FISHER’S MODEL AND THE GENOMICS OF ADAPTATION: RESTRICTED PLEIOTROPY, HETEROGENOUS MUTATION, AND PARALLEL EVOLUTION

Luis-Miguel Chevin,1,2 Guillaume Martin,3 and Thomas Lenormand4
1Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire, SL5 7PY, United Kingdom
2E-mail: I.chevin@imperial.ac.uk
3Institut des Sciences de l’Evolution, UMR 5554, Université Montpellier II, 34095 Montpellier Cedex 5, France
4Centre d’Ecologie Fonctionnelle et Evolutive, UMR CNRS 5175, 34293 Montpellier, France

Received February 25, 2010
Accepted May 31, 2010

Genetic theories of adaptation generally overlook the genes in which beneficial substitutions occur, and the likely variation in their mutational effects. We investigate the consequences of heterogeneous mutational effects among loci on the genetics of adaptation. We use a generalization of Fisher’s geometrical model, which assumes multivariate Gaussian stabilizing selection on multiple characters. In our model, mutation has a distinct variance–covariance matrix of phenotypic effects for each locus. Consequently, the distribution of selection coefficients $s$ varies across loci. We assume each locus can only affect a limited number of independent linear combinations of phenotypic traits (restricted pleiotropy), which differ among loci, an effect we term “orientation heterogeneity.” Restricted pleiotropy can sharply reduce the overall proportion of beneficial mutations. Orientation heterogeneity has little impact on the shape of the genomic distribution, but can substantially increase the probability of parallel evolution (the repeated fixation of beneficial mutations at the same gene in independent populations), which is highest with low pleiotropy. We also consider variation in the degree of pleiotropy and in the mean $s$ across loci. The latter impacts the genomic distribution of $s$, but has a much milder effect on parallel evolution. We discuss these results in the light of evolution experiments.

KEY WORDS: Adaptation, extreme value theory, genetic constraints, mutation, parallel evolution, pleiotropy, random matrix theory.

It is difficult to tell both when a beneficial mutation will occur and what its effect will be. Because of these two forms of uncertainty, adaptive trajectories are hardly predictable. However, depending on the reservoir of possible beneficial mutations (few mutations of large effects vs. many mutations of weak effects), different regimes may be distinguished, leading to different kinds of unpredictability of adaptation. The dynamics of adaptation may appear either as a steady increase in fitness but with unpredictable genetic architecture (when numerous alleles of small effects are available at many loci), or as a more erratic fitness trajectory but predictably relying on a few alleles of larger effects at a handful of loci (Lenormand et al. 2009). Ultimately these regimes are determined by the mutation rate, but also by how many loci (and which ones) may potentially respond to selection at any given time. If all traits under selection can be changed by mutation at any locus, that is, if all loci exhibit a similar and large degree of pleiotropy, many loci may produce beneficial mutations and respond to selection, so alleles of small effects will tend to accumulate at different loci, resulting in the steady regime (predictable trajectory with unpredictable genetic architecture). In contrast if
mutations at each locus only affect a very limited number of traits, only a few candidate loci may be able to respond to selection, resulting in the erratic regime (unpredictable trajectory but with a more predictable genetic architecture).

At the phenotypic level, there is growing empirical evidence that most genes can affect several traits simultaneously, that is, have pleiotropic mutational effects (Dudley et al. 2005; He and Zhang 2006; Wagner et al. 2008) but the complete array of phenotypic effects available for mutation at each gene is still unclear. Because genes code for proteins with specific interactions and for enzymes that catalyze specific reactions, they should differ in the primary traits that they can affect, although through integration into a network of regulations they may in fine modify almost all macroscopic traits under direct selection.

Regarding adaptive evolution, the idea that genes vary in their adaptive potential is strongly supported by the large body of evidence for parallel genetic evolution, that is, the repeated and independent fixation of mutations at the same gene (or of the same mutation) in different populations in response to similar environmental challenges. This phenomenon has been documented in the wild for insects (Weill et al. 2003; fish (Chen et al. 1997; Cresko et al. 2004; Colosimo et al. 2005), birds (Mundy et al. 2004), and plants (Christin et al. 2007), notably. Although some of these cases are caused by the recurrent use of the same allele from standing variation (Barrett and Schluter 2007), those that imply very distant taxa rather reflect genuine parallel evolution, where beneficial mutations arose several times independently at the same gene. Experimental evolution in the laboratory using replicate populations of microbes initiated by the same clone has also demonstrated the importance of parallel genetic evolution in bacteria (Elena et al. 1996; Crozat et al. 2005; Woods et al. 2006) and viruses (Bull et al. 1997; Rokyta et al. 2005). Repeated use of the same gene more than predicted by chance, as demonstrated statistically by Woods et al. (2006) for instance, indicates that at any given time, all genes are not equivalent for adaptation. Understanding how the predictability and repeatability of evolution at the genetic level depends on the effects of mutations on traits under selection requires new theoretical developments about the genetics of adaptation.

Genetic theories of adaptation study properties of selection coefficients of single mutations to understand the population process of adaptive evolution, when it is limited by mutation (Fisher 1930; Gillespie 1983, 1991; Orr 1998, 2002). So far, these theories mostly treat the genome as a homogeneous reservoir of mutations, without specifying a set of loci at which those mutations occur. In phenotypic landscape models, tracing back to Fisher’s geometrical model of adaptation (Fisher 1930), each trait under selection is an axis in a multidimensional phenotypic space, and the beneficial mutations are those that allow the population to get closer to the unique neighboring optimum. This approach is not rooted in a clear genetic context (How many alleles? Which loci?), but explicitly considers the pleiotropy of mutation. By contrast, genotypic landscape models attribute fitness to multilocus genotypes (Wright 1932) or to one-step mutants (“mutational landscape,” Gillespie 1983, 1991). There, each dimension corresponds to a locus (or nucleotide) and mutations take discrete values that correspond to the alternative alleles. This approach is not rooted in a clear phenotypic context (How many traits? Which ones?), but explicitly considers genetic changes, although homogeneity over the genome is still assumed in these models.

Bridging the gap between purely phenotypic and purely genetic models would require explicitly attributing mutations to specific loci that may affect different traits. Contrary to the existing models, this approach would allow addressing such questions as: Are beneficial mutations getting fixed at different loci? What determines which locus is more likely to be used at a given time, and what is the probability that the same locus is used repeatedly? And how are those processes influenced by the degree of pleiotropy of mutation?

In this article, we introduce explicit loci in Fisher’s geometrical model, to study the consequences of heterogeneity in mutation effects across loci on the genetics of adaptation from de novo mutations. Our first aim is to understand whether earlier results from mutation models that overlook the loci of adaptation still hold when those are included in the analyses. We also address new issues that arise with heterogeneous mutational effects. We show how heterogeneity in mutational phenotypic effects among loci naturally creates differences in adaptive potential among genes, affecting the likelihood of polygenic response to selection and the predictability of adaptive evolution, when mutation is its limiting factor.

Model

**LOCUS-SPECIFIC DISTRIBUTION OF FITNESS EFFECTS \( f_s(s) \)**

We use a phenotypic model in which the selection coefficients of mutations arise from their effects on traits under selection, contrary to mutational landscape models in which fitness effects of mutations are defined a priori (Gillespie 1983; Orr 2002, 2003). We rely on a general multivariate model of mutation and stabilizing selection on multiple traits, allowing for selective and mutational correlations among traits. This model has been used earlier to study the maintenance of quantitative genetic covariance at mutation—recombination—selection equilibrium (Lande 1980), and the distribution of fitness effects of single mutants, but without mutational heterogeneity (Martin and Lenormand 2006a). It is similar to Fisher’s geometrical model of adaptation (Fisher 1930) in that it deals with adaptation of multidimensional phenotypes close to a single adaptive peak in a smooth fitness landscape. It
is however hopefully more applicable because it also takes into account selective and genetic correlations between traits, and is expressed in terms of measurable effects of mutation on fitness (Martin and Lenormand 2006a).

Adaptation of an organism to its environment is assumed to depend on n phenotypic traits. Each genotype is represented by an n-dimensional column vector $z$ that contains its breeding value (phenotypic value averaged over a residual component of variation) for each of these n traits. Fitness is a multivariate Gaussian function of the vector $z$ with optimum set at $\mathbf{0}$ (the null vector) without loss of generality. The fitness of a genotype with breeding value $z$, relative to the fitness of the best possible genotype, is

$$W(z) = \exp\left(-\frac{1}{2}z^T Sz\right), \quad (1)$$

where $T$ denotes transposition, and $S$ is the selection matrix. The diagonal terms in $S$ are the strengths of stabilizing selection on the breeding value for each trait, whereas the nondiagonal terms quantify correlative selection among traits. We will assume that the rank (number of nonzero eigenvalues) of the selection matrix $S$ is $n$, such that there is no fitness ridge in any direction of the phenotypic space. This means that $n$ is the actual number of orthogonally selected traits, rather than an arbitrary number of measurable traits.

Actual phenotypes are not fully determined by their genotypes, and instead exhibit a residual component of variation caused by microenvironmental variation and developmental noise (Lynch and Walsh 1998). Assuming that this residual is (multivariate) normally distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{e}$, that there are no genotype–environment correlations or interactions, and that the fitness function on phenotypic values is also Gaussian with matrix of selection strength $\Omega$, then $S = (\Omega^{-1} + e^{-1})$ (Lande 1979, 1980).

Each mutation changes the breeding value from $z_0$ to $z_0 + dz$, where $z_0$ is the breeding value of the wild-type genotype. Its selection coefficient can be expressed as

$$s = \ln W(z_0 + dz) - \ln W(z_0) = -\frac{1}{2}dz^TS dz - z_0^TS dz. \quad (2a)$$

In a continuous-time model with overlapping generations, equation (2a) is the relevant measure of selection, that is, a difference in Malthusian fitness (Crow and Kimura 1970). In a discrete-time model, equation (2a) is equivalent for small $s$ to the more usual definition of selection coefficients $s = W(z_0 + dz)/W(z_0) - 1$.

In the present article, we need to distinguish the distribution of fitness effects for individual loci from that at the genome level. More specifically, our aim is to consider the consequences of the heterogeneity of mutational phenotypic effects among loci on the distribution of mutation fitness effects at the genome level. Throughout the article, we will denote with a subscript “$\mathbf{L}$” the distribution (density $f_L$ or moments $E_L$, $V_L$) for a single locus $L$, and with a subscript “$\mathbf{g}$” the genomic distribution (density $f_g$ or moments $E_g$, $V_g$).

The distribution of fitness effects $f_L(s)$ at each individual locus $L$ is similar to that derived by Martin and Lenormand (2006a) at the genomic level without mutational heterogeneity (i.e., when $f_L(s) = f_g(s)$ for all $L$). Assuming that the vector of mutational phenotypic effects $dz$ for a given locus $L$ is drawn from a multivariate Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $M_L$, then $f_L(s)$ can be approximated by a displaced negative gamma distribution,

$$f_L(s) = \frac{e^{-s_{\max}L}(-(s_{\max}L - s)_+)^{\beta_L-1}e^{-(s_{\max}L - s)\beta_L}}{\Gamma(\beta_L)}, \quad (2b)$$

where $\Gamma(\cdot)$ is the Euler gamma function. The displacement parameter in this distribution is $s_{\max,L}$, the maximum possible selection coefficient of a mutation at locus $L$. In Martin and Lenormand (2006a) $s_{\max,L} = s_0 = 1/2\Omega_S z_0$, the initial fitness distance to the optimum, such that the best possible mutation allows reaching the optimum, but it need not be so as we will show below. The shape parameters of this distribution is $\beta_L = \beta^+_L (1 + t_L)^2/(1 + 2q_L s_{\max})$, and its scale parameter is $\alpha_L = \alpha^+_L (1 + 2q_L s_{\max})/(1 + t_L)$, with $\beta^+_L$ and $\alpha^+_L$ the shape and scale parameters when the population is at the optimum, and $t_L = s_{\max,L}/E_L(s)$ (with $E_L(s)$ the mean selection coefficient at locus $L$). The term $q_L = \Omega^+_L S z_0\Omega^+_L S [\Omega^+_L S]^T$, (where “$\mathbf{tr}$” denotes the trace of a matrix, the sum of its diagonal elements, or of its eigenvalues) captures the influence of the directionality of $z_0$ with respect to the orientation of the selection matrix $S$, for a given fitness distance to the optimum $s_0$. Martin and Lenormand (2006a) assumed that $q_L = 1$ in their approximations for the moments of $f(s)$ without mutational heterogeneity (see their eqs. [6a] and [6b]). In practice $q_L$ cannot be related to empirically measurable quantities, and therefore represents a noise parameter.

The moments of $f_L(s)$ are entirely determined by $s_{\max,L}$ by the eigenvalues of $\Omega^+_L S$, which summarize the joint effects of mutation and selection on all traits, accounting for correlations at both levels. The mean selection coefficient of mutations at locus $L$ is $E_L(s) = -1/2\mathbf{tr}(\Omega^+_L S)$. A key result of Martin and Lenormand (2006a) is that this mean selection coefficient is unaffected by $s_{\max,L}$ or by the initial maladaptation $s_0$. The variance of selection coefficients is $V_L(s) = V^+_L(s)(1 + 2q_L s_{\max})$, where $V^+_L(s)$ is the variance at the optimum. When the population has the optimum breeding value ($z_0 = \mathbf{0}$), the ratio of $E^+_L(s)$ to $V^+_L(s)$, which is also the shape parameter of $f_L(s)$ from standard properties of the gamma distributions, is $E^+_L(s)/V^+_L(s) = \beta^+_L = n_{e,L}/2$, where $n_{e,L}$ is the effective number of traits (or effective complexity) for locus $L$. By definition, $n_{e,L}$ is the number of traits that would lead to the same shape parameter for $f_L(s)$ at the optimum if selection and mutation were isotropic (i.e., equally strong along all phenotypic directions, and without correlations). Martin and
Lenormand (2006a) showed that $n_{c,L}$ also relates to the eigenvalues $\lambda_L$ of $SM_L$, such that $n_{c,L} = n/(1 + CV_L^2(\lambda_L))$, where $CV_L$ denotes the coefficient of variation for a given locus. Mutational and selective correlations cause variation in the eigenvalues of $SM_L$, so $n_{c,L} < n$ and $n_{c,L}$ decreases with increasing mutational and selective correlations. The definition of $n_{c,L}$ also implies that the variance at the optimum can be written as $V_L^*(s) = 2E_L^2(s)/n_{c,L}$, and that the scale parameter at the optimum is $a_L^* = -2E_L(s)/n_{c,L}$ (because $a_L^* \beta^*_L = -E_L^2(s)$ for a reverse gamma distribution).

HETEROREGENEITY OF MUTATIONAL PHENOTYPIC EFFECTS ACROSS LOCI

In Martin and Lenormand (2006a) as in most landscape models of adaptation, the distribution of mutational effects are defined at the genomic level, without explicitly referring to the loci at which mutations occur. Here, we wish to study the consequences of heterogeneity of mutational effects across loci, so we need to specify the way mutational effects vary among loci.

Random matrices

In our model, covariance matrices $M_L$ of mutational effects on phenotypic traits are of dimension $n \times n$ at each locus, but their rank $m_L$ can be lower than $n$. Hence at each locus, mutations affect all traits under direct selection, but because of mutational correlations they may affect only $m_L$ linear combinations of those traits independently. A natural way of modeling a distribution of covariance matrices with a given rank $m_L$ and a controlled level of correlation is using random matrix theory (Appendix 2 in Martin and Lenormand 2006a). Specifically, we assume that

$$M_L \sim \frac{1}{m_L}W(c_L I_n, m_L),$$

where $W(c_L I_n, m_L)$ is the Wishart distribution (Wishart 1928) with $m_L$ degrees of freedom and covariance matrix $c_L I_n$, with $I_n$ the identity matrix. Similarly we assume that the selection matrix is drawn from

$$S \sim \frac{1}{n}W(d I_n, n),$$

such that its rank (number of traits under orthogonal selection) is $n$. Under these assumptions, it can be shown that for large $n$, the mean selection coefficient of mutations at locus $L$ converges to

$$E_L(s) = -\frac{\text{tr}(SM_L)}{2} \rightarrow -n_{c,L}d/2$$

and the effective number of dimensions converges to

$$n_{c,L} \rightarrow \frac{n}{1 + n(1/m_L + 1/n)} = \frac{n}{2 + n/m_L}$$

from equation (8) in Martin and Lenormand (2006a). With smaller $n$, or if the rank of the selection matrix $S$ is lower than $n$, such that the strength of stabilizing selection varies strongly among traits, randomness of mutation matrices also causes sampling variance around the expectations in equation (3c,d).

Forms of mutational heterogeneity

Several forms of heterogeneity of mutational phenotypic effects can be envisaged, as illustrated in Figure 1A. First, loci could differ in their pleiotropy levels, that is, in the number of traits (or linear combinations of traits) that can change independently by mutation at each locus. We will describe this situation as “pleiotropy heterogeneity.” This is equivalent to considering that each mutation at each locus has its own rank $m_L$, with heterogeneous mutation matrices and pleiotropy $m = 1$. The breeding value $z_0$ of the wild-type genotype is represented by the small open circle, and the optimum by a triangle. Strong mutational correlations that limit pleiotropy (here $m = 1$ while $n = 2$) cause mutations at each locus to be restricted to a subspace (continuous lines) with a local optimum represented by the diamonds, with maximum selection coefficient $s_{\text{max},L} < 0$.

Figure 1. Heterogeneous mutation and adaptive landscape. (A) Three forms of mutational heterogeneity of mutational phenotypic effects are illustrated for mutation matrices with $n = 2$. (B) Multivariate Gaussian selection close to a phenotypic optimum is shown in a phenotypic space with $n = 2$ dimensions, for two loci with heterogeneous mutational effects and pleiotropy $m = 1$. The breeding value $z_0$ of the wild-type genotype is represented by the small open circle, and the optimum by a triangle. Strong mutational correlations that limit pleiotropy (here $m = 1$ while $n = 2$) cause mutations at each locus to be restricted to a subspace (continuous lines) with a local optimum represented by the diamonds, with maximum selection coefficient $s_{\text{max},L} < 0$. The breeding value $z_0$ of the wild-type genotype is represented by the small open circle, and the optimum by a triangle. Strong mutational correlations that limit pleiotropy (here $m = 1$ while $n = 2$) cause mutations at each locus to be restricted to a subspace (continuous lines) with a local optimum represented by the diamonds, with maximum selection coefficient $s_{\text{max},L} < 0$. The strength of stabilizing selection varies strongly among traits, randomness of mutation matrices also causes sampling variance around the expectations in equation (3c,d).
HBENEBUSU MUTATION EFFECTS ACROSS LOCI

cL across loci; indeed, changing cL will affect the mean selection coefficient of mutations (eq. 3c) without changing nL (eq. 3d). Finally, the different locus-specific mutation matrices may have the same degree of pleiotropy (mL = m for all L) and the same mean fitness effect (E(s) = E(s) = f for all L), but still affect different combinations of traits. We describe this last situation as “orientation heterogeneity.”

Under orientation heterogeneity, if m is much lower than n, very few linear combinations of traits are affected by mutation at each locus, and those mutable directions may differ a lot from one locus to another. In contrast when m is larger, the phenotypic effects of mutation are more similar across loci. This can be visualized by plotting a measure of the coefficient of variation of the covariance matrices of mutational phenotypic effects against m

In actual biological systems, overall mutation is probably characterized by a combination of these three forms of heterogeneity (pleiotropy, magnitude, and orientation heterogeneity). However in the present article, we will chiefly focus on the third type (orientation heterogeneity), because it is the simplest manifestation of the fact that mutations at different loci affect distinct phenotypic traits. Moreover it is a minimum form of heterogeneity, which should occur even when other forms are absent. This allows us to derive analytical results that capture important properties of this kind of heterogeneity, and are still qualitatively relevant when combined with pleiotropy or magnitude heterogeneity. These other two forms of heterogeneity are then included in the analysis, to investigate how they amplify the consequences of orientation heterogeneity.

**SIMULATIONS**

To test the predictions of the model and the robustness of our approximations, we simulated mutations at nL loci with mutational heterogeneity. To draw matrices mL from the distribution in equation (3a), mL column vectors vl were drawn from a multivariate Gaussian distribution with mean 0 and covariance matrix cL In, and then mL was obtained as mL = 1/N ∑m l = 1N vlvlT. A similar method was used to draw S at the beginning of the simulation, but drawing n vectors with covariance matrix mL. To model the case in which all loci have the same mean selection coefficient E(s), we scaled mutation matrices by a scalar such that tr(SMl)/2 = -E(s).

The breeding value z0 of the wild-type genotype was then chosen such that the initial maladaptation was s0, and that z0 had no preferred orientation in the phenotypic space. This was done by drawing a multivariate Gaussian distribution with mean 0 and covariance matrix In, and then scaling it by a scalar so that z0Sz0 = s0. Then for each locus, mutations dz were drawn in a multivariate Gaussian distribution with mean 0 and covariance matrix mL, and their selection coefficients were calculated following equation (2a). All simulations and analyses were performed on R (R Development Core Team, 2007).

**Distribution of Fitness Effects of Mutations**

Our first aim is to investigate how heterogeneity in the phenotypic effects of mutations affects the distribution of selection coefficients of mutations. We will describe the effect of orientation heterogeneity first, and then include pleiotropy or magnitude heterogeneity in the analysis.

**RESTRICTED PLEIOTROPY AND MAXIMUM SELECTION COEFFICIENT**

In our model, n orthogonal traits are under selection, but mutations at each locus can only affect nL linear combinations of those traits independently. We call this situation “restricted pleiotropy.” Because of restricted pleiotropy, it is impossible for mutations at any given locus to reach the optimum, because they are “enclosed” to a subspace of lower dimensionality than the total phenotypic space, as illustrated in Figure 1B. The “mutable” phenotypic axes at each locus only allow reaching a local optimum (represented by diamonds in Fig. 1B) where fitness is necessarily lower than at the global optimum. We define smax to as the selection coefficient of the best mutation available at locus L, which allows reaching the corresponding local optimum (Fig. 1B). The situation depicted in Figure 1B, where strong genetic correlations cause all mutations to occur along a single phenotypic direction, might seem extreme. However this particular example was only chosen for an illustrative purpose. In less-extreme situations, restricted pleiotropy still leads to smax ≤ s0 as long as the rank mL of mutation matrices is lower than the number n of traits under selection. In practice, it is quite likely that mutational correlations are such that some traits under selection can always be obtained as linear combination of other traits. A similar constraint was studied by Kirkpatrick and Lofsvold (1992), but focusing on correlations in the standing variation of traits rather than in mutational effects, and without mutational heterogeneity.

Although mutation effects at each locus L are enclosed to a subspace of dimensionality mL < n, this subspace differs among loci, so the overall dimensionality of selection is still n. As a consequence, the local optima differ across loci, as shown in Figure 1B. In the Appendix, we show that the maximum selection coefficient of mutations at any locus can be interpreted as a distance to the optimum along “mutable axes,” in a transformed phenotypic space where selection is isotropic (i.e., uncorrelated and equally strong along all axes). As a consequence, if all loci have the same pleiotropy level m << n, the distribution of the maximum selection coefficient smax of mutations among loci has
a simple form

\[ s_{\text{max}} \sim \frac{z_0^T S^2 z_0}{2 \text{tr}(S)} \chi_m^2, \quad (4a) \]

where \( \chi_m^2 \) is a chi-square-distributed random variable with \( m \) degrees of freedom (Appendix). In particular the genomic mean of the locus-specific maximum selection coefficient for loci with pleiotropy \( m \) is

\[ E_p(s_{\text{max}}) = (m/2)z_0^T S^2 z_0 / \text{tr}(S). \quad (4b) \]

Moreover from equation (A7) in the Appendix, when the rank of the selection matrix \( n \) is large, we may also use the approximation

\[ s_{\text{max}} \approx 2 \frac{s_0}{n} \chi_m^2, \quad (4c) \]

\[ E_p(s_{\text{max}}) \approx 2 s_0 m / n \quad (4d) \]

for all \( m \ll n \). Note that equation (4a–d) is independent of the mean fitness effect of mutations at each locus, so they are valid even if magnitude heterogeneity combines to orientation heterogeneity of mutational phenotypic effects. This is because with unbounded continuous distribution of mutational phenotypic effects along mutable directions, the maximum selection coefficient is only determined by the orientation of the mutation matrix, not by its trace (Fig. 1B).

Figure 2 compares simulated distributions of the maximum selection coefficients \( s_{\text{max}} \) for many loci with the same pleiotropy level \( m \) to the expected distributions in equation (4a,c). Several observations can be made. First, the maximum selection coefficient for every locus is positive, but much lower than the actual fitness-distance to the optimum \( s_0 \), when pleiotropy of mutation is restricted (\( m < n \)). How much smaller than \( s_0 \) the mean \( s_{\text{max}} \) is depends linearly on the level of pleiotropy (rank \( m \) of the mutation matrix) if \( m \ll n \). Second, the shape of the distribution of maximum selection coefficients changes with \( m \), becoming less skewed and with a lower coefficient of variation when \( m \) increases, as expected for a chi-square distribution with \( m \) degrees of freedom. Third, the shape distribution of \( s_{\text{max}} \) is independent of the initial maladaptation \( s_0 \); it only depends on \( m \). Fourth, the approximation of equation (4c) performs well when the rank of mutation matrices is low (\( m = 2 \)), but it becomes less accurate than the exact formula of equation (4a) when \( m \) increases. Moreover the chi-square approximations of equation (4a,c) perform better for lower \( m \) (Appendix). When the rank \( m \) of mutation matrices is not constant (pleiotropy heterogeneity), the distribution of the maximum selection coefficient \( s_{\text{max}} \) across loci can be obtained as a mixture of chi-square distributions with \( m \) degrees of freedom over the distribution of \( m \).

**Figure 2.** Distribution of \( s_{\text{max}} \) under orientation heterogeneity. The distribution of the maximum selection coefficient \( s_{\text{max}} \) of mutations (calculated from eq. A2 in the Appendix) for 3500 loci with the same pleiotropy level \( m \) and same mean fitness effect (histogram) is shown together with the expected chi-square distributions using the exact multiplying constant (eq. 4a, continuous line) or the approximation (eq. 4c, dotted line). The first row corresponds to mild maladaptation \( (s_0 = -E(s)/5) \) whereas the second row corresponds to intermediate maladaptation \( (s_0 = -E(s)) \), and the rank of mutation matrices (pleiotropy) is 2, 5, and 10 from left to right. The mutation matrices were drawn from Wishart distributions with \( m \) degrees of freedom and normalized so as to yield a constant mean selection coefficient \( E(s) = -0.1 \) by imposing that \( \text{tr}(SM_l) = -2E(s) \). The total number of dimensions is \( n = 200 \).

**HETEROGENEITY IN \( f_L(s) \)**

Variation in the phenotypic effects of mutations across loci entails that the locus-specific distributions of fitness effects \( f_L(s) \) also differ from one locus to another. Figure 3 shows examples of such distributions under orientation heterogeneity, for several degrees of pleiotropy \( m \), with or without maladaptation of the wild-type genotype.

The first row in Figure 3 corresponds to perfect adaptation of the wild-type genotype \( (s_0 = 0) \), such that all mutations are deleterious. In this case, under orientation heterogeneity, the distribution of fitness effects at each locus only depends on its effective number of dimensions \( n_{e,L} \) (“Model” section). Because of heterogeneity in \( M_L \) among loci, the eigenvalues of \( S M_L \) and their coefficient of variation (which determines \( n_e \)) change to some extent from one locus to another. However mutation matrices with a given rank (pleiotropy) \( m \) should vary little in their effective number of dimensions. Indeed, for matrices drawn from the distribution in equation (3a), with \( c_L \) and \( m_L \) constant, \( n_e \) converges for large \( n \) to the value in equation (3d). In Figure 3, the continuous line is the negative gamma distribution with shape \( n_e/2 \), with \( n_e \) assumed to follow equation (3d), and mean \( E_L(s) = E(s) \) (constant mean selection coefficient across loci). The simulated distribution for each locus is shown with dots. The simulated distributions are very similar among loci for all \( m \), and fit the prediction very well.
HETEROGENEOUS MUTATION EFFECTS ACROSS LOCI

Figure 3. Locus-specific distributions of fitness effects $f_L(s)$. Simulated and predicted locus-specific distributions of fitness effects $f_L(s)$ are plotted for five loci with pleiotropy $m = 2, 5$ and $10$ and mean fitness effect $E(s) = -0.1$ under no ($s_0 = 0$) or strong ($s_0 = -20E(s)$) maladaptation of the wild-type genotype. For each locus, a mutation matrix $M_L$ was drawn as in Figure 2, and then 10,000 mutations were drawn from a multivariate normal distribution with covariance matrix $M_L$. The points are counts of number of mutations in each corresponding class of values of $s$, normalized to sum up to $1$. For each locus, the line is the density of the displaced gamma approximations from equation (2b), with $s_{\text{max},L}$ calculated from the distance to the optimum in the transformed space (Appendix), using the asymptotic $n_{e,L}$ in equation (3d) and, assuming $q_L = 1$.

Therefore, locus-specific distributions of fitness effects $f_L(s)$ for loci with a given pleiotropy level $m$ can be well approximated by assuming that their effective complexity $n_e$ is constant across loci.

The second row in Figure 3 corresponds to strong maladaptation, when the population is away from the optimum ($s_0 = -20E(s)$). In this case, although the effective number of dimensions $n_e$ does not vary for loci with the same pleiotropy, the orientation of mutation matrices with respect to the direction of the optimum and to the shape of the adaptive landscape causes heterogeneity in distributions of fitness effects. The orientation of a mutation matrix has two main effects on the locus-specific distribution of fitness effects $f_L(s)$. First, it affects the maximum selection coefficient $s_{\text{max},L}$ of mutations at the corresponding locus if pleiotropy is restricted (previous section). And second, it affects how variance of fitness effects and the shape of $f_L(s)$ depend on $s_{\text{max},L}$ through the parameter $q_L$. Although $q_L$ is a noise parameter that cannot be measured, analysis and simulation results (not shown) have shown that $E(q_L) = 1$ and that $\text{CV}(q_L) < 1$. In the second row of Figure 3, the continuous lines have been plotted with a constant $n_e$ and assuming that $q_L = 1$, while the simulations (points) make no assumption for $q_L$. The distributions of fitness effects $f_L(s)$ are well predicted without knowledge of $q_L$ (Fig. 3). Hence, the main effect of heterogeneity in the orientation of mutation matrices $M_L$ on variation of locus-specific distributions of fitness effects $f_L(s)$ is to cause variation in their maximum selection coefficients $s_{\text{max},L}$, thus affecting both the shape and displacement of $f_L(s)$ relative to $0$. Because heterogeneity in $f_L(s)$ is most pronounced on their right tails, the proportions and mean effects of beneficial mutations can vary substantially across loci, which will have consequences for the genetics of adaptation, as described below.

If pleiotropy heterogeneity also occurs, it causes further variation among $f_L(s)$, by combining distributions with distinct effective number of dimensions $n_e$, and variable distributions of $s_{\text{max}}$. This boils down to combining different $f_L(s)$ illustrated in Figure 3 (for different values of $m$). Magnitude heterogeneity has a more subtle effect. For a given pleiotropy level $m$, the shape of $f_L(s)$ depends on $\epsilon_L = s_{\text{max},L}/|E_L(s)|$ (“Model” section). Variation in the mean selection coefficient $E_L(s)$ across loci is thus similar to an increased variation in $s_{\text{max}}$, in terms of shapes of $f_L(s)$ (not shown). Note that these two kinds of heterogeneity cause variation in $f_L(s)$ even when the population is at the optimum.

GENOMIC DISTRIBUTION OF FITNESS EFFECTS $f_L(s)$

The heterogeneity in distributions of mutation fitness effects $f_L(s)$ across loci may affect $f_L(s)$, the overall distribution of selection coefficients at the genome level, conflicting with the results of earlier work in which loci were not modeled explicitly. Martin and Lenormand (2006a) showed that, if the genomic distribution of phenotypic effects is multivariate Gaussian with covariance $M_g$, and with no underlying heterogeneity among loci (i.e., $M_L = M_g$ for all $L$), then $f_L(s)$ is generally well approximated by a (displaced) negative gamma. However here, because each locus-specific distribution of fitness effects $f_L(s)$ is itself a displaced negative gamma with its own shape and displacement parameters, the genomic distribution $f_L(s)$ is a mixture of $n_L$ gammas (where $n_L$ is the number of genes). As a consequence, $f_L(s)$ may differ from the distribution described in Martin and Lenormand (2006a). We wish to know whether the gamma approximation still holds for the genomic distribution of fitness effects $f_L(s)$ in this case, and if so, what determines its shape and scale.

Orientation heterogeneity

Under “orientation heterogeneity” of mutation matrices across loci, the genomic distribution of fitness effects $f_L(s)$ is a mixture of displaced negative gamma distributions with constant mean fitness effect $E_L(s) = 3$ and constant effective complexity $n_{e,L} = n_e$, but variable scaled distance to the optimum $\epsilon_L = s_{\text{max},L}/|E_L(s)|$ (“Model” Section). With low pleiotropy ($m << n$), the meta-parameter $\epsilon_L$ of this mixture is chi-square-distributed with $m$ degrees of freedom, from equation (4a). Because $\bar{F}$ is the same for all loci in this case, each central moment of $f_L(s)$ can be derived by integrating the locus-specific central moments over the distribution
of η. For the first three central moments, this yields

$$E_x(s) = \tau$$  \hspace{1cm} (5a)

$$V_x(s) = 2 \frac{\sigma^2}{\mu_e} (1 + 2E_x(s)).$$  \hspace{1cm} (5b)

$$\mu_{3x}(s) = \frac{2V_x(s)^2}{\tau(1 + E_x(s))} R$$  \hspace{1cm} (5c)

$$E_x(s) = \frac{m}{2} \frac{\sum_{i=1}^{2} s_i^2}{CV_m |S|}.$$  \hspace{1cm} (5d)

Not surprisingly, with constant mean fitness effect across loci, the genomic mean selection coefficient in equation (5a) is the same as that at each locus. The variance of selection coefficients (5b) is similar to that expected with no mutational heterogeneity, but with $E_x(s)$ instead of $E_L(s)$ as the displacement parameter. Note that $q_L$ has been replaced in (5b) by its expectation $E_x(q_L) = 1$. Therefore, the relationship between $V_x(s)$ and $E_x(s)$ at the genome level is more accurate than that between $V_L(s)$ and $E_L(s)$ for each specific locus, if $q_L$ is unknown. This is because the effects of directionality of mutation matrices other than $s_{\text{max}}$ cancel out when averaged over many loci.

For the third central moment (eq. 5c), the first factor on the right-hand side is the value that would be expected if $f_x(s)$ was a displaced negative gamma as predicted without mutational heterogeneity. The second factor $R$ in equation (5c) thus quantifies how much orientation heterogeneity of mutational effects causes the genomic distribution of fitness effects to depart from a gamma distribution, in terms of skewness. The term $R$ can be obtained by integrating $\mu_{3L}(s) = -2\beta_x^2 \alpha_x^3$ over the distribution of $s_{\text{max}}$ given in equation (4a). It has a complicated expression (not shown), but always lies within the interval [1, 1] when $m$ and $E(s)$ are constant across loci. It is largest when $m$ is close to $|\tau/n/(8\xi_n)|$, and when this latter term is very small. Hence under orientation heterogeneity of mutational effects, the excess skewness caused by mutational heterogeneity across loci remains small as long as maladaptation is not very strong and the total number of dimensions is large. If so, we may approximate the genomic distribution of fitness effects $f_x(s)$ by a displaced negative gamma

$$f_x(s) = \frac{e^{-\frac{s_{\text{max}}-s}{\beta_x}} \left( E_x(s_{\text{max}}) - s \right)^{\beta_x-1} \alpha_x^{-\beta_x}}{\Gamma(\beta_x)},$$  \hspace{1cm} (5e)

with scale and shape parameters

$$\alpha_x = \frac{V_x(s)}{E_x(s_{\text{max}}) - \tau} = \frac{2\tau (1 + 2E_x(s))}{\mu_e (1 + E_x(s))}. \hspace{1cm} (5f)$$

$$\beta_x = \frac{(E_x(s_{\text{max}}) - \tau)^2}{V_x(s)} = \frac{\tau_e (1 + E_x(s))^2}{2 (1 + 2E_x(s))}. \hspace{1cm} (5g)$$

Figure 4 shows the estimated shape parameter $\beta_x$ (Fig. 4A) and third central moment $\mu_{3x}$ (Fig. 4B) of selection coefficients in simulations of 500 loci (with 100 mutations per locus), together

The second terms in equation (5f–g) show how the scale and shape parameters of $f_x(s)$ away from the optimum can be estimated empirically in experiments in which the fitness effects of spontaneous or induced mutations are measured (Elena et al. 1998; Sanjuan et al. 2004; reviewed in Eyre-Walker and Keightley 2007), and a displaced gamma as in equation (5e) is fitted to their distribution.

Figure 4 shows the estimated shape parameter $\beta_x$ (Fig. 4A) and third central moment $\mu_{3x}$ (Fig. 4B) of selection coefficients in simulations of 500 loci (with 100 mutations per locus), together
with their predictions assuming a displaced gamma with parameters given by equation 5(e–g). The prediction for the shape parameter is very accurate, but performs less well as pleiotropy m and the distance to the optimum s0 increase (Fig. 4A). Perhaps more interestingly, the third central moment is also very well predicted by assuming that \( f_s(s) \) is gamma distributed, even for large maladaptation (\( s_0 = 2 \), Fig. 4B). Hence, with “orientation heterogeneity” of mutation matrices among loci, the genomic distribution of fitness effects \( f_s(s) \) remains approximately a displaced negative gamma as described in Martin and Lenormand (2006a), at least up to the third moment. However there are two main differences with the results in Martin and Lenormand (2006a). First, because of restricted pleiotropy, the displacement parameter of this distribution is not \( s_0 \) but \( E_g(s_{\text{max}}) \) as given in equation (4b). Second, with mutational heterogeneity, the genomic distribution of fitness effects \( f_s(s) \) is not directly related to the covariance matrix of mutational effects on phenotypic traits at the genomic level \( E_p = E_g(M_L) \). The overall distribution of phenotypic effects shows little mutational correlations if the orientation of locus-specific mutation matrices is random in the phenotypic space, as we assume here. Such a lack of mutational correlation would cause a large \( n_e \) if there were no mutational heterogeneity among loci. However with mutational heterogeneity and \( m < < n \) at each locus, each individual locus does exhibit strong mutational correlations and the actual \( f_s(s) \) has a small \( n_e \).

**Other forms of mutational heterogeneity**

If there is also pleiotropy heterogeneity, the genomic distribution of fitness effects \( f_s(s) \) is a mixture of distributions \( f_m(s) \) for each level of pleiotropy \( m \), where each \( f_m(s) \) is similar to the \( f_s(s) \) with constant pleiotropy in equation (5a–f). Combining \( f_m(s) \) with different \( m \) implies mixing negative gammas with different effective complexities \( n_e \) and distributions of maximum selection coefficients. From equation (3d), \( n_e \approx m \) if \( m < < n \) and if the mutation matrices are drawn from the distribution in equation (3a). Assuming that the rank of most mutation matrices remains sufficiently low for the chi-square approximation to hold for \( s_{\text{max}} \), then from equation (5b,d) \( V_m(s) \approx 2\pi^2/m + 2a[\| \right \rangle \), where \( a = \sum_b S^2 \sum_{n_e} \right \langle \| \right \rangle \), for all loci with rank \( m \). Because we assume that the mean fitness effects of mutations can be then approximated as \( V_m(s) \approx 2\pi^2 E_g(1/m) + 2a[\| \right \rangle \). Hence with variable pleiotropy (but constant mean fitness effects) of mutations across loci, the genomic variance of fitness effects of mutations increases with the harmonic mean pleiotropy. That is, loci with low pleiotropy weigh more in the genomic variance than loci with high pleiotropy.

Variation in pleiotropy may also cause the genomic distribution \( f_s(s) \) to differ more from a gamma than under orientation heterogeneity only. This is measured by the ratio \( R \) of the measured third central moment to its predicted value if \( f_s(s) \) was gamma-distributed (eq. 5c). Figure 4C shows how \( R \) changes with the coefficient of variation of pleiotropy across loci \( V_{g}(m) \), for several values of the mean pleiotropy \( E_g(m) \) and of the distance to the optimum. We drew \( m \) from a binomial distribution displaced by 1 (to avoid \( m = 0 \)), and numerical integration was used for each \( m \). The departure from the gamma distribution increases with \( V_{g}(m) \), whereas \( s_0 \) has little effect. Moreover for a given \( V_{g}(m) \), \( R \) is larger for smaller \( E_g(m) \), even though the variance in pleiotropy \( V_{g}(m) \) is smaller in this case.

With magnitude heterogeneity, the moments of \( f_s(s) \) cannot be obtained by integrating the corresponding locus-specific moments over the distribution of \( s_{\text{max}} \), because the mean selection coefficient \( E_g(s) \) also varies among loci. However they can be obtained from the law of total cumulance. The genomic variance is \( V_{g}(s) = E_g(V_{L}(s)) + V_{L}(E_g(s)) \), and the genomic third central moment is \( \mu_{3g}(s) = E_g(\mu_{L}(s)) + \mu_{3L}(E_g(s)) + 3\text{cov}_g(E_{L}(s), V_{L}(s)) \), where \( \text{cov}_g \) denotes the genomic covariance across loci. Figure 4D shows how the departure of \( f_s(s) \) from a gamma distribution (as measured by the parameter \( R \)) is affected by the coefficient of variation of locus-specific mean selection coefficients \( E_{L}(s) \), for several distances to the optimum \( s_0 \) and pleiotropy \( m \). We used a gamma distribution for \( E_{L}(s) \), with shape \( b \) and scale \( E_g(s)/b \), such that its coefficient of variation is \( 1/\sqrt{b} \). This distribution was chosen for its flexibility, however its specific form is of no particular importance to illustrate the effect of \( V_{g}(E_{L}(s)) \). \( R \) increases with the intensity of magnitude heterogeneity of mutation effects, and this effect is stronger with larger \( m \) and larger \( s_0 \). Overall, the examples in Figure 4C, D show that \( R < 3 \) even under strong pleiotropy or magnitude heterogeneity of mutation effects.

Full distributions of fitness effects have been measured in a handful of species (VSV virus, *E. coli*, *S. cerevisiae*, *D. melanogaster*, see table 2 in Martin and Lenormand (2006a)). In all cases the genomic distribution of fitness effects of random mutations is close to a negative gamma (see Fig. 4 in Martin and Lenormand (2006a)). More precisely, the average \( R \) is 0.74 (\( SE = 0.13 \), range 0.43–1.55), that is, slightly below the expectation for a Gamma distribution (\( R = 1 \), although this discrepancy is likely not significant owing to the small number of studies. These estimates were computed assuming that \( s_0 = 0 \) in all cases, which biases the estimation downward if the actual \( s_0 > 0 \) (see eq. 5c). Moreover, studies of mutation fitness effects tend to miss some highly deleterious mutations, which should also reduce the estimated skewness by culling the left tail of the distribution. Alternatively, \( R < 1 \) may be caused by distributions of fitness effects at the optimum \( (s_0 = 0) \) that slightly differ from a gamma at each locus. In any case, the data available so far do not suggest that \( R \) strongly exceeds one for the genomic distribution of mutation fitness effects, as would be expected under strong
magnitude or pleiotropy heterogeneity among loci. Note that this is not at odds with the observed large variation in nonsynonymous divergence rates (or dN/dS) across genes (or functions) in the genome. Indeed, such a variation is likely to mainly result from orientation heterogeneity: loci (or group thereof) with low $s_{\text{max}}$ exhibiting strong purifying selection (and low dN/dS), whereas loci with high $s_{\text{max}}$ exhibiting more adaptive substitutions (and higher dN/dS). The observation that dN includes a substantial proportion of adaptive substitutions (> 40% in the Drosophila lineage, Eyre-Walker 2006; Welch 2006), which varies among functional categories of genes (e.g., Obbard et al. 2009), is consistent with this view.

Adaptive Potential and Parallel Evolution

Although heterogeneity of mutational phenotypic effects across loci has little impact on the functional form of the genomic distribution of fitness effects $f_{L}(s)$ (displaced gamma) as long as magnitude or pleiotropy heterogeneity are not very strong, it does introduce variability in the right tail of the locus-specific $f_{L}(s)$, as seen in Figure 3. In particular, variation in $s_{\text{max}}$ entails that the proportion and mean selection coefficient of beneficial mutations vary from one locus to another. A minority of all mutations that affect fitness are beneficial, but they deserve particular attention because they are the raw material for adaptive evolution. In the following section, we address how mutational heterogeneity in the phenotypic effects of mutations translates into variation in the adaptive potentials of loci, and increases the likelihood of parallel evolution at the genetic level. We focus on populations that are large enough that only beneficial mutations can fix.

COMPONENTS OF THE ADAPTIVE POTENTIAL

We define the adaptive potential $P_{L}$ of a locus $L$ as the probability that a mutation occurring at this locus is beneficial and goes to fixation. Denoting $\pi(s)$ the fixation probability of a mutation with selection coefficient $s > 0$, the adaptive potential of locus $L$ is

$$P_{L} = \int_{0}^{s_{\text{max},L}} f_{L}(s)\pi(s)ds,$$

which can also be written

$$P_{L} = p_{b,L}p_{f,L},$$

where $p_{b,L} = \int_{0}^{s_{\text{max},L}} f_{L}(s)ds$ is the proportion of beneficial mutations at locus $L$ and $p_{f,L} = \int_{0}^{s_{\text{max},L}} \pi(s)f_{L}(s)ds / \int_{0}^{s_{\text{max},L}} f_{L}(s)ds$ is their mean fixation probability. In the simulations that follow, we calculate the adaptive potential as the product of the proportion of beneficial mutations and their mean fixation probability following (6b). Similarly in our analysis, we first study the behavior of $p_{b}$ and $p_{f}$, before turning to the adaptive potential.

Mean fixation probability of beneficial mutations

When there are few beneficial mutations and $s_{\text{max}} << 1$, $p_{f}$ has a simple expression. In a large population, the fixation probability of a beneficial mutation with selection coefficient $s << 1$ is $\pi(s) \approx 2s$ (Haldane 1927; Fisher 1930). Hence the mean probability of fixation of beneficial mutations at locus $L$ can be approximated as $p_{f,L} \approx 2E_{L}(s \mid s > 0)$, where $E_{L}(s \mid s > 0)$ is the mean selection coefficient of beneficial mutations at locus $L$. When beneficial mutations are rare, Martin and Lenormand (2008) showed that a tail approximation based on extreme value theory could be used for the distribution of beneficial fitness effects in the generalized multivariate version of Fisher’s geometrical model. In particular, if the distance to the optimum is $s_{\text{max},L}$, the mean selection coefficient of beneficial mutations at locus $L$ is $E_{L}(s \mid s > 0) \approx 2s_{\text{max},L}/(2 + m_{L})$. Hence the mean fixation probability of beneficial mutations at locus $L$ is proportional to the maximum possible selection coefficient at this locus,

$$p_{f,L} \approx 4s_{\text{max},L}/(2 + m_{L}).$$

Therefore, with orientation heterogeneity of mutation effects across loci, the distribution of $p_{f}$ across loci is (from eq. 4a)

$$p_{f} \sim 2\frac{\chi_{m}^{2}}{m\nu(2 + m)}\chi_{m}^{2},$$

where $\chi_{m}^{2}$ is a chi-square-distributed random variable with $m$ degrees of freedom. The expression in equation (7b) also holds with magnitude heterogeneity of mutational phenotypic effects. Indeed, the proportional relationship between $p_{f,L}$ and $s_{\text{max},L}$ does not depend on $E_{L}(s)$, as long as the tail approximation is valid for beneficial mutations. The chi-square distribution from equation (7b) is a good approximation when $mn$ is small, although discrepancies appear as $mn$ increases (e.g., for $mn \geq 10^{200}$ on Fig. S3). This occurs for two reasons. First, the assumptions that led to the distribution of $s_{\text{max}}$ in equation (4a) become violated as $mn$ increases. Second, the tail approximation for $p_{f}$ as a function of $s_{\text{max}}$ assumes that $s_{\text{max}} << 1$, which becomes wrong as $m$ increases, as expected from equation (4b).

Proportion of beneficial mutations

The tail approximation used above is less robust for the proportion of beneficial mutations than for their mean fixation probability (Martin and Lenormand 2008). We thus recover $p_{b}$ from the displaced gamma approximation for $f_{L}(s)$. The proportion of beneficial mutations at locus $L$ is then given by the cumulated density function of the gamma distribution,

$$p_{b,L} \approx 1 - \frac{\Gamma(\beta_{L},s_{\text{max},L}/\alpha_{L})}{\Gamma(\beta_{L})}$$

where $\Gamma(\cdot)$ used with two arguments is the incomplete gamma function, and $\alpha_{L}$ and $\beta_{L}$ are given after equation (2b). Figure 3 has
shown that the nuisance parameter \( q_L \) has little effect relative to \( s_{\text{max},L} \) on locus-specific distributions of fitness effects. Therefore in all the following, we will assume that \( q_L = 1 \) in \( a_L \) and \( p_L \) for all loci when computing the predicted \( p_{b,L} \). However the effect of \( q_L \) is included in the simulations, so the validity of this assumption can be assessed. Under orientation heterogeneity, the mean proportion of beneficial mutations decreases when \( m \) increases, for a given distance to the optimum \( s_0 \) (Fig. S3), which illustrates one aspect of the well-known cost of complexity (Fisher 1930; Orr 2000; Welch and Waxman 2003; Martin and Lenormand 2006a). The shape of the distribution of \( p_b \) has a complex form; it remains highly skewed and changes little from \( m = 2 \) to 10 (Fig. S3).

**VARIATION IN ADAPTIVE POTENTIAL AND PARALLEL GENETIC EVOLUTION**

Variation in adaptive potential across loci is particularly important for the genetics of adaptation, because it determines to what extent different genes are equally likely to contribute to adaptive evolution. This bears on the number of genes that can potentially respond to selection at any given time, and on the probability of parallel evolution.

We define parallel genetic evolution as the repeated fixation of beneficial mutations at the same locus in replicate populations starting from the same breeding value \( z_0 \) and adapting to the same environment. The probability that the next beneficial mutation that reaches fixation occurs at locus \( L \) is \( P_L u_L / \sum P_k u_k \), where \( u_L \) is the rate of mutations with phenotypic effects at locus \( L \). The expected amount of parallel genetic evolution \( p_{//} \) is the sum of the probabilities that each locus contributes twice, that is,

\[
p_{//} = \sum \left( \frac{P_L u_L}{\sum P_k u_k} \right)^2 = \frac{E_g \left( P_L u_L^2 \right)}{n_L E_g \left( P_{L,H} \right)^2}, \tag{8a}
\]

where the sums are taken over the \( n_L \) loci. There is no reason a priori to think that mutation rates should correlate with adaptive potentials (and similarly for their squares), so (8a) can be simplified to

\[
p_{//} = \frac{1 + CV_g^2(P)}{n_L} \left( 1 + CV_\mathcal{L}^2(u) \right), \tag{8b}
\]

where \( CV_g(.) \) denotes coefficient of variation across loci in the genome.

If all loci have the same mutation rate and adaptive potential, then they are equivalent and interchangeable with respect to adaptation, and the expected amount of parallel evolution is at its minimum \( \min(p_{//}) = 1/n_L \). Heterogeneities of mutation rates and phenotypic effects decrease the number of genes that can respond to selection at any given time, and hence reduce the probability of a polygenic response to selection. Defining the effective number of genes \( n_L \) as the number of equivalent genes that would yield the same constraint on polygenic response to selection (for a given genomic mutation rate), then from the above \( n_L = 1/p_{//} \).

Here we wish to understand how the heterogeneity of mutation effects across loci contributes to the probability of parallel evolution. Therefore we need to compare the expected amount of parallel evolution with mutational heterogeneity to that when all loci have the same adaptive potential \( (CV_g^2(P) = 0) \). We name the ratio of these two quantities the relative increase in parallel evolution caused by mutational heterogeneity,

\[
r_{//} = 1 + CV_g^2(P). \tag{8c}
\]

To intuit how \( r_{//} \) should vary with \( m \) and \( s_0 \) under orientation heterogeneity, it is useful to first focus on its components \( p_b \) and \( p_f \). A heuristic argument can be devised from the geometrical illustration in Figure 5. The major effect of the distance to the optimum is to change the curvature of the fitness landscape (or more precisely, of the isofitness surface) for mutations available from \( z_0 \): almost flat when very far away (Fig. 5, first row) and strongly curved close to the optimum (second row). This affects heterogeneity in \( p_b \) (proportion of the ellipse in grey in Fig. 5). In the limit of infinite maladaptation, all loci have \( p_b = 1/2 \) (no heterogeneity), regardless of their pleiotropy \( m \) (cf. Fisher 1930, p. 40), so \( p_b \) does not contribute to \( r_{//} \) (Fig. 5, first row). In contrast, when the distance to the optimum is smaller, the curvature of the isofitness surface generates heterogeneity in \( p_b \) among loci (Fig. 5, second row). This effect of the curvature is stronger if loci have larger \( m \), because the curvature then acts in more dimensions (compare the left and right bottom panels in Fig. 5). Therefore overall, heterogeneity in \( p_b \) among loci should increase with increasing \( m \) and decreasing \( s_0 \). We now turn to \( p_f \). As discussed
above, \(p_f\) is closely related to \(s_{\text{max}}\) (illustrated by the dotted lines in Fig. 5). When \(m\) is small (low degree of pleiotropy), differences in orientation among loci generate strong heterogeneity in the best available mutations (\(s_{\text{max}}\)), which causes strong heterogeneity in \(p_f\) (compare arrows in right and left panels). Heterogeneity in \(p_f\) among loci thus decreases with \(m\) at any distance to the optimum. Mathematically, this effect is readily seen from equation (7b), which implies \(\text{CV}^2(p_f) = 2m\). Joining \(p_b\) and \(p_f\), this argument suggests that \(r_{ff}\) should decrease with the distance to the optimum (because variation in \(p_b\) vanishes). Further, \(r_{ff}\) should decrease with \(m\) away from the optimum, where only heterogeneity in \(p_f\) matters. Closer to the optimum, \(r_{ff}\) may increase with \(m\), at least in part of the parameter range, as a consequence of the larger contribution of heterogeneity in \(p_b\).

**Probability of parallel evolution**

The simulated \(r_{ff}\) as a function of pleiotropy \(m\) (under orientation heterogeneity) is plotted in Figure 6 for several values of \(s_0\), the maladaptation of the wild-type genotype. Higher probabilities of parallel evolution occur when the population is closer to the optimum (lower \(s_0\)), as suggested by Figure 5. The probability of parallel evolution is generally higher when pleiotropy is low. When \(m = 1\), \(r_{ff}\) can be about five times as large as without mutational heterogeneity. Overall, \(r_{ff}\) decreases rapidly with \(m\) (note the logarithmic scale on abscissa), although not always monotonically for small maladaptation (e.g., \(s_0 = 0.2\)), consistent with the predictions from the previous paragraph. With large maladaptation (\(s_0 = 2\)), \(r_{ff}\) decreases strongly with \(m\), as expected if most variation in adaptive potentials is caused by variation in \(p_f\). The dashed line in Figure 6 is the expected \(r_{ff}\) caused by variation in \(p_f\) only, assuming equation (7b) is still valid, such that \(r_{ff} = 1 + 2/m = (m + 2)/m\).

The black lines in Figure 6 are approximations assuming that \(p_b, L\) is obtained from \(s_{\text{max}}, L\) using equation (7c), whereas the gray lines further assume chi-square-distributed \(p_f\) following equation (7b). These approximations perform well for large or moderate maladaptation, but are less efficient as \(s_0\) decreases. However they do capture the overall tendencies for the effect of \(m\) and \(s_0\) on the probability of parallel evolution.

Although Figure 6 was obtained for a specific set of the parameters \(s_0\), \(n\), and \(E_g(s)\), its results apply with little change to other parameter values, keeping \(s_0/n(\pi)\) constant and as long as \(m << n\) for all \(m\). Indeed, from equation (6a) and the definition of \(f_L(s)\) (“Model” section), the adaptive potential only depends on the effective number of traits \(n_e = n/(2 + n/m)\) and on the scaled maximum selection coefficient \(e = \text{max} \|\pi\|\). From equation (4a) the distribution of \(s_{\text{max}}\) for a given \(m\) only depends on \(s_0/n\), if \(m << n\). It can then be shown that for a given \(m\) the adaptive potential (and hence the probability of parallel evolution) is mostly determined by \(s_0/n(\pi)\). Figure S4 plots \(r_{ff}\) against the mean adaptive potential of loci \(E_g(P)\) (as a surrogate for the likelihood of adaptive evolution), for similar values of \(s_0/n(\pi)\) but changing \(n\) and the range of \(s_0\) across panels. Both \(r_{ff}\) and \(E_g(P)\) were obtained by numerical integration of equation (6a) over the distribution of \(s_{\text{max}}\) from equation (4a). Black lines connect points obtained with a given value of \(s_0/n(\pi)\). The patterns obtained with different values of \(n\) can barely be distinguished for large \(n\) (compare \(n = 200\) and \(n = 500\) if \(s_0/n(\pi)\) does not change). Under orientation heterogeneity only, \(r_{ff}\) does not exceed 5 in the region with moderately small adaptive potentials (e.g., \(E_g(P) > 10^{-4}\)), where adaptation is likely to occur.

**Pleiotropy and magnitude heterogeneity**

Figure 7 shows that the probability of parallel evolution may be further increased by magnitude or pleiotropy heterogeneity. However with moderate coefficients of variation of \(m\) or \(E_g(s)\) across loci (<0.4), \(r_{ff}\) does not exceed 10 for values of \(s_0/n(\pi)\) where adaptation is not negligible (\(E_g(P) > 10^{-4}\)). Moreover for \(E_g(P) > 0.005\), that is, for distances to the optimum such that the probability of adaptation is higher than 0.5% at each locus, \(r_{ff}\) is mostly determined by orientation heterogeneity.
HETEROGENEOUS MUTATION EFFECTS ACROSS LOCI

Figure 7. Parallel evolution under pleiotropy or magnitude heterogeneity. The relative amount of parallel evolution caused by phenotypic mutational heterogeneity, $r_{ij}$, is plotted against the mean adaptive potential of loci $E_g(P)$ for two values of the mean pleiotropy $E_g(m)$ (4 or 10), in the presence of variation in $m_L$ (A) or in $E_u(s)$ (B) across loci. The same range [0.004,0.4] of values of $SS/(nE_g(s))$ was covered in all case. We computed $r_{ij}$ numerically from the definition of $p_t$ in equation (6a), and using the same type of distributions of $m$ (or $E_u(s)$) as in Figure 4C–D. For a given $E_g(m)$, pleiotropy or magnitude heterogeneity increases with the darkness of lines, with the light gray line illustrating orientation heterogeneity only ($CV_g(m) = 0$ and $CV_g(E_u(s)) = 0$). The dark gray/black lines illustrate, respectively: $CV_g(m) = 0.31/0.43$ for $E_g(m) = 4$ and $CV_g(m) = 0.21/0.30$ for $E_g(m) = 10$ in (A); $CV_g(E_u(s)) = 0.25/0.4$ in (B). Other parameters are $n = 200$ and $E_u(s) = -0.1$.

of mutation matrices, because lines with different degrees of pleiotropy or magnitude heterogeneity but same $E_g(m)$ and $E_u(s)$ overlap.

Discussion

When loci in the genome differ in the way they affect traits under selection, they are not equally likely to contribute to adaptive evolution at any given time. We have investigated the influence of heterogeneity in mutation effects across loci on the genetics of adaptation, using a phenotypic model of mutation on multiple correlated characters under stabilizing selection. We mainly focused on “orientation heterogeneity,” whereby all loci have the same pleiotropy level, but differ in the combinations of traits that they can affect by mutation. With strong correlations of mutational effects on phenotypic traits, pleiotropy is restricted: each locus can only affect a limited number of linear combinations of traits under selection. Mutations at each locus then do not allow reaching the global phenotypic optimum, but only a local optimum that differs among loci. The maximum possible selection coefficient of mutations $s_{\text{max}}$, thus varies across loci. The distribution of $s_{\text{max}}$ for loci with same pleiotropy $m$ is shown to be proportional to a chi-square with $m$ degrees of freedom. When maladaptation of the wild-type genotype is not very large, the genomic distribution of fitness effects $f_g(s)$ is similar to that without mutational heterogeneity, that is, it can be approximated by displaced negative gamma as in Martin and Lenormand (2006a). However restricted pleiotropy entails that the shape of $f_g(s)$ depends on the genomic mean of locus-specific maximum selection coefficient $E_g(s_{\text{max}})$ rather than on $s_{\text{fit}}$, the actual fitness distance to the optimum phenotype. Moreover variation in $s_{\text{max}}$ implies that loci have different probabilities to be “used” during adaptation (which we term their adaptive potentials). This variation is highest for loci with low pleiotropy, which therefore exhibit a higher probability of parallel genetic evolution. Heterogeneity in pleiotropy or in mean fitness effects across loci may further increase the variability of adaptive potentials and the probability of parallel evolution, but mostly in situations in which adaptation is rare.

Link to earlier theoretical results

A situation similar to what we term restricted pleiotropy has been studied by Welch and Waxman (2003), who investigated the influence of modularity on the rate of adaptation and the cost of complexity. By modularity they meant that the $n$ original traits of the phenotypic space could be divided into $n/m$ groups of $m$ traits independently affected by mutation. Hence in their model, loci also have pleiotropy $m < n$, but this does not arise from mutational correlations, and mutation matrices do not rotate freely from one locus to another (see Fig. 1A). Their model is thus similar to a discretized version of ours, where the orientation of mutation matrices in the phenotypic space would still be random, but distributed in a limited number ($n/m$) of discrete classes (or modules), with several loci per module. Note however that if mutational correlations were allowed to vary across loci, the orientation of mutation matrices would change within a module even in this case. They showed that modularity did not necessarily increase the rate of adaptation, sometimes even decreasing it, and thus could not help reduce the cost of complexity, contrary to what had been suggested verbally. Our result that the maximum selection coefficient of mutations decreases with decreasing pleiotropy, resulting in lower mean $p_t$, is in qualitative agreement with their findings.

Orr (2005) investigated the probability of parallel evolution in the context of Gillespie (1983)’s mutational landscape model. He showed that the probability that the same exact mutation fixes twice in replicate populations starting from the same genotype is $2/k$, where the wild-type sequence is the $k$th best possible genotype. Recently, Unckless and Orr (2009) extended this analysis to several substitution “steps” of an adaptive trajectory, assuming a fitness landscape with no epistasis, such that the selection coefficients of all alleles remain unchanged throughout adaptive evolution (an assumption that is not valid here as stabilizing selection generates strong epistasis). The results of these two papers apply to parallel molecular evolution, where the same substitution occurs in several populations starting from the same genotype. Here we focused on parallel genetic evolution, where two populations substitute a different mutation at the same gene, so our results are complementary to those in Orr (2005) and
Unckless and Orr (2009). Parallel genetic evolution as studied here arises from functional properties of loci (heterogeneity in their mutational phenotypic effects), whereas parallel molecular evolution only results from the wild-type sequence being one of the best possible genotypes. In the most careful statistical analysis to date of parallel evolution in experimental evolution data, parallel genetic evolution occurred more than expected by chance, but parallel molecular evolution did not (Woods et al. 2006). Empirical evidence (e.g., Mundy et al. 2004) also suggests that recurrent use of the same gene may be more common in the wild than fixation of the same mutation.

Wright and Rausher (2010) recently investigated which genes are more likely to substitute beneficial mutations in the context of metabolic control theory (Kacser and Burns 1973). They studied a linear metabolic pathway leading to the production of a molecule whose rate of production (or flux) is under stabilizing selection. They showed that enzymes that catalyze reactions upstream in the metabolic pathway evolve higher control over the metabolic flux, and substitute more beneficial mutations (and with larger beneficial effects) during adaptation to a new optimum. The results of this “bottom-up” approach apply to the situation in which an organism can only increase its fitness by adjusting the flux of one particular metabolite. In contrast, our “top-down” approach focuses on multidimensional adaptation requiring the joint evolution of multiple traits under selection. It is less specific and should thus apply to a broader range of biological contexts (as proven by investigations of mutation fitness effects from viruses to animals in Martin and Lenormand (2006a)). Note that our results and those of Wright and Rausher (2010) are not necessarily mutually exclusive, but their formal comparison is difficult because their output quantities are quite different.

Our prediction that the probability of parallel evolution is higher at lower maladaptation is, to our knowledge, new. It seems at first counter-intuitive: it could be expected that stronger selection would lead to more marked differences in selective effects of mutations among genes. However under Gaussian stabilizing selection, the mean curvature of the isofitness surface is larger when closer to the optimum, for a given distribution of mutational phenotypic effects (Fig. 5). This accentuates the fitness consequences of orientation heterogeneity in phenotypic effects of mutations across genes, and hence the probability of parallel evolution.

**Empirical implications**

Our results have several bearings on the genetics of adaptation. First, restricted pleiotropy entails that the proportion of beneficial mutations may not increase a lot with maladaptation of the wild-type genotype. In Martin and Lenormand (2006a)’s model, the maximum selection coefficient of mutations is equal to the initial maladaptation $s_0$, and the proportion of beneficial mutations can become large when maladaptation is strong relative to the mean fitness effect of mutations $E_g(s)$. This prediction seems to disagree with the observation that beneficial mutations generally are quite rare, and that their proportion may not increase a lot even under substantial stress (Wang et al. 2009). The reason for this pattern may be that mutations at each locus do not have access to the global optimum because of restricted pleiotropy, such that the best genotype available at each locus is actually not much better than the wild-type. Second, we predict that the distribution across loci of the maximum selection coefficient available at each locus should be determined by their pleiotropy, but not by their mean fitness effect. Third, the probability of parallel evolution should be related to the heterogeneity in mutational effects across loci, and to their pleiotropy level. It should be possible to use experimental evolution with microbes to experimentally test these hypotheses, and to estimate some of the parameters of the model.

The first type of approach applies to the genomic distribution of fitness effects $f_g(s)$. In a survey of the experimental literature, Martin and Lenormand (2006b) showed that the mean fitness effect of mutations remains constant under environmental change, whereas the variance in fitness effects increases with the initial maladaptation $s_0$, as predicted by Fisher’s geometrical model. Recently, Wang et al. (2009) also found support (although not significantly) for a pattern of increasing variance of selection coefficients with initial maladaptation. Here we predict that the genomic variance of fitness effects of mutations $V_g(s)$ increases in proportion not to maladaptation $s_0$, but to the average locus-specific maximum selection coefficient $E_g(s_{\text{max}})$ (eq. 5b, recalling that $E_g(s) = E_g(s_{\text{max}}) / |E_g(s)|$ with $|E_g(s)|$ a constant). Interestingly from equation (4d), $E_g(s_{\text{max}})(2s_0)$ is a rough estimate of $m/n$, the ratio of the number of traits affected by mutation at each locus (pleiotropy) to the actual number of traits under selection, when this ratio is small (typically for $m/n < 0.2$). This ratio is a major determinant of the effective number of dimensions $n_e$ in Fisher’s generalized geometrical model, because $n_e = m(2 + n/m)$ (eq. 3d). Martin and Lenormand (2006a) estimated effective complexities $n_e$ for several organisms based on mutagenesis data, and found generally low values (lower than 3). This could be due either to low dimensionality of selection (small $n$), or to low pleiotropy of mutation while selection is highly dimensional (small $m/n$). These two hypothesis can be tested empirically with the framework we developed here, because only the second one implies that $E_g(s_{\text{max}})$ is smaller than $s_0$. The experiment would involve measuring fitness effects of single mutations (as in e.g., Elena et al. (1998); Sanjuan et al. (2004)) in several environments causing increasing maladaptation of the wild-type genotype. By fitting a displaced gamma as is equation (5e) to the empirical distribution of fitness effects, $E_g(s)$ and $E_g(s_{\text{max}})$ can be estimated in each environment from the shape parameter in equation (5g), after checking that $E_g(s)$ remains constant across environments, as predicted by the
model and validated by several experimental studies (Martin and Lenormand 2006b). The difference between the Malthusian fitness of the wild-type genotype in each environment and that in the reference environment is a measure of $s_0$. The slope of the linear regression of $E_s(s_{\text{max}})$ on $s_0$ across environments, if small, can then be used as an estimate of $\ln(n)$ (from eq. 4d). Such an experiment could thus help elucidate the reasons for the observed low effective dimensionality of selection across taxa (Martin and Lenormand 2006a). Even in cases in which the approximation in equation (4b,d) does not hold (for larger $\ln(n)$) the slope of the regression of $E_s(s_{\text{max}})$ on $s_0$ still quantifies how much restricted pleiotropy limits adaptation, but it cannot be directly related to the influence of $\ln(n)$ on the effective complexity $n_e$.

The second type of experiments would directly address variation in $f_i(s)$ across loci. It should be possible to empirically measure individual $s_{\text{max}}$ at least for a few loci, by using in-vivo targeted mutagenesis, or by transforming cells with genes that have been mutated through error-prone PCR (as in, e.g., Cambray and Mazel (2008)). Such experiments could be combined with replicate adaptations to the same environment starting from clones, and monitoring of parallel genetic evolution, as in Woods et al. (2006), to relate the distribution of $s_{\text{max}}$ to the probability of parallel evolution. Variation in $s_{\text{max}}$ depends on variation in adaptive potentials in a complex way, and our simulations showed that many loci may be required to accurately picture this relationship. However combining such experiments with information on locus-specific $f_i(s)$ may yield a clearer picture of what determines the heterogeneity of adaptive potentials among loci and the genetic predictability of adaptation.

Other factors that may contribute to parallelism

Our results show that mutational heterogeneity multiplies the probability of parallel evolution by 5 to 10, in situations in which adaptation is likely. Although this factor may seem small, it does not relate in a straightforward manner to tests of repeated use of specific genes as carried out by Woods et al. (2006). One of the reasons is that Woods et al. (2006) tested for parallel evolution in genes known to have substituted a beneficial mutation in at least one population along a full adaptive trajectory, which likely inflates their mean adaptive potential. Aside from this, several biological features that were not modeled here may further increase the probability of parallel evolution.

First, variation in mutation rates $u_i$ across loci can cause $p_{ij}$ to be larger, as shown in equation (8b). Second, we considered that there is a continuum of possible phenotypic effects of mutation at each locus, but this approximation is less accurate when mutation at any gene only gives access to a few phenotypes. Indeed, this will cause a sampling effect that will inflate variation in adaptive potential across loci and hence the probability of parallel evolution. However the number of possible mutant phenotypes for a given gene is unclear, and its relationship to the number of possible genetic changes—which can be very large considering point mutations, indels, changes in regulatory regions, etc—is not straightforward. Third, in highly polymorphic asexual populations, variation in adaptive potentials could be increased by clonal interference (Gerrish and Lenski 1998), whereby alternative beneficial mutations compete for ultimate fixation. Because mutants with small fitness advantages are outcompeted by fitter mutants in this case, loci with a larger $s_{\text{max}}$ should be overrepresented among mutations that fix. This may accentuate the importance of differences in $s_{\text{max}}$ across loci, thus increasing $r_{ij}$. Fourth, in a spatially heterogeneous environment, only mutations with large—but spatially antagonistic—fitness effects may be able to overcome gene flow and cause local adaptation, whereas mutations with smaller effects are prevented to fix by gene swamping (Lenormand 2002). This could increase parallelism in a way similar to clonal interference. Finally, some types of environmental changes (embodied in $z_{0,i}$ in our model) may result in higher parallel evolution for a given adaptive landscape, especially when the strength of stabilizing selection is very heterogeneous across traits (matrix $\mathbf{S}$ with rank lower than $n$). Although a full treatment of the effects of these factors is beyond the scope of the present article, they can all be addressed using the model we developed here.

Conclusion

There has been long-lasting interest in the repeatability and predictability of evolution (Travisano et al. 1995; Grant and Grant 2002; Weinreich et al. 2006; Stern and Orgogozo 2009). How much evolution is determined and repeatable, or merely historical and contingent, impinges on the nature of evolutionary biology as a historical or predictive science. If the same environmental challenges lead to similar solutions in several independent instances of adaptation, then evolution may be perhaps predictable, whereas if not, it is idiosyncratic and can only be appreciated in retrospect (Gould 1989; Lenormand et al. 2009). The extent to which evolutionary paths are repeatable also affects the possibility of reverse evolution, when a population is exposed to ancestral selective pressures (Teotonio and Rose 2000, 2001).

At the gene level, the repeatability of evolution manifests itself through parallel genetic evolution. Here we showed that when evolution is limited by mutation, parallel evolution is genetically determined by the heterogeneity of mutation across loci, and that its probability generally increases with decreasing pleiotropy or increasing variance in pleiotropy or mean fitness effect.

However, the selection coefficient of a mutation in isolation may not be relevant to predict its actual dynamics and fate in very polymorphic populations of sexual organisms, where mutations may segregate simultaneously at many loci (Chevin and Hospital 2008). Future work will tell how mutational heterogeneity across
ACKNOWLEDGMENTS

We thank O. Tenaillon and P.-A. Gros for discussions, P. Gladieux for comments on an earlier draft, and two anonymous reviewers for helpful criticisms. This work was supported by a Newton International Fellowship from the Royal Society to L.-M. C. and a starting grant from European Research Council to T.L.

LITERATURE CITED


Associate Editor: J. Hermisson

Appendix

Here we derive the distribution across loci of the maximum selection coefficient $s_{\text{max}}$ of mutations available at each locus.

EQUIVALENT PHENOTYPIC SPACE AND DEFINITION OF $s_{\text{max}}$

We first show how the original phenotypic space can be transformed into an equivalent space (in terms of the fitness effects of mutations) where directions are directly translated into selection coefficients. The axes of this space are linear combinations of the original axes, and are such that selection is isotropic—that is, all directions are equivalent for selection and there are no selective correlations—and mutational effects are independent but vary in magnitude across axes. Formally, this is equivalent to (1) replacing the selection matrix with $I_m$, the identity matrix, and (2) replacing the mutation matrix with a diagonal matrix $A$. Note that this transformation is analogous, but reverse, to that presented in Appendix 1 of Martin and Lenormand (2006a), because their transformation resulted in isotropic mutation and variable selection among axes.

It can be shown that such a transformation always exists, and its derivation is known as the generalized eigenvalue problem (Mathai and Provost 1992). The aim is to find an invertible transformation matrix $F$ such that $z = Fx$, where the column vector $z$ contains the breeding values in the original phenotypic space and $x$ is the equivalent vector in the transformed space, and such that the two conditions defined above are satisfied in the transformed space. In the rest of this section, although we develop an argument for locus-specific mutation effects, the substrings “L” have been removed to simplify notations.

Reformulating equation (2a) in the transformed phenotypic space, the selection coefficient of a mutation of effect $dz$ is

$$s = -x_i^2 F^T S F dz - \frac{1}{2} \frac{dx^T}{dz} F^T S F dx,$$

where $x_0$ is the original breeding value (before mutation) and $dx$ is the mutational effect, both in the transformed phenotypic space. From (A1) the first condition is equivalent to $F^T S F = I_m$. The covariance matrix of phenotypic mutational effects in the transformed space is $E(dx dz^T) = F^{-1} E(dz dz^T) (F^{-1})^T$. Because $M = E(dz dz^T)$ by definition, the second condition is equivalent to $F^{-1} M (F^{-1})^T = A$. Defining a matrix $A$ such that $A A^T = S$, it can be shown that $AMA^T$ and SM are similar (i.e., they share the same eigenvalues). In practice, $A$ can be obtained by Cholesky decomposition of $S$ in statistical software such as Mathematica or R. Then if $O$ is the matrix composed of the eigenvectors of $AMA^T$, $O$ defines an orthonormal basis (such that $O O^T = O^2$), and it is easy to show that $F = A^{-1} O$ satisfies conditions (1) and (2). In the transformed space, the covariance matrix of mutations is $A$. It harbors the eigenvalues of the matrix $S M$ on its diagonal (in decreasing order), and zeros at all nondiagonal terms. The coordinates in the transformed space are obtained from those in the original space by $x = F^{-1} z$ if $SM$ is positive definite (i.e., if $SM$ only has nonzero eigenvalues). If $SM$ is positive semi-definite (i.e., if it has some null eigenvalues), the transformation is still valid, except $F^{-1}$ must be replaced by the left pseudo-inverse (or Moore-Penrose inverse) of the matrix $F$.

In the transformed space, fitness is directly related to the (squared) distance to the optimum, and the initial maladaptation is $s_0 = x_0^T x_0 / 2$. If $M$ is semi-definite with rank $m < n$, then $SM$ is also of rank $m < n$, and therefore the last $(n - m)$ diagonal elements of $A$ are null. The dimensions from $m + 1$ to $n$ in the transformed space thus correspond to combinations of the original traits that cannot change by mutation, because they have no mutational variance. The best possible mutation is then the one for which the distance to the optimum vanishes in all mutable directions (from 1 to $m$). Its selection coefficient is

$$s_{\text{max}} = \frac{1}{2} \sum_{i=1}^{m} x_{0i}^2 = \frac{u^T u}{2},$$

where the column vector $u$ contains the $m$ first elements of $x_0$. Figure S2 shows that $s_{\text{max}}$ predicted from $x_0$ in the transformed
space perfectly matches the maximum selection coefficient of mutations drawn from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{M}$ when $m$ is small (left panel). With larger $m$, the relationship is still very good, but the maximum selection coefficient in simulations is always lower than the one predicted by (A2) (right panel). This is because the distribution of selection coefficient at each locus is a gamma with a shape parameter that increases with $m$ (Martin and Lenormand 2006a). When $m$ increases, the distribution of selection coefficient changes from very leptokurtic to more platykurtic (Fig. 3), so its maximum becomes more difficult to estimate. The discrepancy between simulations and predictions for larger $m$ thus reflects empirical insufficiency of the simulations in estimating $s_{\text{max}}$, rather than an error in the predicted $s_{\text{max}}$: simulations would match predictions provided enough mutations were drawn. For instance, we checked that the shape of the displaced gamma for $f(x)$ (Fig. 3) is more accurate when using equation (A2) than when using the empirical maximum for $x$ (not shown).

An equivalent expression for $s_{\text{max}}$ can be derived by recalling that $\mathbf{x}_0 = \mathbf{F}^{-1}\mathbf{z}_0 = \mathbf{O}^T\mathbf{A}\mathbf{z}_0$ (because $\mathbf{O}^{-1} = \mathbf{O}^T$). We can then write $\mathbf{u} = \mathbf{P}^T\mathbf{y}$ where $\mathbf{y} = \mathbf{A}\mathbf{z}_0$ is a column vector of length $n$ that does not depend on mutation, and $\mathbf{P}$ is the $n \times m$ matrix whose columns are the first $m$ eigenvectors of $\mathbf{AMA}^T$ (first $m$ columns of $\mathbf{O}$). Then (A2) is equivalent to

$$s_{\text{max}} = \frac{\mathbf{y}^T \mathbf{Q} \mathbf{y}}{2}, \quad (A3)$$

where $\mathbf{Q} = \mathbf{PP}^T$ has dimension $n \times n$.

**DISTRIBUTION OF $s_{\text{max}}$ UNDER “ORIENTATION HETEROGENEITY”**

The distribution of $s_{\text{max}}$ across loci can be derived analytically under orientation heterogeneity of mutation covariance matrices. As equation (A3) shows, the maximum selection coefficient $s_{\text{max},L}$ for each locus $L$ depends of the eigenvectors of $\mathbf{AMA}^T$, where $\mathbf{M}_L$ is drawn from the Wishart distribution in equation (3a) and $\mathbf{A}$ is a constant matrix. Then from the properties of Wishart matrices, and recalling that under orientation heterogeneity $m_L = m$ and $c_L = c$ for all $L$, then

$$\begin{align*}
\mathbf{AMA}^T &\sim \frac{1}{m} \mathbf{W}(c \mathbf{A} \mathbf{A}^T, m), \quad (A4)
\end{align*}$$

It has been shown that the distribution of eigenmatrices (matrices of eigenvectors) of Wishart $\mathbf{W}(m, m)$ matrices converges for large $n$ to the Haar distribution, such that each eigenvector is an independent draw from a multivariate normal distribution with mean $\mathbf{0}$ (the null vector) and covariance matrix $\mathbf{I}_m / \text{tr}() = (1/n)\mathbf{I}_n$ (Silverstein 1990; Bai et al. 2007). Here, we focus on the eigenmatrices of $\mathbf{W}(\mathbf{V}, m)$ Wishart matrices, where $\mathbf{V} \neq \mathbf{I}_m$. By analogy with the result above, we conjecture that each eigenvector can be approximated in this case by an independent draw from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{V} / \text{tr}()$ (where the normalization by $\text{tr}()$ ensures that the eigenvectors are unitary). Replacing $\mathbf{V}$ by its expression in equation (A4), $\mathbf{V} = (c/m)\mathbf{A} \mathbf{A}^T$, and recalling that $\mathbf{AA}^T$ and $\mathbf{S}$ have the same trace (because they are similar matrices), then $\mathbf{V} / \text{tr}() = \mathbf{AA}^T / \text{tr}()$. Although we were unable to find a theorem that demonstrates this statement, our simulations confirmed that it yields the right result for the distribution of $s_{\text{max}}$.

If we replace $\mathbf{y}$ with $\mathbf{A}\mathbf{z}_0$ in equation (A3), the maximum selection coefficient of mutations available at locus $L$ can be written as a quadratic form of the initial phenotype $\mathbf{z}_0$ by a matrix $\mathbf{A}^T \mathbf{Q}_L \mathbf{A}$. Because at each locus $L$, the $m$ columns of matrix $\mathbf{P}_L$ are independent draws from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{AA}^T / \text{tr}()$, then by construction $\mathbf{Q}_L = \mathbf{P}_L \mathbf{P}_L^T$ has Wishart distribution with expectation $\mathbb{E}(\mathbf{Q}) = m\mathbf{AA}^T / \text{tr}()$. Because $\mathbf{A}$ is a constant, $\mathbf{A}^T \mathbf{Q}_L \mathbf{A}$ is then drawn from a Wishart with $m$ degrees of freedom and expectation $\mathbf{A}^T \mathbb{E}(\mathbf{Q}) \mathbf{A} = m\mathbf{S} / \text{tr}()$. By the standard properties of Wishart matrices (e.g., http://en.wikipedia.org/wiki/Wishart_distribution), it follows that

$$s_{\text{max}} \sim \frac{\mathbf{z}_0^T \mathbf{S} \mathbf{z}_0}{2 \text{tr}()} \chi_m^2, \quad (A5)$$

where $\chi_m^2$ is a chi-square distributed random variable with $m$ degrees of freedom. Note that, although we chose the Wishart distribution as a natural model of random covariance matrices, the result that $s_{\text{max}}$ is chi-square-distributed holds generally for any type of random mutation covariance matrices of rank $m$, as long as their eigenvectors are (multivariate) normally distributed.

An even simpler approximation can be derived by recalling that $\mathbf{z}_0^T \mathbf{S} \mathbf{z}_0 = 2s_0$. We can then replace $\mathbf{z}_0^T \mathbf{S} \mathbf{z}_0 / (2 \text{tr}())$ with $s_0 / \text{tr}()$, which after some algebra and assuming that the elements of $\mathbf{z}_0$ are independent, can be approximated by

$$s_{\text{max}} \approx s_0 \frac{\text{tr}()} \mathbb{E}(\chi_m^2) \chi_m^2, \quad (A6)$$

where the $\lambda_s$ are the eigenvalues of the selection matrix $\mathbf{S}$. Interestingly this results is independent of the initial phenotype, and only depends on the initial maladaptation and on properties of the fitness landscape. If the selection matrix is drawn from a Wishart distribution with $m_S$ degrees of freedom (with $m_s \geq n$ such that there are indeed $n$ traits under selection), then for large $n$ the eigenvalues of $\mathbf{S}$ converge to the so-called Marčenko-Pastur distribution (Bai 1999). In particular, $\mathbb{E}(\lambda_s^2) / \mathbb{E}(\lambda_s)^2 = 1 + m_s / n$, and the distribution of maximum selection coefficient across loci can be approximated as

$$s_{\text{max}} \approx \frac{s_0}{n}(1 + m_s / n) \chi_m^2. \quad (A7)$$

Note that formulas (A5) to (A7) are derived assuming that the eigenvectors are independent of each other, which is actually wrong but should be a good approximation as long as
$m << n$. These approximations logically perform less well when $m/n$ increases, because overall the eigenvectors are not independent.

If pleiotropy is variable, the distribution of $s_{\text{max}}$ is a mixture of chi-square distributions with variable degrees of freedom. In contrast, variation of $E_L(s)$ across loci does not influence the distribution of $s_{\text{max}}$, because for a given $m$ the distribution of $s_{\text{max}}$ is independent of $E_L(s)$.

**ADDITIONAL LITERATURE CITED**


---

**Supporting Information**

The following supporting information is available for this article:

**Figure S1.** Influence of the rank of mutation matrices on their phenotypic heterogeneity, under orientation heterogeneity.

**Figure S2.** Accuracy of $s_{\text{max}}$ predicted from the transformed phenotypic space.

**Figure S3.** Properties of beneficial mutations.

**Figure S4.** Scaling effects on adaptive potentials and parallel evolution.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.
Supplementary Figures to:

Chevin, Martin and Lenormand
(Evolution)
Fisher’s model and the genomics of adaptation: restricted pleiotropy, heterogeneous mutation and parallel evolution
Figure S1. Influence of the rank of mutation matrices on their phenotypic heterogeneity, under orientation heterogeneity. The coefficient of variation of mutation covariance matrices, defined as $\text{CV}(M) = \sqrt{\text{tr} \left( \mathbb{E} \left[ (M - \overline{M})' (M - \overline{M}) \right] \right) / \text{tr}(M)}$, is plotted against their rank. For each rank $m$, 2000 matrices were drawn from a Wishart distribution with $m$ degrees of freedom, and scaled so as to yield a constant mean selection coefficient $\bar{s} = 0.1$, by imposing that $\text{tr}(SM) = -2\bar{s}$. The total number of dimensions of the phenotypic space is $n = 50$. 
Figure S2. Accuracy of $s_{\text{max}}$ predicted from the transformed phenotypic space. The maximum selection coefficient out of 5000 mutations with covariance matrix $M_i$ is plotted against the $s_{\text{max}}$ predicted from the vector $x_0$ of the wild phenotype in the transformed space (equation A2), for 5000 mutation matrices. The straight line is the first bisector (abscissa = ordinate). Mutation matrices were drawn from Wishart distributions with $m = 2$ (left panel) or $m = 5$ (right panel) degrees of freedom, and scaled as in Fig. S1. For each matrix, $x_0$ was calculated by performing the transformation explained in the appendix. The total number of dimensions of the phenotypic space is $n = 50$ and the initial fitness distance to the optimum is $s_0 = 0.025$. 
Figure S3. Properties of beneficial mutations. The distributions across loci of the mean fixation probability of beneficial mutations $p_f$ and the proportion of beneficial mutations $p_b$ are shown for 2000 loci with pleiotropy $m$ and mean selection coefficient $\bar{s} = -0.1$, and mutation matrices drawn as in figure S2. For $p_f$, the continuous line is the chi-square approximation from eq. (7b). For $p_b$, the continuous line is obtained by performing random draws from the chi-square distribution in eq. (3a) for $s_{\text{max}}$ and applying the gamma approximation from eq. (7), using $n_e = mn/(2m+n)$ for the shape parameter. Other parameters are $n = 200$ and $s_0 = 0.6$. 
Figure S4. Scaling effects on adaptive potentials and parallel evolution. The amount of parallel evolution caused by orientation heterogeneity in mutational phenotypic effects $r_{//}$ is plotted against the mean adaptive potential of loci $E_g(P)$, for several degrees of pleiotropy $m$ (colors) and total number of orthogonal traits under selection $n$ (panels). The fitness distances to the optimum $s_0$ were chosen such that $s_0/(n|E_g(s)|)$ explored the same range [0.004-0.4] in the three panels. Black lines correspond the same values of $s_0/(n|E_g(s)|)$ across panels, evenly spaced on a logarithmic scale, with thickness increasing with this compound parameter. In all cases $E_g(P)$ and $r_{//}$ were obtained by numerically integrating (6a) (with $n_c = mn/(2m+n)$ and $\tilde{s} = -0.1$) over the distribution of $s_{\text{max}}$ assumed to be correctly given by (4a).