 2006), is $n_e = n/(1 + cv(\lambda)^2)$ where $cv(\lambda)$ is the coefficient of variation of the eigenvalues of $\mathbf{M}'$. With our M-P law approximation we have $cv(\lambda)^2 = 1/\beta'_e$ so that

$$n_e \underset{n \ll p}{\approx} \frac{n}{(1 + v_1)} \quad . \qquad \text{(A1.20)}$$

In the main text, I refer directly to $v_1 = cv_b^2$ as the coefficient of variation of the eigenvalues of the covariance matrix of the b_{ij} among rows i, to avoid stating the scaling $E(\lambda_{C_1}) = 1$ that was made here for mere notational simplicity. Simulations (not shown) suggest that this new M-P law approximation is indeed accurate as long as v_1 is small and β_e is large ($0 < v_1 < (\beta_e - 1)/\beta_e$).

References:

BAI, Z., and J. W. SILVERSTEIN, 2010 *Spectral Analysis of Large Dimensional Random Matrices, 2nd Edition*.

BENAYCH-GEORGES, F., and R. R. NADAKUDITI, 2011 The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. Advances in Mathematics **227:** 494-521.

KOEPF, W., 1994 Taylor polynomials of implicit functions, of inverse functions, and of solutions of ordinary differential equations. Complex Variables, Theory and Application: An International Journal **25:** 23-33.

MARTIN, G., and T. LENORMAND, 2006 A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. Evolution **60:** 893-907.

TULINO, A. M., and S. VERDÙ, 2004 *Random Matrix Theory and Wireless Communications*. Now Publishers.

WOLFRAM RESEARCH, I., 2012 Mathematica Edition: Version 9.0, pp. Wolfram Research, Inc., Champaign, Illinois.

File S2

Approximations for the DFE

We know that the limit spectral distribution of $\mathbf{M}$ is approximately given by the M-P law under reasonable assumptions (see **Appendix file S1**). Let us see the implications for the distribution of the fitness effect of mutations. Several of the analytical treatments can be checked using the Mathematica® (Wolfram Research 2012) notebook **file S3** (in freely readable [.cdf] format).

Stochastic representation of the DFE in the anisotropic FGM: Eq. (2) provides the stochastic representation of s as a function of each individual value of λ_i. Its derivation can be found elsewhere (Mathai and Provost 1992; Jaschke et al. 2004). It can be obtained in the basis of phenotype space where the mutational covariance is diagonal, via the change of basis $\mathbf{z} = \mathbf{Q}.\mathbf{y}$, where $\mathbf{Q}$ is the eigenbasis of $\mathbf{M}$. In this basis, the random variables dy_i in eq. (1) become dz_i, a set of n independent normal deviates: $dz_i \sim N\left(0, \sqrt{\lambda_i}\right)$. Eq. (1) can thus be rearranged into

$$s = \sum_{i=1}^{n} \frac{z_i^2}{2} - \sum_{i=1}^{n} \frac{\lambda_i}{2} \left(\frac{z_i}{\sqrt{\lambda_i}} + \frac{dz_i}{\sqrt{\lambda_i}} \right)^2 \quad , \qquad \textbf{(A2.1)}$$

where the variables $dz_i/\sqrt{\lambda_i} \sim N(0,1)$ are independent standard normal deviates. By definition of the non-central chi-square distribution: $\left(z_i/\sqrt{\lambda_i} + dz_i/\sqrt{\lambda_i}\right)^2 \sim \chi_1^2[z_i^2/\lambda_i]$, which leads to eq. (2). From this stochastic representation, one can directly obtain the approximations in eqs. (8) in the limit where all $\lambda_i = \tilde{\lambda}$ (or eq. (10) when only the $n-1$ lowest eigenvalues are equal to $\tilde{\lambda}$).

However, it is useful to study the distribution in more details, to understand why this isotropic approximation is in fact robust, considering that the eigenvalues are never exactly all equal. I do so via a generating function of the DFE.

Generating function of the DFE: the stochastic representation in eq. (2) directly yields a closed form expression for the cumulant-generating-function (CGF) of s : $\kappa_s(u) \equiv \log(E_s(e^{u\,s}))$. This CGF fully characterizes the DFE (all its cumulants). It is derived easily from the CGF of the non-central chi-square $\chi_1^2[v]$, which is $\kappa_1(u,v) = u\,v/(1-2\,u) - \log(1-2\,u)/2$. The CGF of the sum of independent variables is the sum of their CGFs, so eq. (2) yields

$$\kappa_s(u) = \frac{1}{2}\left(\sum_{i=1}^{n} \frac{u^2\, z_i^2 \lambda_i}{1+u\,\lambda_i} - \sum_{i=1}^{n} \log(1+u\,\lambda_i) \right) \quad . \qquad \textbf{(A2.2)}$$

As the CGF fully characterizes a distribution, the DFE is fully determined by the joint distribution of the eigenvalues of $\mathbf{M}$ (the λ_i's) and the position of the parental phenotype (the y_i's).

Link to the Shannon transform of the M-P law: At this point, we can note that a central quantity in eq. (A2.2) is $\nu_n(u) = 1/n \sum_{i=1}^{n} \log(1 + u\,\lambda_i)$, the average of $\log(1 + u\,\lambda)$ over the n eigenvalues, namely the Shannon transform of the spectral distribution of $\mathbf{M}$. For an optimal genotype ($s_o = 0$) all $z_i = 0$ and $\kappa_s(u) = -n/2\,\nu_n(u)$. For a suboptimal genotype, it is impossible to derive an equivalent expression. However, below phase transition ($\alpha \approx 1$), we may ignore any potential correlation, across traits i, between the maladaptation terms z_i^2 and the λ_i. Then, we can, first, introduce the derivatives of $\log(1 + u\,\lambda_i)$ with respect to u into $\kappa_s(u)$, then approximate

$$\kappa_s(u) = -\frac{n}{2}\,\nu_n(u) + \sum_{i=1}^{n} \frac{u^2\,z_i^2}{2}\,\partial_u \log(1 + u\,\lambda_i) \approx -\frac{n}{2}\,\nu_n(u) + u^2 \left(\frac{1}{n}\sum_{i=1}^{n} \frac{z_i^2}{2}\right)\left(\partial_u \sum_{i=1}^{n} \log(1 + u\,\lambda_i)\right)$$

Noting that $1/n \sum_{i=1}^{n} z_i^2/2 = s_o/n$ and that $\partial_u(\sum_{i=1}^{n} \log(1 + u\,\lambda_i)) = n\,\nu_n'(u)$, we thus get

$$\kappa_s(u) \approx u^2 s_o \nu_n{}'(u) - \frac{n}{2}\nu_n(u) \quad , \qquad \textbf{(A2.3)}$$

provided that $cov(z_i^2, \lambda_i) \approx 0$ and $\alpha \approx 1$. Therefore, the DFE obtains in terms of the spectral distribution of $\mathbf{M}$ (via $\nu_n(u)$), plus two parameters: s_o and n. The effect of maladaptation ($\mathbf{y}$) on the DFE is thus fully determined by the distance to the optimum ($s_o = \|\mathbf{y}\|^2/2$), not by the actual direction to the optimum: the model behaves *de facto* as an isotropic one. Of course the independence assumed between z_i^2 and λ_i across traits i, is never guaranteed, but it must become an accurate approximation once the λ_i become close to each other (convergence towards more isotropy as $\beta \to \infty$). Note also that (A2.3) is always exact for an optimal genotype ($s_o = 0$).

We can derive the equivalent expression beyond phase transition by simply separating the leading eigenvalue ($\lambda_1 = \alpha\,\tilde{\lambda}$) from the others: $\kappa_s(u) \approx \kappa_{n-1}(u) + \kappa_1(u)$ with

$$\begin{cases} \kappa_{n-1}(u) = u^2 s_{n-1} \nu_{n-1}{}'(u) - \dfrac{n-1}{2}\nu_{n-1}(u) \\ \kappa_1(u) = u^2 s_1 \nu_1{}'(u) - \dfrac{1}{2}\nu_1(u) \end{cases} \quad . \qquad \textbf{(A2.4)}$$

where $s_1 = z_1^2/2$ and $s_{n-1} = \sum_{i=2}^{n} z_i^2/2 = s_o - s_1$, while $\nu_{n-1}(u)$ is the average $\log(1 + u\,\lambda)$ over the $n-1$ smallest eigenvalues and $\nu_1(u) = \log(1 + \lambda_1 u)$ with $\lambda_1 = \alpha\,\tilde{\lambda}$ (eq. (6)).

The general formula above is not fully determined because $\nu_n(u)$ is an average, which varies as n eigenvalues vary randomly. To overcome this, we simply take a limit when $n, p \to \infty$, in which case (i) the distribution of λ_i converges to a non-random limit (the M-P law approximately) and (ii) the expectations $\nu_n(u)$ and$\nu_{n-1}(u)$both converge to the Shannon transform of this M-P law as $n \to \infty$:

$$\nu_n(u) \underset{n,p\,\to\,\infty}{\to} E(\log(1 + u\,\lambda)) = \nu_{\mathbf{M}}(u) = \int_0^{\infty} \tilde{\rho}_{\mathbf{H}.\mathbf{H}^*}(\lambda)\,\log(1 + u\,\lambda)\,d\lambda \approx \nu_{\mathbf{H}.\mathbf{H}^*}(u) \quad , \qquad \textbf{(A2.5)}$$

in the context of the M-P law approximation, $\nu_{\mathbf{M}}$ is approximately given by eq. (A1.3), with effective ratio index $\beta_e = p_e/n$ and effective scale $\zeta_e = \tilde{\lambda}/\beta_e$ as given by eq. (4).

Moments of the DFE: Based on the limit obtained for $\nu_n(u)$ (eq. (A2.5)), we obtain a non-random limit for the CGF via eq. (A2.4). Taking the derivatives of $\kappa_s(u)$ with respect to u taken at $u = 0$ yields the cumulants of the distribution. In particular, defining $(n - 1 + \alpha)\,\tilde{\lambda}/2 = \bar{s}$, $\theta = n/(n - 1 + \alpha) - 1$, $\epsilon_o = s_o/\bar{s}$ and $\epsilon_1 = s_1/\bar{s}$, we obtain, for the mean and variance of the DFE:

$$\begin{cases} E(s) = \kappa_s{}'(0) = -\bar{s} = -\,\zeta_{\mathbf{W}}\,p/2 \\ V(s) = \kappa_s''(0) = \dfrac{2\,\bar{s}^2}{n}\,\dfrac{(n-1)(1+\theta)^2 + p_e(1 + (n-1)\theta^2 - 2\,n\,\theta\,\epsilon_1 + 2(1+\theta)\epsilon_o)}{p_e} \end{cases} , \quad \textbf{(A2.6)}$$

We obtain a simpler expression below phase transition ($\alpha = 1$ so that $\theta = 0$ and $\bar{s} = n\,\tilde{\lambda}/2$), especially when taking the leading order in $n \gg 1$:

$$V(s) \underset{\alpha \to 1}{\to} \frac{2\,\bar{s}^2}{n}\left(\frac{1}{\beta_e} + (1 + 2\,\epsilon_o)\right) + o\left(\frac{1}{n}\right) \qquad \cdot \qquad \textbf{(A2.7)}$$

It can be checked that eq. (A2.7) yields the exact expression from the purely isotropic FGM (see Martin and Lenormand 2006) whenever $\beta_e \gg 1$.

Isotropic approximation: The isotropic approximation that we use in the main text consists in equating the $n - 1$ lowest eigenvalues $\lambda_{i>1}$ to a constant $\tilde{\lambda}$, the mean of the M-P law, which amounts to setting $p_e \to \infty$. The expressions above do not make such an assumption; they only rely on the convergence to the M-P law for the spectral distribution of $\mathbf{M}$ (eq.(A2.5)), and on ignoring any potential correlation between z_i^2 and λ_i (eqs. (A2.3) and (A2.4)). However, to characterize the DFE more explicitly (stochastic representation or pdf), it proves critical to further rely on the isotropic approximation. This approximation proves accurate even though the actual system is clearly anisotropic, as illustrated on **Figure 5.a**, where the spectrum of $\mathbf{M}$ is quite spread. A tentative explanation for this robustness can be proposed, below phase transition (when $\alpha = 1$). Beyond phase transition ($\alpha > 1$), the problem boils down to whether the isotropic approximation is accurate in the eigenspace associated with the $n - 1$ lower eigenvalues, so it is an equivalent issue.

Even when $\alpha = 1$, the actual model is of course never isotropic. The ratio index β must be finite, and because of metabolic correlations in $\mathbf{W}$, the equivalent ratio β_e can be substantially smaller than β (eq. (4)). **Figures 3 and 4** confirm that even with relatively large β, the spectral distribution of $\mathbf{M}$ shows substantial variance, in a manner captured by the M-P law approximation. More precisely, the coefficient of variation of the eigenvalues of $\mathbf{M}$ is approximately $CV(\lambda_i) = 1/\sqrt{\beta_e}$. Therefore, the eigenvalues λ_i are not equal and phenotypic directions are not equivalent. However, this anisotropy remains mild, and proves to have approximately no influence on the DFE, as long as p_e is large enough. This can be understood by looking at the CGF and its approximate expression in eq. (A2.3).

The pdf of a distribution can be obtained as an inverse Fourier transform of the characteristic function of this distribution. In our context, this characteristic function is given by $\psi(t) = e^{\kappa_s(\mathbf{i}\,t)}$, where $\kappa_s(.)$ is the CGF given in eq. (A2.2) when $\alpha = 1$ and $\mathbf{i}$ is the unit complex number ($\mathbf{i}^2 = -1$). Therefore, to find a suitable approximation for the pdf of s one must approximate $e^{\kappa_s(\mathbf{i}\,t)}$. When we can ignore correlations between z_i^2 and λ_i or at the optimum, $\kappa_s(.)$ is approximately given by eq. (A2.3). Part of the anisotropy then vanishes already: only the distance to the optimum s_o has an impact, not its direction. However, to obtain the isotropic approximation exactly (n traits all equivalent) still requires to seek an approximation for $\nu_{\mathbf{H}.\mathbf{H}^*}(u)$, more precisely for $\psi(t) = e^{-n/2\,\nu_{\mathbf{H}.\mathbf{H}^*}(\mathbf{i}\,t)}$at the optimum. The corresponding expression in the isotropic model (all $\lambda_i = \tilde{\lambda}$) is simply $\psi_{iso}(t) = \left(1 + \tilde{\lambda}\,\mathbf{i}t\right)^{-n/2}$ which is the characteristic of a negative gamma distribution $s \sim -\Gamma(n/2, \tilde{\lambda})$. The characteristic function, at the optimum ($s_o \to 0$), is equal to this isotropic approximation, to leading order in $\zeta_e = \tilde{\lambda}/\beta_e$. Indeed, recalling that $\tilde{\lambda} = 2\bar{s}/n$ and $\beta_e = p_e/n$, the ratio between the exact and approximate characteristic functions satisfies

$$\frac{\psi(t)}{\psi_{iso}(t)} \approx \left(1 - \frac{t^2\,\bar{s}^2}{p_e}\right) + o\left(\frac{\bar{s}^2}{p_e}\right) \quad , \qquad \textbf{(A2.8)}$$

The relative error in equating $\varphi(t) \approx \varphi_{iso}(t)$ is thus small under fairly mild conditions. Even when $\beta_e = p_e/n$ is not very large, so that variation across λ_i is substantial, it suffices that p_e be large enough and that mutation effects be mild enough ($\bar{s}^2/p_e \ll 1$) for the isotropic approximation to perform satisfyingly. This accuracy of the isotropic approximation is illustrated in **Figure S1** where $\nu_{\mathbf{H}.\mathbf{H}^*}(u)$ is compared to its equivalent in the isotropic approximation $\nu_{iso}(u) = \log(1 + u\,\tilde{\lambda})$.

This whole argument is merely intuitive as it relies on approximate results, but it does give an intuition on why, even when the spectral distribution of $\mathbf{M}$ is fairly spread (e.g. **Figure 5**.**a**), eqs. (8-11) prove accurate.

References:

Jaschke, S., C. Kluppelberg, and A. Lindner. 2004. Asymptotic behavior of tails and quantiles of quadratic forms of Gaussian vectors. J. Multivar. Anal. 88:252-273.

Martin, G. and T. Lenormand. 2006. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. Evolution 60:893-907.

Mathai, A. M. and S. B. Provost. 1992. Quadratic forms in random variables. Marcel Dekker, New York.

Wolfram Research, I. 2012. Mathematica Edition: Version 9.0. Wolfram Research, Inc., Champaign, Illinois.

File S3

Supplementary Notebook

Available for download as a .cdf file at
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.160325/-/DC1