Stability and Response of Polygenic Traits to Stabilizing Selection and Mutation

Harold P. de Vladar[1] and Nick Barton
Institute of Science and Techology-IST Austria, A-3400 Klosterneuburg, Austria

ABSTRACT When polygenic traits are under stabilizing selection, many different combinations of alleles allow close adaptation to the optimum. If alleles have equal effects, all combinations that result in the same deviation from the optimum are equivalent. Furthermore, the genetic variance that is maintained by mutation–selection balance is $2\mu/S$ per locus, where μ is the mutation rate and S the strength of stabilizing selection. In reality, alleles vary in their effects, making the fitness landscape asymmetric and complicating analysis of the equilibria. We show that that the resulting genetic variance depends on the fraction of alleles near fixation, which contribute by $2\mu/S$, and on the total mutational effects of alleles that are at intermediate frequency. The interplay between stabilizing selection and mutation leads to a sharp transition: alleles with effects smaller than a threshold value of $2\sqrt{\mu/S}$ remain polymorphic, whereas those with larger effects are fixed. The genetic load in equilibrium is less than for traits of equal effects, and the fitness equilibria are more similar. We find that if the optimum is displaced, alleles with effects close to the threshold value sweep first, and their rate of increase is bounded by $\sqrt{\mu S}$. Long-term response leads in general to well-adapted traits, unlike the case of equal effects that often end up at a suboptimal fitness peak. However, the particular peaks to which the populations converge are extremely sensitive to the initial states and to the speed of the shift of the optimum trait value.

UNDERSTANDING quantitative genetics in terms of population genetics is crucial for both scientific and practical reasons. However, the development of a consistent theory for long-term evolution has had limited success, because the polygenic basis of quantitative traits makes the prediction of their response to selection immensely intricate, even under the simplest assumptions (*e.g.*, additivity, equal effects of an allele on the trait, and linkage equilibrium) (Barton and Turelli 1989; Keightley and Hill 1990; Turelli and Barton 1994). Most traits seem to be under some form of stabilizing selection, either by the direct action of selection on a trait whose extreme values are unfit or indirectly by compromising individual fitness due to pleiotropic detrimental effects (Keightley and Hill 1990; Mackay 2001, 2010; Hill and Zhang 2012). The joint effects of stabilizing selection and mutation lead to very complicated allele frequency equilibria and evolution, and it is not obvious how much genetic variation they can maintain (Turelli 1984, 1988; Bürger 2000).

An exact analysis in terms of allele frequencies is lacking for polygenic traits with loci of unequal effects. This is desirable, as data from genome-wide association studies (GWAS) yield information about the distribution of single-nucleotide polymorphisms (SNPs) relevant to several traits based on population sequence data (Hindorff *et al.* 2009; Visscher *et al.* 2012). This makes it urgent to understand how variation at the molecular level explains phenotypic variation. How quantitative variation depends on the number of loci and on the distribution of allelic effects is not clear and is the central question of this article.

Barton (1986) showed that when a population is at equilibrium with the trait mean at the optimum and the alleles are very close to fixation, then the genetic variance, ν, is $2n\mu/S$, where n is the number of contributing loci, μ is the per-locus mutation rate, and S is the strength of stabilizing selection. In general, the genetic load L is due to both deviation from the optimum and genetic variance, $L \propto \Delta z^2 + \nu$, where Δz is the deviation of the trait mean from the optimum. The contribution of the genetic variance is much more significant because compared to it, the deviations from the optimum are very small. Moreover, if the trait is not at

doi: 10.1534/genetics.113.159111
Manuscript received November 5, 2013; accepted for publication April 2, 2014; published Early Online April 7, 2014.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.159111/-/DC1.
[1]Corresponding author: IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria. E-mail: hpvladar@ist.ac.at

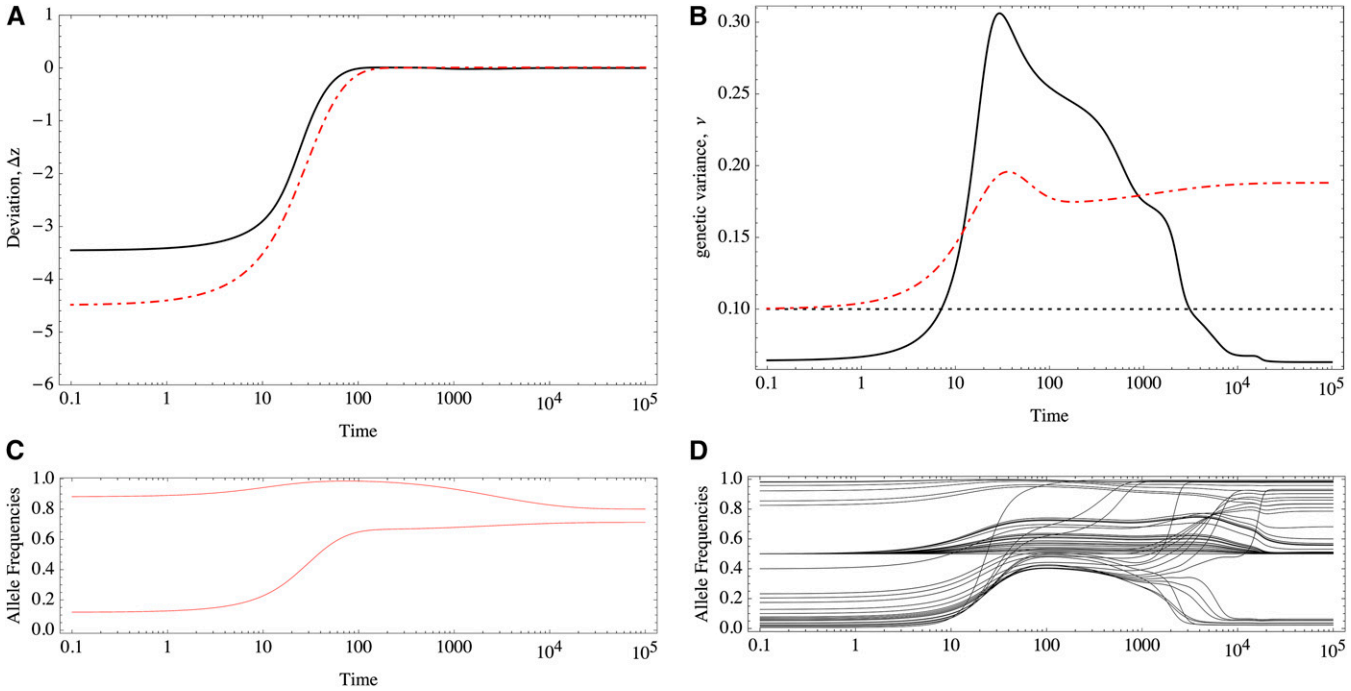

Figure 1 Response to selection of traits determined by 50 loci of equal (dashed-dotted red line) and unequal effects (black solid line). (A) Deviation of the trait mean from the optimum value. (B) Genetic variance. The dotted black line show the HoC variance. (C and D) Allele frequencies under (C) equal and (D) unequal effects. The equal effects have $\gamma = 1/10$, and the unequal effects are distributed as an exponential (mean = 1/10). Mutation rate $\mu = 10^{-4}$, selection intensity $S = 10^{-1}$. The dynamics are numerical solutions to ordinary differential equations (Equation 6 in the text).

the optimum, higher variance can be maintained. These calculations assumed a trait with diallelic loci of equal effects. However, we show that under unequal effects, deviations from the optimum can also maintain less variance.

The response to a shift in the optimum trait value will be radically different under equal and unequal effects. For example, Figure 1 shows the response of two equivalent populations that differ only in their distribution of allelic effects. Note that although the traits match the optimum almost perfectly in both cases (Figure 1A), under equal effects much more variation is maintained than under unequal effects (Figure 1B), which implies a greater mutation load.

We will see that under unequal effects, the equilibria depend on the magnitude of allelic effects. With equal effects, there is a high degree of symmetry in the sense that many allelic combinations match a given optimum value, making it easier to characterize the possible equilibria (Barton 1986). However, this analysis fails under unequal effects because the symmetry is absent. For example, Figure 1, C and D, shows the response of the allele frequencies; in Figure 1, C and D, the alleles have equal and unequal effects, respectively. Initially, the alleles rest at a stable equilibrium that has comparable mean and variance. We see that the response under unequal effects is more heterogeneous in Figure 1D (unequal effects), whereas the alleles respond homogeneously under equal effects (Figure 1C). This difference in the response accounts for the eventual maladaptation of traits with loci of equal effects. Moreover, alleles that have very large effects are at very low frequency and might take substantial time to achieve a higher representation in the population. Thus, anticipating when they will reach intermediate frequencies and make a notable contribution to the genetic variance is difficult. Also, we do not know which alleles contribute preferentially to the response and eventual adaptation of the trait. Under equal effects, all alleles have the same contribution, but the symmetry of the solutions effectively reduces the genetic degrees of freedom, which in turn limits the possible paths to find a global fitness optimum.

The house of cards (HoC) is a mutation–selection balance model that assumes that each allele is new and arises independently from the previous allele from which it mutated, so that the effects of new mutations are uncorrelated from the previous ones. In equilibrium, the variance of allelic effects is larger than the genetic variance and is predicted to be $2n\mu/S$, where n is the number of loci, μ is the per-locus mutation rate, and S is the strength of stabilizing selection (Kingman 1978; Turelli 1984; Bürger 2000, Chap. IV). Exactly this amount of genetic variance is maintained for traits with several diallelic loci of equal effects, when they are adapted to the optimum (Barton 1986). However, numerical experiments such as the one shown in Figure 1B reveal that with unequal effects, the genetic variance is decreased even farther below this bound. It is not immediately clear why this difference between traits with equal and unequal effects occurs.

Although the HoC assumes a continuous production of new alleles with varying effects (Turelli 1984), this model

can be interpreted as a limit of a trait consisting of many loci (Barton 1986), in which case each locus composing a trait z under stabilizing selection evolves according (as is explained in detail below) to

$$\frac{dp}{dt} = -S\gamma p(1-p)(2\Delta z + \gamma(1-2p)) + \mu(1-2p), \qquad (1)$$

where p is the frequency of the "+" allele, γ is its allelic effect, and Δz is the deviation of the trait mean from the optimum. Detailed analyses of this system under equal effects were performed by Barton (1986). At equilibrium, the equation above can have one or three solutions for each locus, given by the cubic polynomial on p that results from equating $dp/dt = 0$. If we plot the equilibrium value of p against $\Delta z/\gamma$, we find that there are two types of curves, depending on the mutation rate, μ (Figure 2A). The first type (Figure 2A, thin curve) occurs at high mutation rates; in this case and at small deviations from the trait optimum, the equilibrium is maintained at intermediate frequencies, maintaining substantial variability. The other type of equilibrium (Figure 2A, thick curve) occurs when mutation rates are low compared to the mutational effects: for well-adapted traits either of the two alleles can be near fixation, each one contributing to the genetic variance by $2\mu/S$, as the HoC predicts.

Under equal effects the equilibrium value of the trait depends only on the number of + and "−" alleles, thus allowing many equivalent genetic combinations; there are many other stable but suboptimal combinations (Barton 1986). The particular state to which the population converges is thus strongly determined by its previous history. All these suboptimal combinations trap the population in local fitness peaks that deviate considerably from the optimum trait value. We can see in Figure 2A (thick line) that if the effect of each locus on the trait is fairly large, then deviations from the trait that are at most equally large as the effect can maintain any + or − alleles at equilibrium. Thus, many of the suboptimal combinations are realizable. Also, if the population is resting at an initial equilibrium and the optimum is shifted (either slowly or abruptly), the allele frequencies respond in a coordinated way. Thus, the trait is resilient to perturbations in the sense that all allele frequencies are always equidistant from the bifurcation point where their stability changes and thus resist large deviations from the optimum. Once the bifurcation point is reached, all loci become unstable at once and suddenly jump to a suboptimal state. Therefore, it is unlikely that populations reach an optimal peak.

In this article we show that if the loci that constitute the trait have different effects, there is a more heterogeneous distribution of equilibria, with no symmetry among peaks. There are still many suboptimal states where the population could get stuck, but we will see that under unequal effects, these suboptimal equilibria are much more similar (and closer) to the optimum. However, the trait is also less resilient to deviations from the optimum, and smaller perturbations

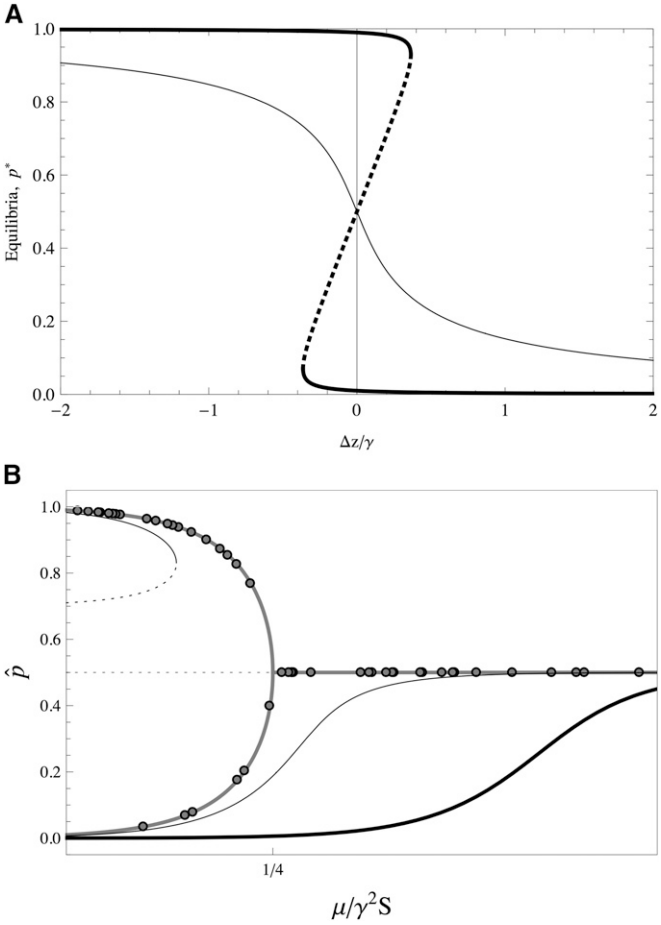

Figure 2 (A) Equilibria of allele frequencies as a function of scaled deviation from the optimum. Thin curve shows equilibria for alleles of small effects ($\gamma^2 < 4\mu/S$); for each value of $\Delta z/\gamma$ there is one stable allele frequency. Thick curve shows equilibria for alleles of large effects, ($\gamma^2 > 4\mu/S$); for small deviations from the optimum, there are two possible equilibria near fixation. The dashed segments are unstable equilibria. (B) Equilibria of allele frequencies as a function of the scaled parameter $m = \mu/\gamma^2 S$. Thick shaded curve shows no deviations from the optimum, $\Delta z = 0$. Thin solid curve shows small deviations from the optimum. Thick solid curve shows large deviations from the optimum. Circles show end point of Figure 1D.

render the configurations unstable. In fact, we see in Figure 2A that alleles of very small effects will make $\Delta z/\gamma$ large, implying that the allelic configurations become unstable. Naturally, the occurrence of small effects is contingent on the distribution of allelic effects, which is unknown in detail; we explore this aspect in this article.

Summarizing, under equal effects precise adaptation to the optimum is harder because the population might get stuck at suboptimal peaks that have large variation and larger mutation load. At equilibrium, selection purges the new mutations and irrespective of their allelic effects, each locus contributes $2\mu/S$ to the genetic variance. In fact, this is an upper bound achieved when the trait is perfectly adapted to the optimum, irrespective of the distribution of genetic effects (as long as these are larger than their contribution to

the genetic variance in equilibrium). However, if the trait mean deviates from the optimum, the genetic variance can differ from that of the HoC (Bürger and Hofbauer 1994) (Figure 1).

A different situation is realized if the allelic effects are smaller than the equilibrium variance, for which the HoC model does not apply. Another classic approximation, which supposes multiple alleles and is often referred as the Gaussian model (GM) (Kimura 1965; Lande 1976), makes the opposite assumption about the allelic effects, *i.e.*, that these are small compared to their contribution to the genetic variance in equilibrium. The GM assumes that there is a continuous production of new alleles that follows a Gaussian distribution of effects at each locus that is centered at the parental genotypic value. Barton (1986) showed that in polygenic diallelic traits under equal effects, changes in the optimum can lead the population toward stable albeit maladapted equilibria that can have much larger variation than that of the HoC and fall into a limit that is better approximated by the GM.

The analyses for polygenic systems with unequal effects that we perform here are more challenging than for equal effects. Our current understanding of unequal effects derives from models that deal with a few loci, from which general results are hard to extrapolate (Turelli 1984; Bürger 2000; Chevin and Hospital 2008; Pavlidis *et al.* 2012). In this article we aim to understand how a trait determined by arbitrarily many loci of unequal effects responds to stabilizing selection and mutation. Putting aside the technical complexities, we regard this problem as fundamental to understanding the bigger picture of the evolution of polygenic traits, namely that of finite populations subject to drift and how these are constrained by pleiotropic effects caused by selection on multiple characters. But first, we need to understand in detail the nature of the equilibria and the response of allele frequencies to factors such as stabilizing selection and mutation. Thus, we address the simplest case of deterministic selection on a single trait.

We start by studying the equilibria and find that there can be multiple loci with high polymorphism, provided that they have small effects. For these alleles of small effect, deviations from the optimum trait value are tolerated without affecting their equilibrium. However, we also find that there is a threshold $\hat{\gamma} = 2\sqrt{\mu/S}$ that objectively defines which alleles are of "small" effect ($\gamma < \hat{\gamma}$) and which ones are of "large" effect ($\gamma > \hat{\gamma}$). The former remain at intermediate frequencies and the latter near fixation most of the time. Alleles of large effect can be sensitive even to small deviations from the optimum. In particular, if the optimum is suddenly shifted, we find that the alleles that respond first are those with effects closer to the threshold value, $\gamma \sim \hat{\gamma}$. In the long term, however, the dynamics are intricate. Different initial equilibrium configurations that are equally well adapted may lead the population to totally different regions of the fitness landscape. However, these different genetic states have very similar phenotypic values.

Model of Stabilizing Selection and Mutation on Additive Traits

We consider the simplest diploid genotype–phenotype map, which assumes an additive trait for diallelic loci, without dominance or epistasis,

$$z = \sum_{i=1}^{n} \gamma_i \Big(X_i + X_i' - 1 \Big), \tag{2}$$

where γ_i is the allelic effect at locus i, n is the number of loci composing the trait, and X and X' are indicators of the $-$ allele ($X, X' = 0$) or of the $+$ allele ($X, X' = 1$). We allow each γ_i to vary across loci. Specific values are drawn from a given distribution (we explore mainly gamma distributed effects), although in every run they are kept constant. Assuming linkage equilibrium, the trait mean and the genetic variance are given by

$$\overline{z} = \sum_{i=1}^{n} \gamma_i (2p_i - 1) \tag{3}$$

$$\nu = 2 \sum_{i=1}^{n} \gamma_i^2 p_i (1 - p_i), \tag{4}$$

where $p_i = \varepsilon[X_i]$, the allele frequency of the $+$ allele, given by the expectation of X_i in the population. (Unless otherwise stated, the expectations are on the population, not on the distribution of effects.)

We assume a Gaussian fitness, $W_z = \exp[-(S/2)(z - z_o)^2]$ so the mean fitness of the population is

$$\overline{W} = \exp\left[-\frac{S}{2} \Delta z^2 - \frac{S}{2} \nu \right], \tag{5}$$

which assumes weak selection. The genetic load is due to both terms: the deviations from the optimum $\Delta z = \overline{z} - z_o$ and the genetic variance ν. The maximum mean fitness is 1, which occurs if an optimal genotype is fixed, with no genetic variance.

In an infinite, random-mating population, the change in allele frequencies is given by the selection–mutation equation

$$\frac{dp_i}{dt} = -S\gamma_i p_i (1 - p_i)(2\Delta z + \gamma_i (1 - 2p_i)) + \mu(1 - 2p_i), \tag{6}$$

for $i = 1, \ldots, n$, and μ, $S \ll 1$ (see, for example, Barton 1986). This equation for the dynamics of allele frequencies assumes linkage equilibrium and weak selection.

To understand the complexities of the fitness landscape and how the trait evolves, we first study the properties of the equilibria. An exact solution of the equilibria of the system defined by Equation 6 for the n loci is possible, but the formulas are complicated, being solutions to coupled cubic equations. Therefore, instead of taking this exact but intricate approach,

we first study the qualitative aspects of the equilibria of Equation 6, by assuming that Δz is given, which uncouples the equations. This gives a solid intuition to understand detailed equilibrium analyses and how much genetic variance can be maintained at the different suboptimal peaks.

Afterward, we explore the dynamics and understand the irregular behaviors by using the intuition derived from the equilibrium analyses. We numerically solve the system for all the allele frequencies at each locus. Since we assume diallelic loci, there are n equations to track. We calculate the genetic variance and trait means from the definitions given by Equations 3 and 4. We typically randomize the initial conditions and the realization of allelic effects, unless otherwise stated.

Allelic Equilibria Have Two Defined Regimes

Equation 6 shows that the equilibrium condition for every locus is given by a set of cubic equations coupled through Δz. When there are many loci, we can assume a particular value for Δz and treat each of the n equations independently. Therefore, for each locus the number of valid roots of the cubic equation depends on four quantities: the deviation from the optimum Δz, the allelic effect γ of the focal locus, the mutation rate μ, and the strength of stabilizing selection S. However, these four variables can be combined into just two scaled parameters, $\delta = \Delta z/\gamma$ and $m = \mu/S\gamma^2$, and the equilibrium solution at each locus is given by the scaled equation

$$p^3 - p^2\left(\frac{3}{2}+\delta\right) + p\left(\frac{1}{2}+\delta+m\right) - \frac{m}{2} = 0. \qquad (7)$$

Figure 2A shows how the equilibrium frequencies depend on the scaled deviation from the optimum, δ. These diagrams also hold for unequal effects, except that the equilibria for each locus are represented by a specific diagram. In *Appendix A* we give the precise expression of the critical points of Equation 7. Figure 2A shows that there may be two types of equilibria: either near fixation of one or the other allele or a single equilibrium at intermediate frequency 1/2. The factor that determines which equilibrium is attained at a given locus is the scaled variable $m = \mu/S\gamma^2$.

Figure 2B shows the equilibrium allele frequencies as a function of m. Consider $\delta = 0$: we see a partitioning of two qualitative regions with stable states that are near fixation ($m < 1/4$, to the left) and intermediate equilibrium ($m > 1/4$, to the right). Note that since m is inversely proportional to γ^2, the smaller the effects are, the more to the right the alleles are represented in Figure 2B, and vice versa. Thus $\hat{\gamma} = 2\sqrt{\mu/S}$ is a threshold that objectively defines alleles of large and small effect: if $\gamma > \hat{\gamma}$, these fall into the category of large effects, and if $\gamma < \hat{\gamma}$, these fall into the category of small effects.

Figure 2B shows how this diagram is modified for $\delta \neq 0$: the bifurcation is shifted, and the intermediate equilibria close to $m \geq 1/4$ are displaced from 1/2. This has two main implications. First, assume a small deviation of $\delta > 0$ ($\delta < 0$); some of the alleles of large effect that would have been close to fixation, at the + (−) state, are forced to sweep to the alternative state. Second, some of the alleles of small effect that would be at the intermediate state, $P = 1/2$, will show reduced (increased) frequencies. Most notably, the alleles that are displaced are those that are close to $m = \hat{m}$, which are those that are close to the threshold $\hat{\gamma} = 2\sqrt{\mu/S}$.

A bigger picture emerges when we consider how the equilibria depend on combinations of both δ and m (see *Appendix A* for the exact calculations), which is shown in Figure 3. On a logarithmic scale, the allelic effects fall on a straight line, with the distribution of effects determining their spread along this line: smaller effects fall toward the right and larger effects fall toward the left of the diagram; different values of δ are represented as parallel lines of slope 1/2.

When effects are large enough that $m < \hat{m} = 1/4$, the alleles can be in a bistable regime: there are two stable points close to fixation and one unstable at intermediate frequency (the thick line in Figure 2A). This is provided that deviations from the optimum are small ($\delta \sim 0$). In this case, the stability is not affected. However (and as we explain below), deviations that are of the order $\geq \delta 1/2$ disturb this equilibrium (see Figure 2B).

The situation is very different for small effects, when $m > 1/4$, since there is only one valid root of the cubic above (the thin line in Figure 2A). When the trait is close to the optimum ($\delta \sim 0$), intermediate frequencies can be maintained, as explained above. Small deviations from the optimum will readjust frequencies slightly, but the stability of the equilibrium is not modified (there is no qualitative change in the stability).

As a consequence, the frequency of alleles of small effect varies smoothly with the deviations from the optimum, whereas those with larger effect experience discontinuous transitions when the magnitude of the deviation approaches half of their respective effects.

The scaling properties indicate that the parameters that really matter are $m = \mu/\gamma^2 S$ and $\delta = \Delta z/\gamma$. Thus, the specific numerical choices of μ and S are not in themselves decisive.

Stability and Variation at Equilibrium

Equilibria under mutation–selection balance

As above, when mutation is present, there might be one or three solutions for each locus, with the stability depending on the particular multidimensional adaptive peak. Consequently, there might be up to 3^n possible equilibria, although only a fraction of these can be stable. Assuming equal effects, it would be enough to count the number of loci that are fixed or intermediate, since the symmetry of the landscape makes it feasible to understand the stability of any of

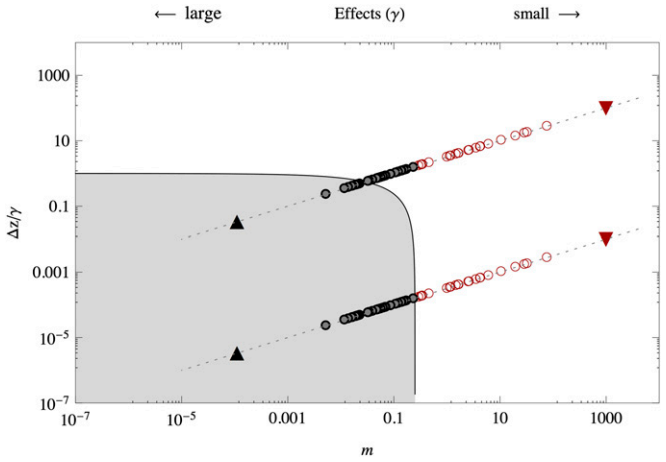

Figure 3 Diagram showing the two regions of qualitatively different equilibria of allele frequencies. For $m < 1/4$ (the gray region) the allele frequencies are near fixation points. When $m > 1/4$, only polymorphisms can be maintained. On a log scale the effects are distributed along parallel lines whose height is determined by $\log(\Delta z) - \log(\gamma)$ and $\log(\mu/S) - 2\log(\gamma)$ and therefore have a slope of 1/2. Effects that fall at the right-hand side of the point $m = 1/4$ can never fall into the bistable regime and correspond to the alleles with the smallest effect. Depicted are (circles) effects distributed, $\gamma \sim \exp(10)$; (upward triangles) traits with equal but large effects ($\gamma = 3$); and (downward triangles) traits with equal but small effects ($\gamma = 0.001$).

these peaks (Barton 1986). With unequal effects, we need to consider each of these configurations separately. These are tractable as long as we assume that $\Delta z = 0$, in which case the equilibria at each locus are given by

$$0 = \left(\mu - S\gamma^2 p(1-p)\right)(1-2p). \tag{8}$$

In this case there are three possible states for each locus, namely

$$p = 1/2 \tag{9}$$

$$p = \frac{1}{2}\left(1 \pm \sqrt{1 - 4\frac{\mu}{S\gamma^2}}\right). \tag{10}$$

In Supporting Information, File S1 we detail the stability analyses that we now summarize. In the absence of mutation, at most one allele can be maintained polymorphic, irrespective of the magnitudes of the allelic effects and of the deviation from the optimum (this result was anticipated by Wright 1935). If mutation rates are small compared to selection, the trait mean is exactly at the optimum, and all alleles are of large effects, then the configurations where all loci are close to fixation will be stable (all eigenvalues are negative). Furthermore, configurations where alleles of large effect are at intermediate frequency will be unstable. We also find that configurations where alleles of small effect are near fixation are unstable.

Why is this? Alleles with large effects increase the genetic variance substantially. We saw that each allele near fixation contributes $2\mu/S$, whereas if it has intermediate frequency, it contributes $\gamma^2/2$ to the load. Since the alleles with large effects fulfill that $\gamma^2/2 > 2\mu/S$, the genetic load would be much larger if the alleles of large effect were maintained polymorphic. A similar argument applies for alleles of small effect. Because $\gamma^2/2 < 2\mu/S$, then the load would be significantly higher if these alleles were near fixation. We can interpret this by thinking that the amount of selection required to fix alleles of small effect would need to be considerably high, to make them fall into a "large effect" class.

Distribution of allelic equilibria

We saw that alleles of large effect will be in near fixation. However, whether they are more likely to be at the + or the − state depends on details such as the position of the optimum and the deviation from it. For example, in File S2 we show that optima positioned toward the largest (smallest) trait value z_x bias alleles to the + (−) states. Can we estimate how likely are alleles to be close to a particular state?

We assume that the trait mean is at the optimum and focus on the state of one particular allele. We study how the probability ρ that the focal allele is at the + state depends on its effect γ. In this case, we take ρ to be a probability calculated over all possible states (peaks), where we assume that the rest of the (background) loci contribute in a way that keeps the trait at the optimum. We assume that for all alleles of large effects the initial conditions are such that $\mathrm{Pr}_0(-) = \mathrm{Pr}_0(+) = 1/2$. Numerically, we perform many runs that start close to uniformly randomly selected peaks and let the system reach equilibrium. Then we count how often alleles of effect γ are in the + state. In *Appendix B* and *Appendix C* we show that

$$\rho_j = \frac{1}{1 + \exp\left[-2(z_\circ/V)\sqrt{\gamma_j^2 - 4(\mu/S)}\right]}, \tag{11}$$

where $V = \sum'_{i\neq j}(\gamma_i^2 - 4\mu/S)$, where $\sum'$ indicates summation on the set of alleles of large effect. Figure 4 shows that the predictions of Equation 11 are consistent with the simulations. The distribution of effects does not affect the probability of an allele being in the + or − state; only the effect of the focal allele matters.

The larger the effect of the focal allele is, the larger the probability it is in the + state (this assumes positive positioning of the optimum; for negative positioning, the converse would be true). The reason is that alleles that are closer to the threshold value are more prone to the instabilities resulting from small deterministic fluctuations around the optimum. Large alleles, on the other hand, are more often at the + state since they are more resilient to perturbations from the optimum value. Thus, once a population attains equilibrium, large alleles with effects close to $\hat{\gamma}$ are much more likely to be stuck in alternative equilibria than larger alleles.

We also find in Figure 4B that the larger the value of z_o is, the larger the probability for all loci to be in the + peak. This

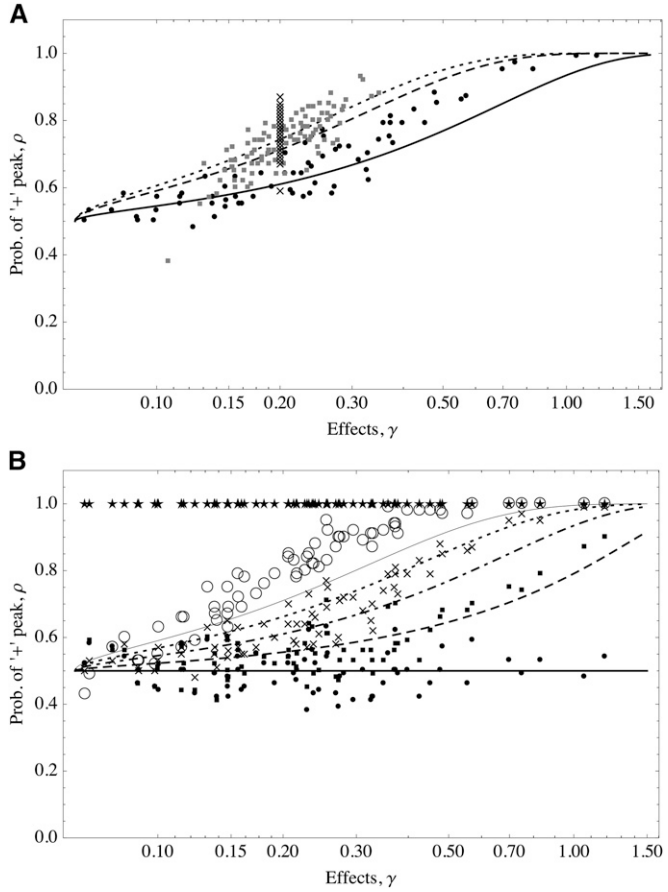

Figure 4 The probability of + alleles increases as the magnitude of their effect gets larger. Lines follow Equation 11; symbols show average of occurrences of the + state from 100 simulations. (A) The optimum is fixed and the distribution of allelic effects is varied. $z_o = 10$ (roughly halfway from the maximum trait value). Solid line and solid circles show exponential distribution (mean = 1/5); dashed line and shaded squares show gamma distribution (shape = 20, scale = 1/100); dotted line and crosses show equal effects, $\gamma = 1/5$. (B) The distribution of allelic effects is fixed and the position of the optimum is varied. Effects are distributed as an exponential (mean = 1/5); thick solid line and solid circles show $z_o = 0$, dashed line and solid squares show $z_o = 5$, dotted-dashed line and crosses show $z_o = 10$, dotted line and open circles show $z_o = 15$, and thin solid line and stars show $z_o = 20$. In all cases the trait is determined by 100 loci. $\mu = 10^{-4}$, $S = 0.1$. The initial conditions were uniformly and independently drawn for each locus.

is expected, because lager trait values require more + alleles. This obvious observation, although supported by the model, is quantitatively underestimated by it. In principle, deviations from the optimum trait value can be accommodated in Equation 11 (*Appendix C*). But this correction, at least to first order on Δz, does not fully account for the underestimation of the model at large optimum values (data not shown). What actually happens is that as the optimum is positioned closer to the range of response of the trait, the distribution of traits is considerably skewed and the Gaussian assumption fails.

Above we saw that alleles with effects that are close to the critical point are more susceptible to small perturbations to the optimum. Thus how stable the trait is to small shifts of the optimum depends on how far the frequencies are from the critical point, which in turn depends on the particular distribution of allelic effects. Under equal effects, all alleles are equally far from the critical point and thus remain stable for a long period until the deviation is large enough. But when the deviation reaches the critical value, all alleles are perturbed at the same instant.

Under unequal effects the picture is more complex. The individual equilibria of each allele are perturbed differently by deviations from the optimum. Moreover, once a given allele is perturbed and placed at an alternative state, the newly attained equilibrium is characterized by a different deviation from the optimum, potentially perturbing yet another allele. The interplay among the complex equilibria is hard to characterize in detail.

Now we determine the size of the deviations from the optimum. By employing perturbation analysis (*Appendix B*) we find that positive deviations from the optimum push allele frequencies closer to fixation. We also prove that the maximum deviation close to a given peak is of the order $\tilde{\Delta}z \simeq \min_{i\in\mathbb{L}}\gamma_i/2$, where $\mathbb{L}$ is the set of large effects (in *Appendix B* we give an exact expression to the maximum deviation). Clearly, this is bounded below by $\widetilde{\gamma}/2$, and $\tilde{\Delta}z$ depends on the particular draw of effects. This limit for the deviation is suggested by the diagram in Figure 2A: we see that the shoulders of the black lines actually occur relatively near $\delta = 1/2$ (as long as effects are large enough). Consequently, we expect that most of the time the traits will be fairly well adapted, and most of the load is given by the genetic variance, rather than by large deviations of the mean from the optimum.

Genetic variance

By direct substitution of Equations 9 and 10 into Equation 4 we see that the genetic variance that is maintained by mutation–selection balance is $\gamma^2/2$ per locus at the intermediate state and $2\mu/S$ per locus near fixation. Contrast this to the genetic variance predicted by the HoC, which is the same as for traits controlled by equal but large effects; *i.e.*, $\nu = 2n\mu/S$. Under unequal effects, if $\Delta z = 0$, the genetic variance is

$$\nu = 2n_f\frac{\mu}{S} + \frac{1}{2}\sum_{k\in\mathbb{S}}\gamma_k^2, \tag{12}$$

where n_f is the number of alleles of large effect, and the set $\mathbb{L}$ contains the n_s alleles with small effects, $\gamma^2/2 < 2\mu/S$; clearly, $n = n_f + n_s$. Note that the first term is due to alleles that are close to fixation, and their contribution to the genetic variance is independent of their effect, and the second term is due to alleles of small effect, which are at intermediate frequency. Equation 12 is one of our central results.

With this result we come back to Figure 1: in Figure 1B the genetic variance of the trait with unequal effects is lower than that of the HoC. That is because 24 alleles are of small effect.

Equation 12 correctly predicts the equilibrium variance, $\nu = 0.064$. However, note that if we use the HoC with n_f (instead of n), then $\nu \simeq 0.052$, >80% of the total variance.

Thus, the HoC variance bounds the genetic variance under unequal effects. Specifically, Equation 12 implies that $\nu_{\mathrm{HoC}}(n_f) \leq \nu \leq \nu_{\mathrm{Hoc}}(n)$, where $\nu_{\mathrm{HoC}}(m)$ is the HoC variance with m loci. With no deviations from the optimum, the load is proportional to the genetic variance. Under the HoC the load is always $L = S\nu/2 = n\mu$. However, because under unequal effects the genetic variance is smaller, the mutational load will also be smaller and dependent on the distribution of alleles.

The equilibrium genetic variance depends on the distribution $\mathcal{P}(\gamma)$ of allelic effects. Even though alleles near fixation contribute to ν independently of γ, the proportion of alleles of large effects will change with $\mathcal{P}$. For example, fixing the expected value of γ at a value larger than $\hat{\gamma}$, but allowing the shape of the distribution to change, will result in different proportions $P = n_f/n$ of alleles of large effect (Figure 5). In this way, we keep the whole range of response of the trait comparable across different distributions of effects. Distributions peaked around the mean will correspond to traits with alleles of large effect, all of which will be near fixation (Figure 5A, yellow stars); thus 100% of the variance is due to alleles of large effects and will match the HoC variance (Figure 5B, yellow curve). For distributions that are more spread, the traits will have mixed effects (Figure 5, red squares/line and green circles/line): n_s increases to 96, with ~7% of the variance due to alleles of small effects, and $n_s = 342$, ~17% of the variance is due to alleles of small effects (red and green, respectively). The extreme case will be for positively skewed distributions, such as the exponential, where the proportion of alleles of large effects is much smaller (Figure 5A, black diamonds), and the genetic variance will be considerably lower than that of the the HoC (Figure 5B, black curve): roughly half of the alleles ($n_s = 478$) are of small effect, but contribute by 20% to the total variance.

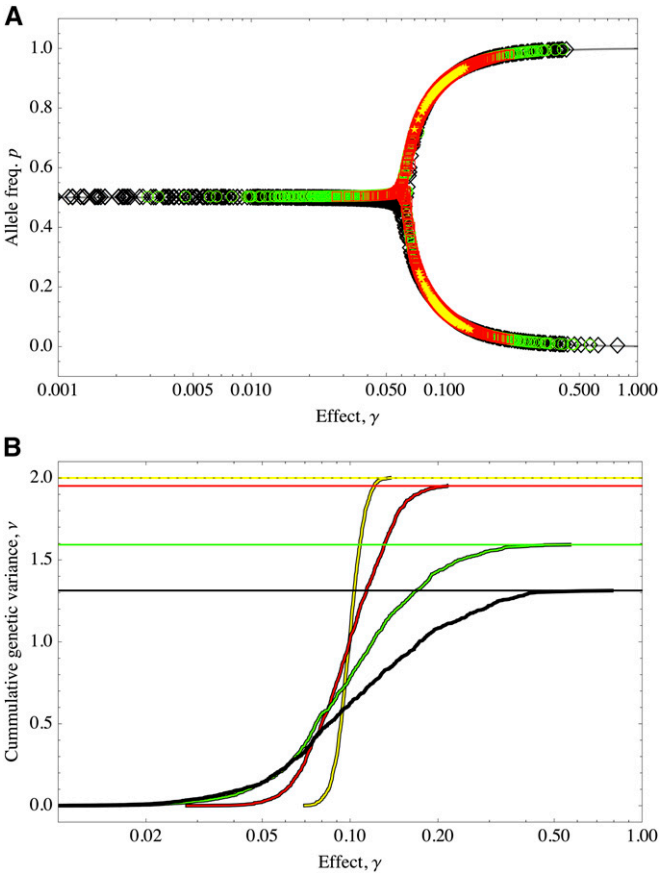

Figure 5 (A) Equilibria under different distributions of allelic effects. Symbols show data from simulations. Initial frequencies were drawn uniformly in (0, 1) and the system numerically evolved to equilibrium. Gray lines show equilibria of allele frequency and symbols show numerical equilibria. (B) Cumulative contribution to the genetic variance under different distributions of allelic effects $\mathcal{P}(\gamma)$. Solid curves show data corresponding to the simulations in A. Solid horizontal lines show equilibrium genetic variance (Equation 12). Dotted line shows genetic variance of the HoC. $\mathcal{P}(\gamma)$ is a gamma distribution of mean = 0.1 with shape parameters π as follows: yellow, stars, $\pi = 10^{-3}$; red, squares, $\pi = 10^{-2}$; green, circles, $\pi = 5 \times 10^{-2}$; and black, diamonds, $\pi = 10^{-1}$. $\mu = 10^{-4}$, $S = 0.1$, $n = 1000$, $z_o = 0$.

Distribution of phenotypic equilibria

For many loci, the number of allelic equilibria can be astronomical. Nevertheless, under equal effects it can be calculated explicitly (Barton 1986). When mapped to trait mean and genetic variance, the number of distinct equilibria is smaller, since many combinations of allelic effects have equivalent, or at least very similar, trait mean and variance. Figure 6 shows how the phenotypic states change when we keep the mean of the allelic effects constant, but increase its variance: the genetic variance decreases (see also Figure 5), and deviations from the optimum have less effect on the genetic variance, making Equation 12 a good approximation. Notably, we find that the number of values of trait mean and genetic variance increase when the distribution of effects spreads. However, these equilibria become more similar and closer to each other.

Although it is hard to count the states precisely, we can study how a given phenotypic equilibrium is affected as the asymmetry of the unequal effects is increased. For instance, suppose under equal effects we track a set of initial conditions Π that lead to a particular point in the "phenotypic" $(\bar{z}, \nu)$-space. The states to which these trajectories converge (basins of attraction) are symmetric in the sense that exchanging + alleles at one locus with − alleles at another does not change the phenotypic states. If we keep constant the set Π, but now change one effect by a small amount, how is the distribution of phenotypic states affected? Many of the trajectories starting at Π that under equal effects converged to equilibria characterized by the same trait mean and variance will now converge to different points in $(\bar{z}, \nu)$-space. The symmetries are broken and exchanging alleles at that locus with another one affects the trait mean and variance. Hence, the phenotypic equilibria show bifurcations when we perturb the effects. We can

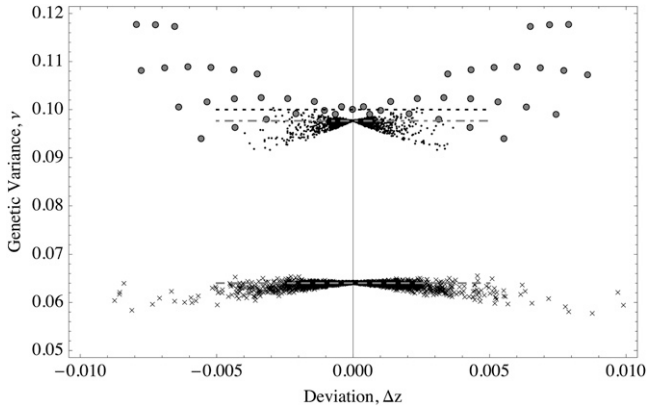

Figure 6 Phenotypic equilibria under different distributions of effects. Shaded circles, equal effects; small solid circles, effects tightly clustered around the mean (gamma distributed with variance = 1/1000); crosses, exponentially distributed effects (variance = 1/100). In all cases the mean effect is 0.1. Points are results from numerical calculations for 11 equidistant optima $z_o \in [-z_x/2, z_x/2], z_x = \overline{\gamma} = 5$, at each point employing 200 runs with uniform random initial conditions. $\mu = 10^{-4}$, $S = 0.1$, $n = 50$.

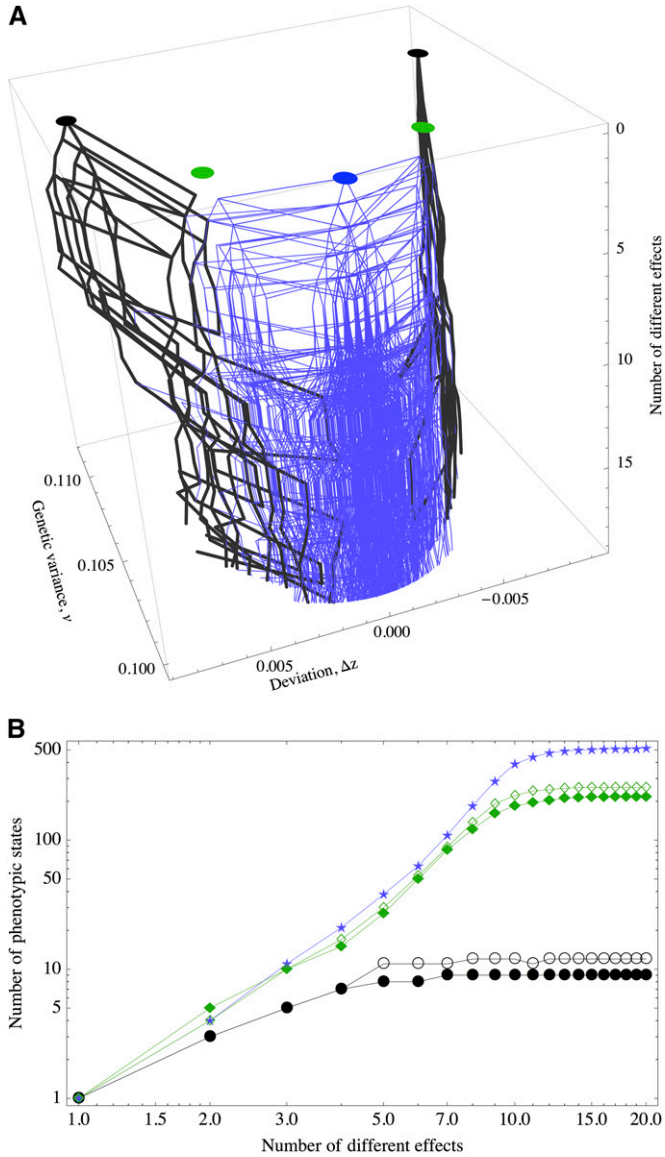

Figure 7 (A) Connectedness of macroscopic fitness peaks as the number of different effects of a polygenic trait is systematically increased from 1 to 20. The thick black lines originate at suboptimal equilibria, and the thin blue lines originate at the optimal equilibrium ($\delta z = 0$). (B) Number of new equilibria derived from the suboptimal equilibria under equal effects (black circles and green diamonds) and at the optimum equilibria (blue stars) as the number of different effects is increased. $\mu = 10^{-4}$, $S = 0.1$, $n = 50$, $z_o = 0$.

repeat this procedure by then perturbing a second locus, and so on. Thus, if we represent the phenotypic states as nodes, and we connect these nodes according to the initial condition that led to their corresponding phenotypic states, we will have a graph that represents the increase in complexity of the adaptive peaks. Unless we have a way to cover the initial space uniformly, this does not ensure complete counting of the number of phenotypic equilibria. However, even if the subsampling of the basins of attraction is poor, the method quantifies how complex the space becomes as we increase the variance of the allelic effects.

If we carry out this procedure for different phenotypic states under equal effects, then we will have several of these graphs. If these graphs share nodes, then it means that the adaptive landscape is more accessible to better-adapted equilibria (because under unequal effects the equilibria have less genetic load, Figure 6). In fact, in Figure 7 we see an example for three graphs derived from the optimal peak and two suboptimal peaks.

Altogether, this exposes that under unequal effects the fitness landscape is more complex or "rough", but the solutions are generally closer to the optimum. Surprisingly, perturbing slightly the effect of only one or two loci is enough to overlap different graphs with common nodes, indicating that unequal effects act like a funnel to guide the trajectories to nearly optimal states.

Initial Response to Selection

We saw that there are two well-defined regimes that clearly separate alleles of large effect from alleles of small effect. If the optimum shifts, which alleles respond first? A related question is, Is the initial rate of change of the trait driven mainly by alleles of large or small effect? Although these two questions are related, they are not the same; even if, for example, alleles of small effect sweep first, they might not drive a substantial displacement of the trait. Conversely, even if an allele of large effect sweeps first, its overall effect might be negligible when compared to a background of very many loci of small effect.

To calculate the rate of response of an allele, assume that the population is at equilibrium at a local peak with no deviation from the optimum that is at z_o. Suddenly, the optimum is placed at another value z_f. Equation 6 implies that at each locus

$$\frac{dp}{dt} = 2S\gamma pq\Delta\Omega, \tag{13}$$

where $\Delta\Omega = z_f - z_o$. For alleles of small effect, the right-hand side is $(S\gamma/2)\Delta\Omega$. Therefore, alleles with infinitesimally small effects will be nearly neutral and will have a vanishingly small rate of response to selection. As the effects become closer to (but still smaller than) $\hat{\gamma}$, the rate of response is larger. Consider now alleles that have infinitely large effects. The right-hand side of Equation 13 is $(2\mu/\gamma)\Delta\Omega$ and implies that since these alleles will be almost fixed, there is little variation to select on. Consequently, their rate of response is also vanishingly small. As the effect become closer to (but still larger than) $\hat{\gamma}$, the rate of response become larger. Therefore, those alleles with effects close to $\hat{\gamma}$ will have the earliest response to selection because they are the most sensitive to deviations from the optimum. Thus, the maximum response for each limit is given by the effect that is exactly at the critical value $\hat{\gamma}$. Evaluating Equation 13 at $\hat{\gamma}$ we find that the rate of response is at most

$$\left(\frac{dp}{dt}\right)_{\max} = \sqrt{\mu S}\,\Delta\Omega, \tag{14}$$

indicating that alleles with effects close to $\hat{\gamma}$ drive the initial response of the trait to selection.

Long-Term Response to Selection

The long-term response of a polygenic character to the displacement of the optimum trait value can be driven by alleles other than those of intermediate effect. Although the alleles of large effect evolve slowly in the beginning, they can eventually gain representation and evolve much faster. A general closed solution for the dynamics is neither possible nor useful, as the behavior of the allele frequencies is rather complicated. The question is whether the theory developed above can be useful to gain insight into the long-term response of the trait.

Abrupt displacement of the optimum

We assume that repositioning the optimum happens always within the range of response of the trait and far from the extreme values given by $z_x = \sum \gamma_i$; *i.e.*, $-z_x \ll z_o \ll z_x$. The equilibrium analyses revealed that the particular position of the optimum is not decisive for equilibria or stability. Instead the deviation from the optimum is the important factor. Thus, if the population eventually adapts to the new optimum value, the genetic variance that is maintained at the newly established equilibrium will be more or less the same as in the beginning. In the transient time, the dynamics will be complicated and depend on the specific initial conditions (the adaptive peaks where the population initially stands) and on the distance to the optimum.

If the optimum changes abruptly and is larger in magnitude than the largest of the effects, all the equilibria will be perturbed, favoring an increase in the frequency of the those alleles that diminish the deviation from the optimum. That is, if the new optimum value is smaller (larger) than that of the original optimum value, − (+) alleles will increase in frequency. This displacement is seen in the diagram in Figure 3 toward the top (where only one stable border is initially beneficial), with a gradual return of the line to low values of Δz. Therefore, we expect to observe a transient increase in the genetic variance. Figure 8 provides an example where these patterns are in fact found.

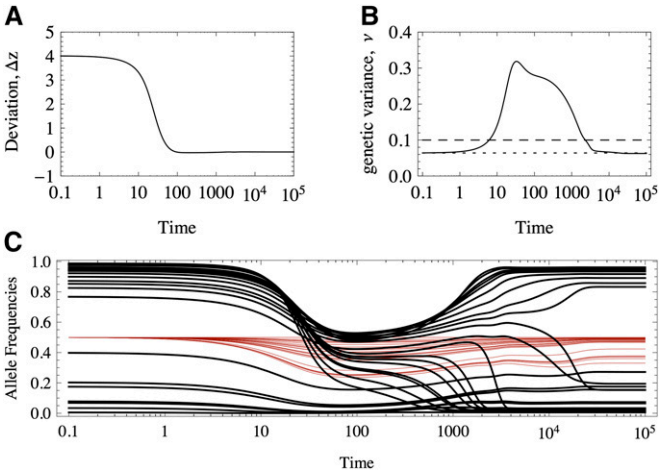

Figure 8 Response to an abrupt displacement of the optimum of a polygenic trait. (A) Deviation of the trait mean from the newly positioned optimum. (B) Genetic variance. Solid black line, exact numerical results; dashed black line, house of cards prediction ($\nu = 2n\mu/S$); dotted black line, exact value from Equation 12. (C) Response of the allele frequencies: black lines, alleles of large effect; thin red lines, alleles of small effect. The trait is constituted by $n = 50$ loci, 26 of large effects and 24 of small effects, distributed as an exponential of mean = 1/10. $\mu = 10^{-4}$, $S = 10^{-1}$, $z_o \simeq -2$.

In the example in Figure 8, of 50 loci, 26 are of large effect and contribute >80% of the initial variance, whereas 24 are of small effect, contributing the remaining 20% of the variation. In Figure 8C we see that many of the large alleles shift in frequency and some sweep, transiently raising the genetic variance. In this case, since the optimum shifts from $z_o = 2$ to $z_o = -2$, the + (−) alleles decrease (increase) in frequency. Alleles of small effect are displaced, but not substantially. Most of the transient variation that is generated is due to sweeps of alleles of large effects.

As hypothesized above, even if transient dynamics are very complex (Figure 8), when a new equilibrium is attained, the final deviations from the optimum are small, and the genetic variance is close to that of Equation 12 (see also Figure 1). As we saw in the Introduction, under equal effects the population will evolve to a suboptimal state where plenty of genetic variation is maintained. Why do populations end up better adapted under unequal effects?

At first we might think that a bulk of alleles of small effects could provide enough background variation, allowing the population to explore the genetic space more efficiently. However, the response of a trait constituted only

by alleles of large effect is virtually the same as that of a trait that contains also alleles of small effect, in as much as the initial genetic variation contributed by the latter alleles is small, and the optimum is not too close to the maximum trait value (File S3). Thus, concerning the response, alleles of small effect can be regarded as nearly neutral.

Another plausible explanation lies in the rate of "beneficial mutations" (in the sense that these are mutations that approach the trait mean to the new optimum). Because under equal effects the response of allele frequencies is synchronized, most mutations are initially beneficial. However, due to the epistatic nature of the fitness landscape, those initially beneficial mutations are not necessarily beneficial once the rarer alleles have increased their representation in the population and might even become detrimental. Furthermore, these alleles arise and increase their frequency on the same timescale. Under unequal effects different mutations arise at different times and can compensate the load contributed by previous mutations. In fact, because of epistasis, we expect and in fact we find (Figure 8C) that some alleles that are initially beneficial increase in frequency, but afterward become detrimental and decrease in frequency again. This "prevention of sweeps" has been observed in polygenic traits with up to eight loci (Pavlidis *et al.* 2012). However, under equal effects allele frequencies remain synchronized along evolution, and it is unlikely that the initial conditions and the shift in the optimum are in general finely tuned in such a way to allow the population to reach a local peak that is close to the global optimum.

Since the previous examples suggest that adaptation is driven mostly by alleles of large effect, an interesting question that follows is, What happens when traits are controlled principally by alleles of small effect? First of all, Equation 12 indicates that the genetic variance will be much lower than that of the HoC. We find that the population eventually adapts (although somewhat slower), but with virtually no change in the genetic variance. In this example, the trait has only three alleles of large effect, which contribute 11% of the genetic variance, and 397 loci with alleles of small effect, which contribute the remaining 89% of the variation. The three alleles of large effect sweep, but ultimately do not affect the genetic variation substantially (assuming their HoC contribution), and some alleles of small effect are strongly shifted.

Figure 9 shows the response of a system of alleles of principally small effects. Although 400 loci determine the trait, we estimate that the effective number of alleles $n_e \simeq 28$ (see File S4). Assuming constant genetic variance given by the HoC (but using n_e), we find that directional selection toward the optimum explains the response of the trait (although it fails to predict a minor final deviation from the optimum). This experiment highlights why we might find stasis of the genetic variance and a sustained response to selection, which is caused by innumerable alleles of small effect. Under these circumstances, although experimental assays would be able to detect only a few major loci (Hindorff *et al.* 2009; Visscher *et al.* 2012), these turn out to be the least relevant to explain the quantitative genetic variation.

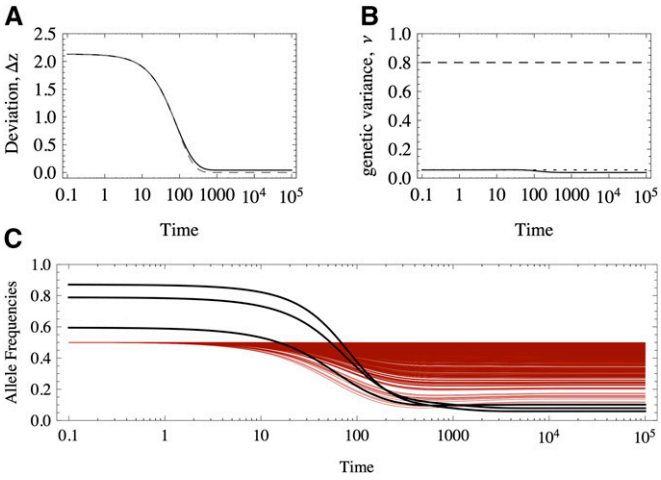

Figure 9 Response to an abrupt displacement of the optimum of a polygenic trait constituted mainly by alleles of small effect. (A) Deviation of the trait mean from the newly positioned optimum. Solid black line, exact numerical results; dashed gray line, approximation assuming an effective number of loci, $n_{eff} = 28$ and constant genetic variance, $\nu \simeq 2n_e\mu/S$ (see text). (B) Genetic variance. Solid black line, exact numerical results; dashed black line, house of cards prediction, ($\nu = 2n\mu/S$); dotted black line, exact value from Equation 12. (C) Response of the allele frequencies: black lines, alleles of large effect; thin red lines, alleles of small effect. The trait has $n = 400$ loci, 3 of large effects and 378 of small effects, distributed as an exponential of mean = 1/80. $\mu = 10^{-4}$, $S = 10^{-1}$, $z_o \simeq -2$.

Slowly moving optimum

If the optimum is shifted slowly enough, the deviation Δz remains very small. We also see that the genetic variance then hardly changes (for example, Figure 10). After some time the population keeps evolving but reaches a stationary state. If the optimum suddenly stops, the population settles at a state characterized by a lower genetic variance, but larger deviation from the optimum as in the case when it adapts to a rapid shift of the optimum, as in the previous section. How can we explain these patterns?

Under the infinitesimal model the population achieves a stationary lag from the optimum given by $\Delta z^* = -\kappa/2S\nu$, where κ is the speed of the moving optimum (Lynch and Lande 1993; Jones *et al.* 2004). However, we see in Figure 11 that this approximation fails for a finite number of loci of unequal effects.

Suppose that a moving optimum changes linearly in time: $z_o(t) := \Omega_0 + (\Omega_f - \Omega_0)t/T$. For simplicity we consider optima starting at $-\Omega$ and ending at Ω. Hence the speed of the moving optimum is $\kappa = 2\Delta\Omega/T$.

By summing Equation 6 over loci and using Equations 3 and 4, we find that during transient evolution the deviation from the optimum is given by

$$\frac{d\Delta z}{dt} = -2\nu\Delta z + Sm_3 - 2\mu\bar{z} - \kappa, \qquad (15)$$

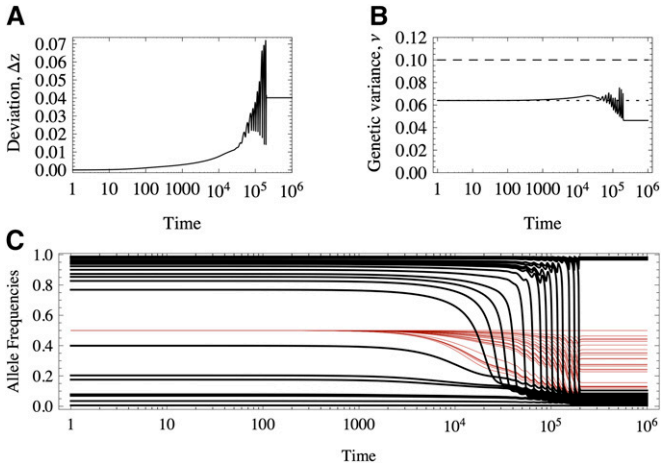

Figure 10 Response to a gradually shifting optimum of a polygenic trait. (A) Deviation of the trait mean from the newly positioned optimum. (B) Solid black line, exact numerical results; dashed black line, house of cards prediction ($\nu = 2n\mu/S$); dotted black line, exact value from Equation 12. (C) Response of the allele frequencies: black lines, alleles of large effect; thin red lines, alleles of small effect. The optimum linearly moves from $-z_o$ to z_o between $t = 0$ and $t = 10^4$ and afterward stays constant at z_o. Other parameters are as in Figure 8.

where $m_3 = \sum_i \gamma_i^3 p_i q_i (2p_i - 1)$ is the third moment of the allelic effects. If the deviation from the optimum reaches a stationary state Δz^* where $d\Delta z^*/dt = 0$, then

$$\Delta z^* = \frac{\kappa + 2\mu\bar{z} - Sm_3}{2S\nu}. \qquad (16)$$

Under the infinitesimal model, the breeding values are normally distributed, which implies that $m_3 = 0$. The genetic variance due to mutational effects is finite, but the mutation rate decreases with the inverse of n and the term $\mu\bar{z}$ can be neglected. Consequently, Equation 16 reduces to the approximation of the infinitesimal model (Lynch and Lande 1993; Jones *et al.* 2004). What limits are then necessary from the point of view of our model with a finite number of loci?

The stationary lag is not a constant; it represents a quasi-equilibrium state, and so we need to know how $\bar{z}$, ν and m_3 change in time. This is not feasible in an exact way, except under restrictive limits such as the infinitesimal model. Even under other simple assumptions, such as the HoC, predicting higher moments is hard (Barton 1986; Barton and Turelli 1987; Bürger 1991).

Figure 11 shows that the third moment of allelic effects, m_3, is relevant for an accurate prediction of the lag. In fact, if all terms of Equation 16 are considered, there is virtually no distinction between the stationary lag approximation and the actual lag. However, neglecting the third moment does affect the prediction substantially. But the extent to which m_3 is relevant depends on the distribution of effects.

Figure 11B shows an example for a trait constituted only by alleles of small effects. The third moment is small, and neglecting it leads to a good approximation of the stationary

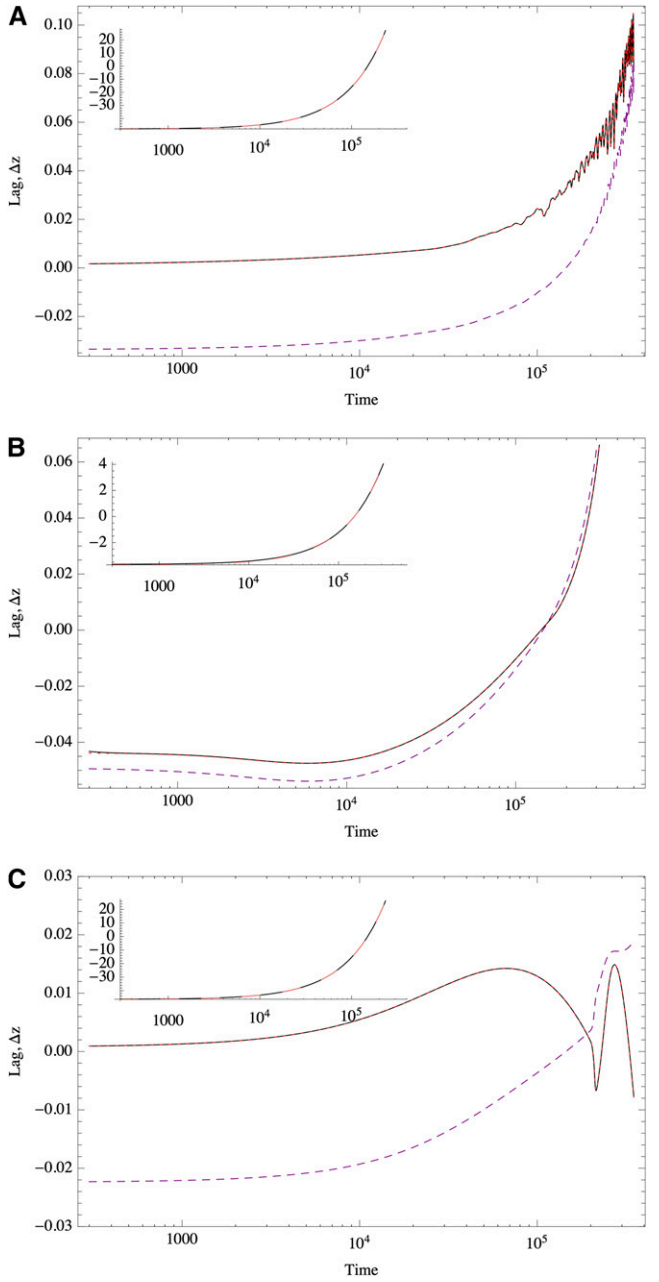

Figure 11 Stationary lag approximation for the response of polygenic traits to a steadily moving optimum. In all cases the trait is constituted by $n = 1000$ loci. The optimum moves steadily, shifting from $-\Omega$ to Ω in $T = 300{,}00$ time units. The value of Ω was chosen to match the random initial condition. Solid black line, lag $\bar{z} - z_o$; dotted red line, stationary lag Δz^* (from Equation 16); dashed purple line, stationary lag neglecting the third moment of the allelic effects. The insets compare $\bar{z}$ (black) and z_o (dashed pink line) along all times. The trait is determined by (A) exponentially distributed effects (mean = 0.1), $n_s = 435$, $\Omega = 48$; (B) exponentially distributed effects (mean = 0.01) $n_s = 1000$, $\Omega = 3.72$; and (C) equal effects with mean $\simeq 0.1$, $n_s = 0$, $\Omega = -46$. In all cases $S = 0.1$, $\mu = 10^{-4}$.

lag. This is consistent with the infinitesimal model as a limit of many loci of small effects.

When traits are determined by alleles of large but equal effects, the distribution of allelic effects is also asymmetric. As the optimum advances, traits with unequal effects allow

many small adjustments. These gradual changes allow fine tuning of the deviation from the optimum. Under unequal effects, this results in high frequency but low amplitude fluctuations of the lag. However, under equal effects the allele frequencies change in a coordinated fashion and the equilibria are more robust to deviations from the optimum. Thus, we observe fewer but larger fluctuations.

Comparing Figure 11, A and B, we find that traits with mixed allelic effects initially respond smoothly, but eventually enter a highly fluctuating phase. This does not happen when all the effects are small. We must point out that strong fluctuations can occur when the optimum is very close to the range of response of the trait. In File S5 we show that different initial conditions converge to the same erratic trajectories and that these are not chaotic, but quasi-periodic (deterministic but unpredictable fluctuations of many different frequencies, characterized by a zero Lyapunov exponent).

The existence of the quasi-periodic phase explains why, if the optimum suddenly halts, the population remains stuck at a local optimum. This is because the populations are not able to wander freely in the fitness space. Being driven by the moving optimum, they are forced to stay in states that keep a certain deviation from it. In turn, in equilibrium, this poses some directional selective pressure, which biases even further the allele frequencies, resulting in a loss of genetic variation.

In File S5 we also study a moving optimum that oscillates smoothly with different frequencies and amplitudes. The lag enters a periodic phase of many frequencies (*i.e.*, it is not smooth), and the fluctuations increase as the frequency and the amplitude increase. We also study damped oscillations. Surprisingly, once the oscillations stop, the population ends up better adapted when compared with linearly moving optima.

Discussion

Our analyses help us to understand the relative contributions of alleles of large and of small effect in the maintenance of genetic variation and in the response to stabilizing selection. Interestingly, our analysis questions whether the specific distribution of allelic effects is relevant. Naturally, a more robust interpretation of these results requires understanding how genetic drift affects the distribution of allele frequencies. In this model, alleles of small effect are at intermediate frequencies. However, these will be fixed by genetic drift. This could induce major changes to our results when there are many alleles of small effect. Thus, selection on many traits and genetic drift might change the picture substantially by maintaining larger variation, even though the details of the distribution of allelic effects might not be relevant.

Under genetic drift, the eventual fate of any allele is fixation or loss. Although drift may seem an additional complication, it also has an interesting effect, namely to allow access to parts of the fitness landscape that were inaccessible from a given state of a deterministic population (De Vladar and Barton 2010). In this sense, genetic drift smooths the landscape, and although stochastic effects are introduced, the expected trajectories are somewhat regularized, because the populations can easily escape suboptimal peaks (Wright 1935; Barton 1989), converging to fitter states. In this sense, drift aids adaptation, allowing alleles to jump across peaks by mutation and genetic drift (Wright 1931, 1932; Coyne *et al.* 1997).

The stochastic HoC (Bürger 2000, p. 270) shows that, on average, each locus at equilibrium contributes to the genetic variance by $2\gamma^2 N\mu/(1 + \gamma^2 NS)$. Thus, the expected genetic variation, $\langle \nu \rangle$, is smaller than the ν_{HoC}. Alleles of small effect will be fixed by drift, but their contribution is still small compared to alleles of large effect. Consequently, on average, the response of the trait is slower in finite populations than in infinite populations, which was already observed by us for the case of equal effects (De Vladar and Barton 2010). However, in a strong selection regime, *i.e.*, $\gamma^2 NS \gg 1$, even alleles of small effect ($\gamma^2 NS < \mu$) will be near fixation. Thus for sufficiently large populations, the HoC approximation should hold. However, the problem is far from trivial because these fixed alleles of small effect can further induce deviations of the trait mean from the optimum.

These analyses assume that at equilibrium the trait is well adapted. Other factors such as asymmetric mutation rates can maintain a deviation from the optimum (Charlesworth 2013). In this case, a stationary population effectively experiences directional selection and maintains even more genetic variance than when the trait matches the optimum. In our analyses we find in some cases that when traits deviate from the optimum, there is more genetic variance. Under asymmetric mutation rates, as in Charlesworth's model, the deviation from the optimum is maintained by two opposing forces: an asymmetric flux of mutations and directional selection toward the optimum. In our model the populations simply stand at a suboptimal peak.

Selection on many traits and pleiotropy

It is often argued that stabilizing selection acts on multiple traits. Under a common polygenic basis, two traits that are subject to antagonistic selective pressures remain at an intermediate value that is a compromise among the optimal solutions. Alleles of large effect that are not subject to these pleiotropic effects can contribute significantly to the response to selection, even though their contribution to the genetic variance is negligible due to the large number of loci (Kelly and Rausher 2009 provide many examples). In this section, we show that our previous results are relevant in this larger context.

A more general calculation for many traits under selection shows that if several traits are all adapted to their optima, there are still two classes of alleles, near fixation and at intermediate frequency, but the criterion for locus i to be near fixation is $\Gamma_i \equiv \sum_k \gamma_{ki}^2 S_k > 4\mu$, where S_k is the selection strength on trait k and γ_{ki} is the allelic effect of locus i

on trait k (*Appendix D*). However, if the optimum for one trait favors alleles at the + state, and the optimum of the other trait favor alleles at the − state, then the net effect of selection on an allele might partly neutralize. In this case, deviations from the optima will exist and both alleles will be maintained at intermediate frequency, and the genetic variance in the population will be high, as they will contribute by $\gamma_{ki}^2/2$ even if $\gamma_{ki} > \hat{\gamma}$.

For multiple traits that share a polygenic basis, only a few principal components will experience strong stabilizing selection; all the other components will be subject to only weak selection; otherwise the genetic load would be prohibitively large (Barton 1990). Hence, it remains unclear when (and unlikely that) a particular focal trait is the main component of fitness (Barton 1990). Therefore, if the stabilizing nature of selection is attributed to pleotropic factors, the equilibrium genetic variance will be decreased (Turelli 1985; Barton 1990; Slatkin and Frank 1990): if the strengths of selection on the M traits are the same, we get that $\nu = \nu_{\mathrm{HoC}}/M$.

The observed differences in fitness can be due to other correlated traits, as explained above, or to pleiotropic effects directly affecting fitness. For morphological traits, the distribution of allelic effects is positively correlated with fitness effect (Keightley and Hill 1990). However, it remains difficult to disentangle whether pleiotropy or multivariate selection is the acting mode of fitness reduction (Barton 1990; Zhang and Hill 2003)

Under antagonistic selection the picture is different. The HoC model for many traits (Turelli 1985) shows that the genetic variance of one trait depends on the strength of selection of the other traits (even if the traits are uncorrelated). In this case each locus near fixation contributes by $2\gamma_{1i}\gamma_{2i}\mu/\Gamma_i$. However, we must consider that the condition $\Gamma_i > 4\mu$ is dependent not only on the distribution of allelic effects, but also on the distribution of selective coefficients S. If the latter has a mean of zero and small variance (weak selection), the fixation condition would be hard to fulfill, and most alleles will be at intermediate frequency, leading to high genetic variance (consistent with Zhang and Hill 2003).

Admixed populations and genetic incompatibilities

Suppose that two populations that are genetically differentiated come into contact. Will a subsequent admixture result in maladapted offspring? In File S6 we show that the admixed population necessarily has larger genetic variance than the source populations, even if the latter have the same trait mean and variance. This is because the populations might be at different adaptive peaks that have the same or very similar phenotypic distribution. However, there are $\tilde{n}$ loci with distinct alleles in each population, which cause the excess variance relative to the parental mean. This will be caused by alleles of large effects, each one contributing by $\gamma_i^2 - 2\tilde{n}(\mu/S)$. This can be interpreted as the expression of genetic incompatibilities between the two divergent populations and emphasizes the role of stabilizing selection and epistasis in the process of speciation (Barton 1989, 2001). (However, this mechanism is of a different nature than the paradigm of Dobzhansky–Müller incompatibilities). After secondary contact the population might develop isolation and readapt to its original state, retaining the incompatible alleles, or it might hybridize and readapt to a new state. This will depend on the initial degree of admixture, but also on the otherwise negligible deviations from the optimum as well as on genetic drift, factors that we have not considered.

SNPs as genomic signatures of stabilizing selection

Under the assumptions of our analyses, most loci with high heterozygosity will have small effects, whereas alleles of large effect will have much lower heterozygosity, a result consistent with early results of the neutral theory (Kimura 1969). In turn, our results support the well-known idea that there can be substantial measurement bias in the estimation of allelic effects from QTL or GWAS: alleles of large effect will be harder to detect than polymorphisms of alleles with small effect. For instance, most effects that we can map are expected to be small. This is consistent with the knowledge that most alleles have small effects. Furthermore, if alleles of large effect are common, our results indicate that they will be close to fixation, and thus rare in the population, and consequently less likely to be detected.

Ignoring drift and equating n_s to the SNPs on a genome of size n and the proportion of fixed alleles to $P = n_f/n$ and assuming that this proportion is homogeneous not only across the genome, but also across the set of loci that affect different traits (questionable suppositions of course), this implies that traits are approximately at a fraction P of the total genetic variation, as we saw above. The ratio n_s/n is on the order of 1:100 or 1:1000, and thus $P \simeq 0.99$. The evolution of these traits is mutation limited rather than by standing genetic variability. These estimations assume linkage equilibrium, as are the SNPs identified by GWAS, which are often spread across the genome. (Clearly, this does not apply within genes or coding or regulatory sequences, as linkage is tight).

Different populations that show similar trait distribution and genetic variation may still differ at individual SNPs, especially if these have large effects. Thus, a particular allele might not be uniquely associated with a particular trait, even if they are causally related. This justifies and is consistent with GWAS findings that several variants can be associated with different alleles. What we have shown is that these causal variants are expected to contribute equally to the genetic variance, irrespective of the specific genetic makeup of the quantitative trait(s).

Acknowledgments

The authors thank Tiago Paixao and Daniel Weissman for discussions. This project was funded by the European Research Council grant ERC-2009-AdG for project 250152 SELECTIONINFORMATION.

Literature Cited

Barton, N. H., 1986 The maintenance of polygenic variation through a balance between mutation and stabilizing selection. Genet. Res. 47(3): 209–216.

Barton, N. H., 1989 The divergence of a polygenic system subject to stabilizing selection, mutation and drift. Genet. Res. 54(1): 59–77.

Barton, N. H., 1990 Pleiotropic models of quantitative variation. Genetics 124: 773–782.

Barton, N. H., 2001 The role of hybridization in evolution. Mol. Ecol. 10(3): 551–568.

Barton, N. H., and M. Turelli, 1987 Adaptive landscapes, genetic distance and the evolution of quantitative characters. Genet. Res. 49(02): 157–173.

Barton, N. H., and M. Turelli, 1989 Evolutionary quantitative genetics: How little do we know? Annu. Rev. Genet. 23: 337–370.

Bürger, R., 1991 Moments, cumulants, and polygenic dynamics. J. Math. Biol. 30(2): 199–213.

Bürger, R., 2000 *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Chichester, UK.

Bürger, R., and J. Hofbauer, 1994 Mutation load and mutation-selection-balance in quantitative genetic traits. J. Math. Biol. 32 (3): 193–218.

Charlesworth, B., 2013 Stabilizing selection, purifying selection, and mutational bias in finite populations. Genetics 194: 955–971.

Chevin, L. M., and F. Hospital, 2008 Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics 180: 1645–1660.

Coyne, J. A., N. H. Barton, and M. Turelli, 1997 Perspective: a critique of Sewall Wright's shifting balance theory of evolution. Evolution 51(3): 643–671.

de Vladar, H. P., and N. H. Barton, 2010 The statistical mechanics of a polygenic character under stabilizing selection, mutation and drift. J. R. Soc. Interface 8(58): 720–739.

Hill, W. G., and X. S. Zhang, 2012 On the pleiotropic structure of the genotype-phenotype map and the evolvability of complex organisms. Genetics 190: 1131–1137.

Hindorff, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta *et al.*, 2009 Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA 106(23): 9362–9367.

Jones, A. G., S. Arnold, and R. Bürger, 2004 Evolution and stability of the G-matrix on a landscape with a moving optimum. Evolution 58(8): 1639–1654.

Keightley, P. D., and W. G. Hill, 1990 Variation maintained in quantitative traits with mutation selection balance—pleiotropic side-effects in fitness traits. Proc. Biol. Sci. 242(1304): 95–100.

Kelly, J., and M. Rausher, 2009 Connecting Qtls to the G-matrix of evolutionary quantitative genetics. Evolution 63(4): 813–825.

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

Kimura, M., 1969 The rate of molecular evolution considered from the standpoint of population genetics. Proc. Natl. Acad. Sci. USA 63(4): 1181–1188.

Kingman, J. F. C., 1978 A simple model for the balance between selection and mutation. J. Appl. Probab. 15(1): 1–12.

Lande, R., 1976 The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genet. Res. 26: 221–235.

Lynch, M., and R. Lande, 1993 Evolution and extinction in response to environmental change, pp. 234–250 in *Biotic Interactions and Global Change*, edited by P. G. Kareiva, J. Kingsolver, and R. B. Huey. Sinauer Associates, Sunderland, MA.

Mackay, T. F., 2001 The genetic architecture of quantitative traits. Annu. Rev. Genet. 35: 303–339.

Mackay, T. F. C., 2010 Mutations and quantitative genetic variation: lessons from Drosophila. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365(1544): 1229–1239.

Pavlidis, P., D. Metzler, and W. Stephan, 2012 Selective sweeps in multilocus models of quantitative traits. Genetics 192: 225–239.

Slatkin, M., and S. Frank, 1990 The quantitative genetic consequences of pleiotropy under stabilizing and directional selection. Genetics 125: 207–213.

Turelli, M., 1984 Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. Theor. Popul. Biol. 25(2): 138–193.

Turelli, M., 1985 Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. Genetics 111: 165.

Turelli, M., 1988 Phenotypic evolution, constant covariances, and the maintenance of additive variance. Evolution 42(6): 1342–1347.

Turelli, M., and N. H. Barton, 1994 Genetic and statistical analyses of strong selection on polygenic traits: What, me normal? Genetics 138: 913–941.

Visscher, P. M., M. A. Brown, M. I. McCarthy, and J. Yang, 2012 Five years of GWAS discovery. Am. J. Hum. Genet. 90 (1): 7–24.

Wright, S., 1931 Evolution in Mendelian populations. Genetics 16: 97–159.

Wright, S., 1932 The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc. Sixth Intl. Cong. Genet. 1: 356–366.

Wright, S., 1935 The analysis of variance and the correlations between relatives with respect to deviations from an optimum. J. Genet. 30(2): 243–256.

Zhang, X. S., and W. G. Hill, 2003 Multivariate stabilizing selection and pleiotropy in the maintenance of quantitative genetic variation. Evolution 57(8): 1761–1775.

Communicating editor: M. K. Uyenoyama

Appendix A

Bifurcation Points

The bifurcation points occur when there is a change in the stability of the equilibrium allele frequencies by varying a parameter. This means that, in addition to the equilibrium condition $dp/dt = 0$, we also require that the eigenvalue vanishes at the point of equilibrium. If we rescale the equations in terms of δ and m, we have to solve the following system:

$$\begin{aligned} m(1-2p) - (1-p)p(2\delta - 2p + 1) &= 0 \\ -2\delta - 2m + 2p(2\delta - 3p + 3) - 1 &= 0. \end{aligned} \tag{A1}$$

By eliminating p from the two equations we get that

$$8\delta^2\left(2\delta^2 + 2(m-5)m - 1\right) = (4m-1)^3. \tag{A2}$$

This formula defines the boundary in the diagram of Figure 3 in the main text. Clearly, if there are no deviations from the trait, $\delta = 0$, the right-hand side gives the critical value $\hat{m} = 1/4$. Also, as m vanishes, $\delta \to 1/2$. In general, the last equation gives the boundary for any arbitrary deviation.

We can also eliminate m and get the allele frequency p at which the bifurcation occurs, given by the solutions to the cubic:

$$-8p^3 + 4(\delta + 3)p^2 + (-4\delta - 6)p + 1 + 2\delta = 0. \tag{A3}$$

Note that to keep allele frequencies $0 \le p \le 1$, it must be fulfilled that $-1/2 \le \delta \le 1/2$.

Appendix B

Perturbation Analysis for Small Deviations from the Optimum

Consider an approximate solution to Equation 6 expressed as $p = P_0 + (\Delta z)P_1 + (\Delta z)^2 P_2 + O[(\Delta z)^3]$. The time derivative of p neglecting terms of order $(\Delta z)^3$ is

$$\begin{aligned} \dot{p} &\simeq \dot{P}_0 + (\Delta z)\dot{P}_1 + (\Delta z)^2\dot{P}_2 \\ &= (1-2P_0)\left[\mu - S\gamma^2 P_0(1-P_0)\right] \\ &\quad + (\Delta z)\left[-S\gamma\left(2P_0(1-P_0)(1-\gamma P_1) + \gamma P_1(1-2P_0)^2\right) - 2\mu P_1\right] \\ &\quad + (\Delta z)^2\left[S\gamma((2P_0-1)P_1(3P_1+2) + 6(P_0-1)P_0P_2 + P_2) - 2\mu P_2\right] . \end{aligned} \tag{B1}$$

The unperturbed solutions for P_0 are those given in the main text, *i.e.*, Equation 9. In equilibrium, we require the terms proportional to Δz^m in the last equation to vanish, which gives the following two solutions for alleles of small and of large effects, respectively:

$$P_1 = \begin{cases} \dfrac{S\gamma}{S\gamma^2 - 4\mu} & \left(\gamma^2 < \dfrac{4\mu}{S}\right) \\ \dfrac{2\mu/\gamma}{4\mu - S\gamma^2} & \left(\gamma^2 > \dfrac{4\mu}{S}\right) \end{cases} . \tag{B2}$$

Note that both quantities are negative.

The second-order perturbations give $P_2 = 0$ for alleles of small effect, and for alleles of large effect

$$P_2 = \pm\frac{4\mu\sqrt{\gamma S}(\mu + \gamma S)}{(4\mu + \gamma S)^{5/2}}, \tag{B3}$$

where the sign indicates whether the allele is in the + or the − state.

Deviations from the optimum

We can estimate the deviation from the optimum by summing over all alleles. This leads to a quadratic equation for Δz with the solution

$$\Delta z = \frac{1-\zeta_1}{2\zeta_2} \pm \left[\left(\frac{1-\zeta_1}{2\zeta_2}\right)^2 - \frac{\zeta_0 - z_\circ}{\zeta_2}\right]^{1/2}, \tag{B4}$$

where the factors ζ_k are

$$\zeta_k = 2\sum_{i=1}^{n} \gamma_i P_{ki} \tag{B5}$$

(P_{ki} is the kth perturbation at the locus i).

Maximum trait deviation

Taking the solution for alleles of large effect, we can calculate what is the maximum allowed deviation from the trait, $\tilde{\Delta}z$. For this, we equate the allele frequency to zero (for positive deviations) or equivalently to one (for negative deviations). Assuming positive deviations,

$$\min p = 0 = \frac{1}{2}\left[1 - \sqrt{1 - \frac{4\mu}{s\gamma^2}}\right] + (\tilde{\Delta}z)\frac{2\mu/\gamma}{4\mu - S\gamma^2}, \tag{B6}$$

which gives a deviation of

$$\tilde{\Delta}z = \gamma\left(\frac{S\gamma^2}{4\mu} - 1\right)\left[1 - \sqrt{1 - \frac{4\mu}{S\gamma^2}}\right]. \tag{B7}$$

Note that the larger effects tolerate larger deviations. Hence we need to take the minimum of all these deviations, to ensure stability for all alleles. This is thus given by the smallest allele of large effect. Assuming that $S\gamma^2 \gg 4\mu$ the expression above simplifies to

$$\tilde{\Delta}z = \frac{\gamma}{2}\left(1 - \frac{3\mu}{S\gamma^2}\right), \tag{B8}$$

and for alleles of extremely large effect, the deviations can be at most of order $\tilde{\Delta}z \simeq \gamma/2$, as reported in the main text.

Appendix C

Probability of Allelic States

In *Appendix C* we derive the probability ρ of finding an allele at the + state, that is, Equation 11 in the main text.

As mentioned in the main text, we assume that the trait mean has a value $Z = \overline{z} = z_\circ$ and calculate the probability ρ that allele X is at the + state. That is, $\rho \equiv \Pr[X = 1|Z = \overline{z}]$. We first decompose this probability, using Bayes' theorem:

$$\Pr[X = 1|Z = \overline{z}] = \Pr[X = 1]\frac{\Pr[Z = \overline{z}|X = 1]}{\sum_y \Pr[Z = \overline{z}|X = y]Pr[X = y]}.$$

Then, we express the trait mean as the sum over loci, $\overline{z} = \sum_i (2x_i - 1)\sqrt{\gamma_i^2 - 4\mu/S}$, where x_i indicates whether the allele is close to $x = 1$ or $x = 0$. Here we assumed that the background alleles are near fixation. Summarizing,

$$\begin{aligned}\Pr\left[\overline{z}|x_j = 1\right] &= \Pr\left[\sum_i (2x_i - 1)\sqrt{\gamma_i^2 - 4\mu/S}\middle| x_j = 1\right] \\ &= \Pr\left[\overline{z} - \sqrt{\gamma_j^2 - 4\mu/S}\right].\end{aligned} \tag{C1}$$

Since the trait is a sum over independent loci, we can approximate that the trait distribution is normal (central limit theorem). Its variance V is given by summing over the background loci of large effect:

$$V = \sum_{i \neq j} \left(\gamma_i^2 - \frac{4\mu}{S} \right). \tag{C2}$$

Now assume that the initial configurations between + and − alleles are chosen uniformly; *i.e.*, $\Pr[X = 1] = \Pr[X = 0] = 1/2$. The sum in the denominator of Bayes' theorem involves only two Gaussian terms. Putting the pieces together, this leads to Equation 11:

$$\rho_j = \frac{1}{1 + \exp\left[-2(z_o/V)\sqrt{\gamma_j^2 - 4(\mu/S)} \right]}. \tag{C3}$$

To accommodate deviations from the optimum trait value, we proceed in a similar way, but using the first-order perturbation on the allele frequencies. Notably, the variance V does not change, since all alleles are displaced proportionally to Δz. We arrive at the expression

$$\rho_j = \frac{1}{1 + \exp\left[-2(1/V)\sqrt{\gamma_j^2 - 4(\mu/S)} \left(z_o + \Delta z S \gamma_j^2 \Big/ \left(S\gamma_j^2 - 4\mu \right) \right) \right]}. \tag{C4}$$

Note that the term proportional to Δz denotes the strength of the directional selection component on allele j. For alleles of very large effect $S\gamma_j^2 \gg 4\mu$, and the term is $\sim \Delta z$, which as we saw before is very small. Thus the deviations from the optimum affect mainly alleles of large effect that are closer to the critical value $\hat{\gamma}$.

Appendix D

Stabilizing Selection on Multiple Traits

We consider a simple extension of our model, where stabilizing selection acts independently on many traits. Call $\mathbf{z} = (z_1, \ldots, z_m)$ an array of m traits that are under selection $\mathbf{S} = (S_1, \ldots, S_m)$ and $\mathbf{z}_o$ their corresponding optima. Calling $\mathcal{S} = \mathrm{diag}(S)$, we define fitness as $W_{\mathbf{z}} = \exp[-(\mathbf{z} - \mathbf{z}_o) \cdot \mathcal{S} \cdot (\mathbf{z} - \mathbf{z}_o)^T/2)$, where "·" represents the inner product. In principle we could accommodate correlation selection in the model by allowing the matrix $\mathcal{S}$ to have nonzero off-diagonal elements, but we leave out that possibility at the moment. Under weak selection, mean fitness becomes

$$\overline{W} \simeq \exp\left[-\frac{1}{2} \left(\Delta \mathbf{z} \cdot \mathcal{S} \cdot \Delta \mathbf{z}^T + \mathbf{S} \cdot \boldsymbol{\nu} \right) \right], \tag{D1}$$

where $\Delta \mathbf{z} = (\overline{\mathbf{z}} - \mathbf{z}_o)$ is the vector of deviations from the optima and $\boldsymbol{\nu}$ is the vector of genetic variances.

The trait means and genetic variances are

$$\overline{z}_k = \sum_{i=1}^{n} \gamma_{ki}(2p_i - 1) \tag{D2}$$

$$\nu_k = 2\sum_{i=1}^{n} \gamma_{ki}^2 p_i q_i, \tag{D3}$$

where γ_{ki} is the allelic effect of locus i on trait k. We are assuming that all n alleles contribute to m traits. This can be relaxed by simply assuming that $\gamma_{ki} = 0$ for some i, k.

The equilibria for this system are given by

$$0 = -p_i q_i [2\beta_i - \Gamma_i(1 - 2p_i)] + \mu(1 - 2p_i), \tag{D4}$$

where

$$\beta_i = \sum_{k}^{m} S_k \Delta z_k \gamma_{ki} \tag{D5}$$

$$\Gamma_i = \sum_k^m S_k \gamma_{ki}^2. \tag{D6}$$

We find that the two quantities above take the role of the deviation from the optimum and allelic effects on the single-trait model. In fact, Equation 7 in the main text holds, when we define $\delta = \beta/\Gamma$ and $m = \mu/\Gamma$. Consequently, the critical points, the scaled equilibria, and their stability are the same.

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.159111/-/DC1

Stability and Response of Polygenic Traits to Stabilizing Selection and Mutation

Harold P. de Vladar and Nick Barton

DOI: 10.1534/genetics.113.159111

File S1

Stability and response of polygenic traits to stabilizing selection and mutation. Supplementary Information

Harold P. de Vladar, Nick Barton

1 Stability analyses

Our goal is to derive criteria for the equilibria. First, we will calculate the eigenvalues for a polygenic system without mutation. Then by perturbation analyses we will consider small mutation rates,

Stability in the absence of mutation.

In the absence of mutation (μ=0) there are three equilibrium points given by a simpler version of Eq. 6 of the main text, namely:

$$0 = -S\gamma_i p_i(1-p_i)[2\Delta z + \gamma_i(1-2p_i)] \ . \tag{1}$$

The solutions are the two fixed states, and the point where the gradient of mean fitness vanishes, which occurs at

$$p_i = \frac{1}{2} + \frac{\Delta z}{\gamma_i} \ . \tag{2}$$

Since Δz couples all the alleles, the stability of any given point is determined by the whole genetic background which determines the value of Δz. The stability of each equilibrium is determined by the Jacobian matrix, $\mathbb{J} = \{\partial \dot{p}_i/\partial p_j\}_{i,j=1}^{n}$, where the elements are shown in Table 1. It follows that if alleles are fixed at all loci, $\mathbb{J}$ is diagonal with eigenvalues

$$\lambda_i = -S\gamma_i \left(\gamma_i \pm 2\Delta z\right) \ , \tag{3}$$

where the signs depend on whether the locus i has the allele '-' or '+', respectively. Hence, if the trait matches the optimum ($\Delta z = 0$) all fixed states

Table 1: Elements of the Jacobian matrix for a polygenic trait under stabilizing selection and no mutation.

p_i	0	$\frac{1}{2} + \frac{\Delta z}{\gamma_i}$	1
$\frac{\partial \dot{p}_i}{\partial p_i}$	$-S\gamma_i\left(\gamma_i + 2\Delta z\right)$	$-2S\gamma_i^2\left[\frac{1}{4} - \left(\frac{\Delta z}{\gamma_i}\right)^2\right]$	$-S\gamma_i\left(\gamma_i - 2\Delta z\right)$
$\frac{\partial \dot{p}_i}{\partial p_j}$	0	$-4S\gamma_i\gamma_j\left[\frac{1}{4} - \left(\frac{\Delta z}{\gamma_i}\right)^2\right]$	0

are stable. If the trait deviates from the optimum, any locus i for which $|\Delta z| > \gamma_i/2$ will become unstable and sweep to an alternative state.

Under equal effects, if the deviation from the trait optimum surpasses $\gamma/2$, all alleles in one fixation corner become unstable simultaneously, whereas in the case of unequal effects this happens only for a fraction of loci, if they fulfill the above condition.

Clearly, large deviations from the optimum will favour fixation of some alleles, which do not contribute to the genetic variance. However, small deviations in principle allow some alleles to be at intermediate frequencies, and thus generate genetic variance. In that case, each polymorphic allele will contribute to the genetic variance by $\frac{1}{2}\left(\gamma_s + 2\Delta z\right)^2$, where s indicates loci that are polymorphic.

Now we will prove that at most one locus can be polymorphic. First we prove exactly for $n_s = 1, 2$, and then for a general case. Without loss of generality we will assume that the vector of allelic effects is ordered, is bounded and positive, i.e. $0 < \gamma_1 \leq \gamma_2 \ldots \leq \gamma_n < \infty$.

Notice that the Jacobian is a block diagonal matrix:

$$\mathbb{J} = \begin{pmatrix} \mathbb{A}_s & \mathbb{A}_o \\ \mathbb{O} & \mathbb{A}_f \end{pmatrix} \tag{4}$$

the sub-matrices for the polymorphic states, $\mathbb{A}_s$ and $\mathbb{A}_0$ contain the Jacobian terms given by the central column of Table A1, and are of dimensions $n_s \times n_s$ and $n_s \times n_f$, respectively. The sub matrix $\mathbb{A}_f$ is a diagonal matrix corresponding to fixed loci, and its elements are given by the left and right columns of Table A1; it has dimension $n_f \times n_f$. $\mathbb{O}$ is a zero matrix of dimensions $n_f \times n_s$.

Because $\mathbb{J}$ is a triangular matrix, its eigenvalues are given by $\|\mathbb{A}_f - \lambda\mathbb{I}\|\|\mathbb{A}_s - \lambda\mathbb{I}\| = 0$. This implies that we can find the eigenvalues for each matrix separately. Therefore, there will be n_f eigenvalues of the form $\lambda_i = -S\gamma_i\left(\gamma_i \pm 2\Delta \mathrm{z}\right)$ for the fixed alleles, just as in Eq. 3.

Assume now that there is only one polymorphic allele, i.e. that $\mathbb{A}_s = -2S\gamma_i^2\left[\frac{1}{4} - \left(\frac{\Delta z}{\gamma_i}\right)^2\right]$ for a given i. Then, as long as the deviation from the optimum is less than half the effect, the polymorphic allele is stable.

Suppose now that $\mathbb{A}_s$ is a 2×2 matrix. Call $\boldsymbol{f} = (f_1, f_2)$, with

$$f_i = S\gamma_i\left[\frac{1}{4} - \left(\frac{\Delta\, z}{\gamma_i}\right)^2\right], \tag{5}$$

and the vector of effect of polymorphic loci $\boldsymbol{g} = (\gamma_1, \gamma_2)$. Then irrespective of the magnitude of the deviation from the optimum we get:

$$\lambda_s = -\boldsymbol{f} \cdot \boldsymbol{g} \pm \sqrt{(\boldsymbol{f} \cdot \boldsymbol{g})^2 + 12 f_1 f_2 \gamma_1 \gamma_2} \tag{6}$$

where $\boldsymbol{f} \cdot \boldsymbol{g}$ denotes the dot product between the two vectors. Thus, one eigenvalue is negative and one is positive, making the configuration with two polymorphic alleles unstable.

A more general case is hard to compute exactly, but we can approach the problem by partitioning the matrix $\mathbb{A}_s$ into two simpler matrices whose eigenvalues can be computed exactly, and then by applying Weyl's inequality (see also Horn's conjecture; Horn, 1962; Knutson and Tao, 2001), we can bound the eigenvalues of $\mathbb{A}_s$ in terms of the eigenvalues of the other two matrices.

Assume that the vectors $\boldsymbol{f}$ and $\boldsymbol{g}$ are of dimension $n_s > 0$. Consider the following two matrices: $\mathbb{B} = -2\boldsymbol{f} \otimes \boldsymbol{g}$ and $\mathbb{C} = \mathrm{Diag}\{f_i \gamma_i\}_{i=1}^{n_s}$, where $\otimes$ denotes the outer product of the vectors. Then $\mathbb{A}_s = 2S(\mathbb{B} + \mathbb{C})$. The matrix $\mathbb{B}$ has rank one, and therefore has eigenvalues $\beta = (0, \ldots, 0, -2\boldsymbol{f} \cdot \boldsymbol{g})$ where there are $n_s - 1$ zeroes. The matrix $\mathbb{C}$, being diagonal, has eigenvalues $\zeta = (\gamma_{n_s} f_{n_s}, \ldots, \gamma_1 f_1)$, where the eigenvalues are written in decreasing order, and therefore ζ has the reverse order of $\boldsymbol{g}$. Then if α are the eigenvalues of $\mathbb{A}_s$, Weyl's inequality states that $\zeta_i + \beta_1 \leq \alpha_i \leq \zeta_i + \beta_{n_s}$, which implies the following two inequalities:

$$0 < g_1 f_1 \leq \alpha_i \leq g_{n_s} f_{n_s} \ , 1 < i < n \tag{7}$$

$$g_1 f_1 - 2\boldsymbol{g} \cdot \boldsymbol{f} \leq \alpha_n \leq g_{n_s} f_{n_s} - 2\boldsymbol{g} \cdot \boldsymbol{f} < 0 \tag{8}$$

The first inequality implies that there are $n_s - 1$ positive eigenvalues, and the second inequality that there is one negative eigenvalue. (We confirmed this result using numerical calculations; data not shown). This, as stated above, requires that $|\Delta z| < \gamma_1/2$; there are no fixed points that allow larger deviations.

In conclusion, only one allele can be maintained polymorphic; any configuration with more than one polymorphic allele is unstable.

Stability under small mutation rates.

Now we will derive an approximation for the eigenvalues of the Jacobian matrix when the trait matches the optimum and mutation rates are small

compared to selection, ($\mu << S$). However, we will assume that all alleles are of large effect, i.e. $\gamma > 2\sqrt{\mu/S}$, and hence contribute to the Jacobian by $-S\gamma_i^2$ in the diagonal, and $-4\mu\frac{\gamma_j}{\gamma_i}$ in the non-diagonal. If mutation is absent, then the Jacobian is simply $\mathbb{J}_0 = -S\text{Diag}\{\gamma_1^2, \ldots, \gamma_n^2\}$, with eigenvalues $\lambda_i = -S\gamma_i^2$, as above (Eqn. 3) when $\Delta z = 0$. For simplicity, we will scale the eigenvalues as $\lambda \to \lambda/S$. We write the Jacobian as $\mathbb{J} = \mathbb{J}_0 + \mu\mathbb{J}_\mu$, where

$$\mathbb{J}_\mu = \begin{cases} -4\frac{\gamma_i}{\gamma_j} & i \neq j \\ 0 & i = j \end{cases}$$

We are therefore looking for the solution to the equation

$$\|\mathbb{J}_0 + \epsilon\mathbb{J}_\mu - \lambda\mathbb{I}\| = 0 \tag{9}$$

where ϵ=μ/S. This equation is a polynomial of degree n on λ. The central idea is to treat the determinant as a function $F(\epsilon, \lambda)$, and to perform a perturbation analysis on ϵ, and solve for λ. The eigenvalues are expanded as $\lambda = \lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \ldots$, making $F(\epsilon, \lambda) = F\left(\epsilon, \lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \ldots\right)$, which in turn is written as $F\left(\epsilon, \lambda_0, \lambda_1, \lambda_2, \ldots\right)$. Following we expand the determinant in series of ϵ up to second order:

$$F\left(\epsilon, \lambda_0, \lambda_1, \lambda_2, \ldots\right) \simeq F\left(0, \lambda_0\right) + \epsilon F'\left(0, \lambda_0, \lambda_1\right) + \frac{\epsilon^2}{2}F''\left(0, \lambda_0, \lambda_1, \lambda_2\right) + \mathcal{O}(\epsilon^3). \tag{10}$$

In general, the derivatives $F^{(n)}$ are polynomials on ϵ. Grouping terms of equal order of ϵ, leads to system of equations that give the λ_i. Since a general formula for the coefficients of the polynomial on λ is complicated, we use matrix algebra as a general framework to perform the approximations, since the compact notation of matrix algebra makes it more convenient to derive general results. For convenience, define the following $n \times n$ matrix:

$$\mathbb{F} = \mathbb{J}_0 - \lambda_0\mathbb{I} + \epsilon\left(\mathbb{J}_\mu - \lambda_1\mathbb{I}\right) - \left(\epsilon^2\lambda_2 + \ldots\right)\mathbb{I} \tag{11}$$

therefore, the characteristic equation is

$$F\left(\epsilon, \lambda_0, \lambda_1, \lambda_2, \ldots\right) = \|\mathbb{F}\| = 0 \tag{12}$$

The first term in the perturbation method is the unperturbed solution (order ϵ^0) which emerges from the first term of the Taylor expansion: $F(0) \Rightarrow \|\mathbb{J}_0 - \lambda_0\mathbb{I}\| = 0$; its roots are as above (Eq. 3).

The second term (order ϵ) follows from the derivative F', where we use Jacobi's formula, i.e. $F' = \frac{d}{d\,\epsilon}\|\mathbb{F}\| = \mathrm{Tr}\left(\mathbb{F}^\dagger \frac{d\,\mathbb{F}}{d\,\epsilon}\right)$, where $\mathbb{F}^\dagger$ is the adjugate (a.k.a. classic adjoint) matrix of $\mathbb{F}$. The only relevant term of $\mathbb{F}'$ is $(\mathbb{J}_\mu - \lambda_1 \mathbb{I})$ (higher order terms vanish when $\epsilon = 0$). Then at ϵ=0 we get that $\left(\mathbb{F}^\dagger \mathbb{F}'\right)_{i\,j} = -\alpha_{i\,j} \prod_{k\neq j}^{n} (\gamma_k^2 + \lambda_0)$ where $\alpha_{i\,i} = \lambda_1$ and $\alpha_{i\,j} = 4\gamma_i/\gamma_j$. This leads to $\mathrm{Tr}\left(\mathbb{F}^\dagger \frac{d\,\mathbb{F}}{d\,\epsilon}\right) = -\lambda_1^{(i)} \prod_{k\neq j}^{n} \left(\gamma_k^2 + \lambda_0^{(i)}\right)$. Since we require that $F'\left(0, \lambda_0^{(i)}, \lambda_1^{(i)}\right) = 0$, this implies that $\lambda_1^{(i)} = 0$, concluding that there are no terms of order ϵ.

Because the first order perturbation vanishes, it suggests that the configuration will be stable. The leading value of the eigenvalue is given by the unperturbed solution, and it will change by order of $(\mu/S)^2$. Thus even if λ_2 is positive, as long as it is bounded the eigenvalue will remain negative.

Now we will calculate the second order term to verify that it remains finite. For the term of order ϵ^2 we need to evaluate F'', that is

$$F'' = \frac{d^2}{d\epsilon^2}\|\mathbb{F}\| = \frac{d}{d\epsilon}\mathrm{Tr}\left(\mathbb{F}^\dagger \frac{d\,\mathbb{F}}{d\epsilon}\right) \tag{13}$$

the the differential operator can be exchanged with the trace, and using the additive property we get

$$F'' = \mathrm{Tr}\left(\frac{d\mathbb{F}^\dagger}{d\epsilon}\frac{d\,\mathbb{F}}{d\epsilon}\right) + \mathrm{Tr}\left(\mathbb{F}^\dagger \frac{d^2\,\mathbb{F}}{d\,\epsilon^2}\right) \tag{14}$$

Since $d^2\mathbb{F}/d\epsilon^2 \simeq 2\lambda_2 \mathbb{I}$, substitution of this identity gives

$$\lambda_2 = -\mathrm{Tr}\left(\frac{d\,\mathbb{F}^\dagger}{d\,\epsilon}\frac{d\,\mathbb{F}}{d\,\epsilon}\right) \Big/ 2\mathrm{Tr}\left(\mathbb{F}^\dagger\right) \tag{15}$$

At $\epsilon = 0$ and readily using that $\lambda_1 = 0$ we have that $d\mathbb{F}/d\epsilon = \mathbb{J}_\mu$ and that

$$\left\{\frac{d\mathbb{F}^\dagger}{d\epsilon}\right\}_{ij} = \begin{cases} -4\frac{\gamma_j}{\gamma_i}\prod_{k\neq i,j}^{n}(\gamma_k^2 + \lambda_0) & i \neq j \\ 0 & i = j \end{cases}$$

By multiplying the last two matrices we get the numerator of Eq. 15; the denominator is $\mathrm{Tr}\left(\mathbb{F}^\dagger\right) = \sum_{j=1}^{n}\prod_{k\neq j}^{n} - \left(\gamma_k^2 + \lambda_0^{(i)}\right)$. Substituting the corresponding λ_0 gives the second order terms of the eigenvalues, which are:

$$\lambda_2^{(i)} = -\frac{8\sum_{j\neq i}^{n}\sum_{m\neq i}^{n}\prod_{k\neq j,m}^{n}(\gamma_k^2 - \gamma_i^2)}{\sum_{j\neq i}^{n}\prod_{k\neq j}^{n}(\gamma_i^2 - \gamma_k^2)}. \tag{16}$$

Assuming that there $\gamma_i \neq \gamma_j \forall i, j$, the last expression simplifies to

$$\lambda_2^{(i)} = -16 \sum_{j \neq i}^{n} \frac{1}{\gamma_j^2 - \gamma_i^2}. \tag{17}$$

Thus the second order term is finite. If two or more effects are the same, the previous expression does not apply. Nevertheless, multiplicities can be incorporated in the more general expression of Eq. 16, but we leave that case aside as it is unlikely that two effects are exactly the same.

Summarizing, the perturbation parameter is $\epsilon{=}\mu/S$, thus the eigenvalues are as

$$\lambda^{(i)} = -\gamma_i^2 + \left(\frac{\mu}{S}\right)^2 \lambda_2^{(i)} + \mathcal{O}\left(\left[\frac{\mu}{S}\right]^3\right) \tag{18}$$

which completes the approximation up to second order.

2 Allele Frequency Spectra

At equilibrium, alleles of large effect are close to fixation. However, whether most of these are at the '+' or '-' state depends on the relative position of the optimum, and on how much the trait deviates from it. Intuitively, we expect that if $z_\circ > 0$ (conversely, $z_\circ < 0$) most alleles will be at the '+' ('-') state. Figure S1 presents the allele frequency spectrum from numerical runs showing that the position of the optimum strongly biases the allelic states.

The proportion of alleles of large effect will determine the relative height of the central peak (central column in Fig. S1) with respect to the peaks close to the borders. For instance, for traits with many alleles of small effect the central peak will be substantial. In contrast, traits composed mainly of alleles of large effect will have a moderate or practically absent central peak. This is best visualized in the central column for Fig. S1. If $\sqrt{\mu/S}$ is increased, the spectrum is biased towards fixation, and as it is decreased, the spectrum is less dense at the fixation borders.

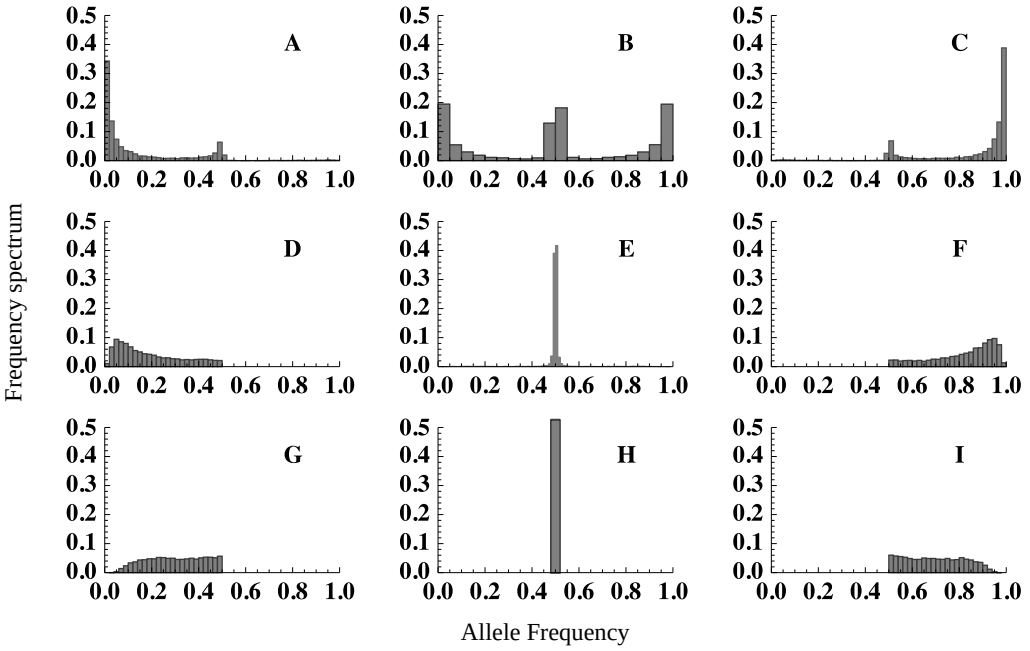

Figure S1: Allele frequency spectrum for different proportions of alleles of large effect, and for different positions of the optimum. Optimum value set at: 90% of the maximum trait value to the left (left column), close to zero (central column), 90% of the maximum trait value to the right (right column). Proportion of alleles of large to small effects is: 90% (top row), 50% (central row), 10% (lower row). Each plot is composed of the end points of 50 runs with random initial condition and realization of effects. $\mu = 10^{-4}$, $S = 0.1$, $n = 100$.

3 Response to selection

3.1 Response of traits composed of alleles of only large effects

By comparing the response of a trait composed only by alleles of large effect but otherwise evolving under identical conditions as another trait with mixed effects, we find that alleles of large effects are the most important driving the response to selection. Figure S2 is nearly identical to Fig. 8 in the main text, but without the alleles of small effects.

To have a more comprehensive demonstration, we proceed as in the previous situation. That is, we compare the response of two populations that are identical, except that in one the trait is composed only by the alleles of large effects. If we compute the absolute difference,

$$D = \int_0^T |\bar{z}_1 - \bar{z_2}| / \sqrt{(\nu_1 + \nu_2)/2}\, dt$$

we expect D to be very small except in the limit where $n_s >> n_f$, and close to the limit of response of the trait. Figure S3 shows the divergence between these trajectories for different realizations of allelic effects, which differ only on the number of alleles of small effect. The deviations that are large (e.g. more than two standard deviations) correspond to cases where the optimum is placed close to the maximum trait values.

Summarizing, alleles of large effect not only contribute most of the variation, but also drive the dynamics of the trait.

3.2 Response of traits under different distribution of effects

In the main text we showed that the number of alleles that can be near fixation depends on the distribution of allelic effects. Since these alleles of large effect are the ones that generate most of the genetic variance, the response to selection will be also affected by the details of the distribution of allelic effects. By integrating the amount of genetic variance that has been produced along an evolutionary run, that is $V_T = \int_0^T \nu dt$, we can measure how the distribution of allelic effects affects the response to selection. At the same time, this measure can be regarded as the effectiveness of directional selection towards

the optimum. Recall that $d\bar{z}/dt = \nu\beta$. Integrating in the time interval $[0, T]$, on both sides we have that $\Delta\bar{z} = \beta V_T \beta$, and consequently $\beta V_T = \Delta\bar{z}/\beta$. The last expression is the relative amount of displacement of the mean trait value per selection. In Fig. S4 we study how different distributions of allelic effects affect the total genetic variance V_T that is produced along a run, and compare it against the number of contributing loci n_f. The positive, power-law relationship indicates that selection works more efficiently by fixing alleles of large effect than by displacements of allele frequencies of alleles of small effect.

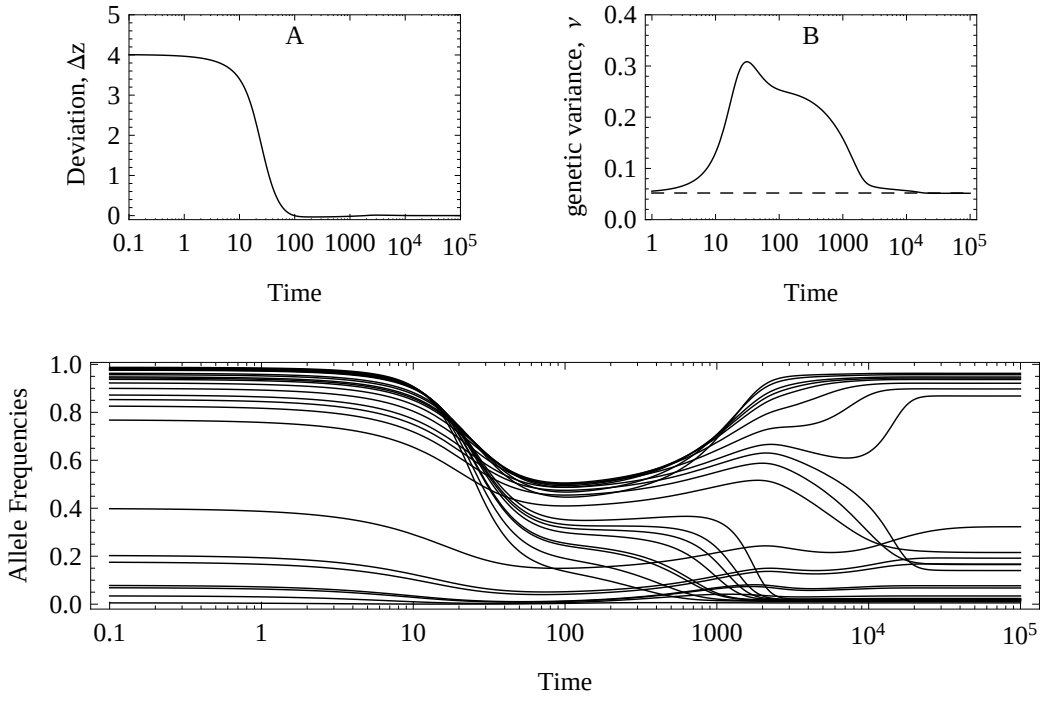

Figure S2: Response to an abrupt displacement of the optimum of a polygenic trait. (A) Deviation of the trait mean from the newly positioned optimum. (B) Genetic variance. Black: exact numerical results. Dashed black line: equilibrium genetic variance, ($\nu = 2n\mu/S$). (C) Response of the allele frequencies. The trait is composed by $n = 26$ all of large effects, distributed as an exponential of mean $= 1/10$. $\mu = 10^{-4}, S = 10^{-1}, z_\circ \simeq -2$.

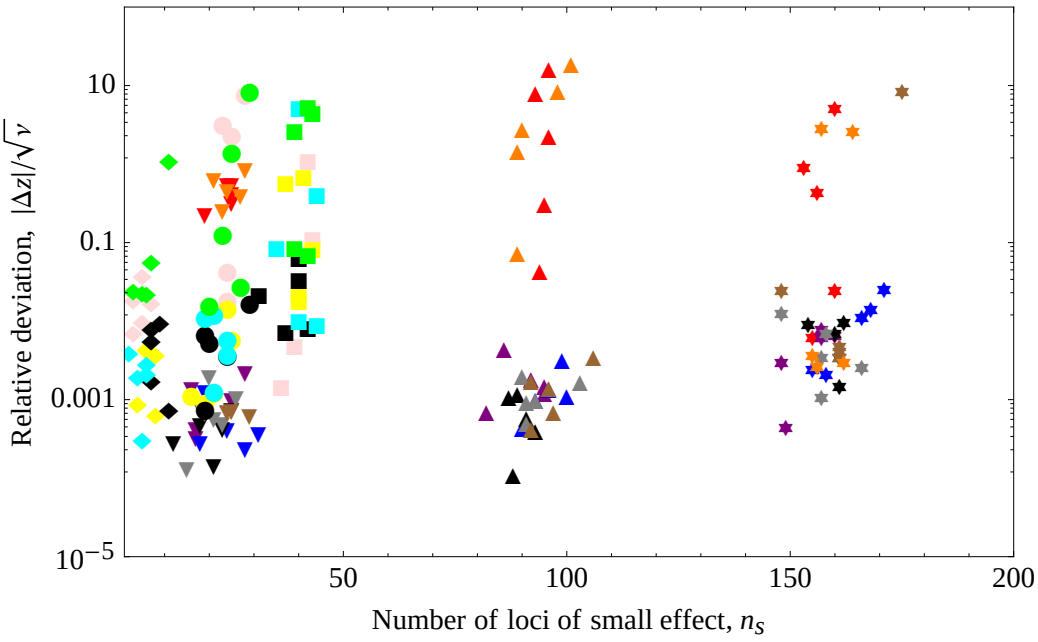

Figure S3: Deviations of the trajectories of traits composed by mixed effects and large effects, against the number of alleles of small effect. The allelic effects are drawn from a gamma distributions of mean 0.1 and shape parameter $\pi = 0.1$ (stars, squares), $\pi = 1$ (up-triangles, circles), $\pi = 10$ (down-triangles, diamonds). The position of the optima are color coded. For loci with $n = 50$ (stars and triangles) the maximum trait value is $z_x \simeq 5$, and the optima are placed at $z_\circ$ = -4 (pink), -2 (yellow), 0 (black), 2 (cyan), and 4 (green). For traits with $n = 200$ (squares, circles and diamonds) the maximum trait value is $z_x \simeq 20$, and the optima are placed at $z_\circ$ = -18 (red), -12 (purple), -6 (blue), 0 (black), 6 (gray), 12 (brown), 18 (orange). $\mu = 10^{-4}, S = 10^{-1}$

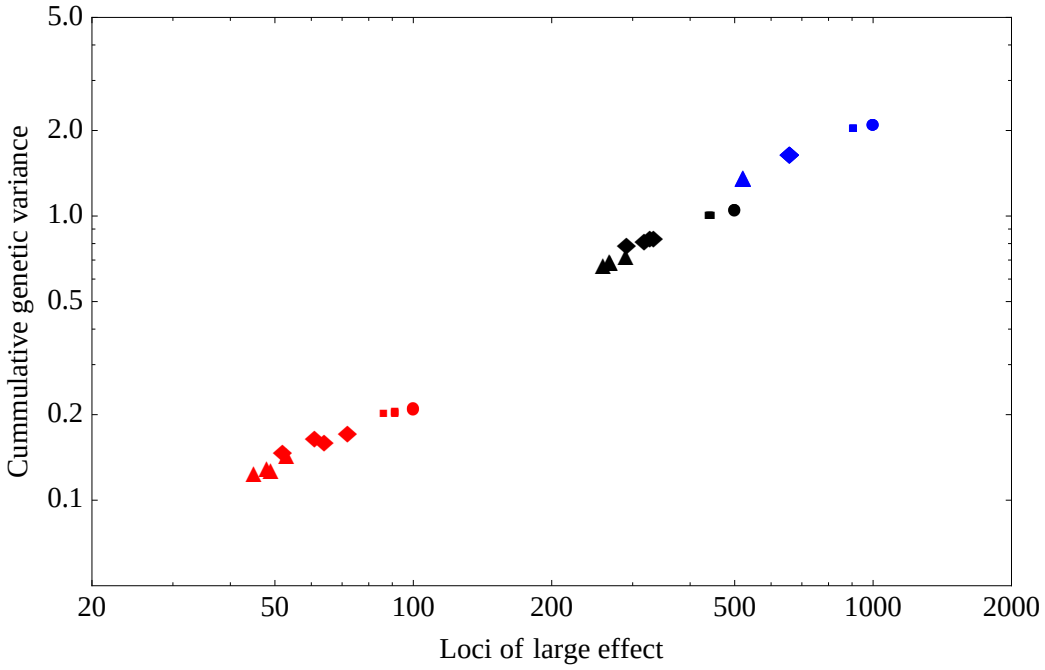

Figure S4: Cummulative genetic variance for different distributions of allelic effects and different amounts of loci. The effects are drawn from a gamma distribution of mean 0.1 and distinct shape parameters $\pi = 0.001$ (circles), $\pi = 0.01$ (squares), $\pi = 0.1$ (diamonds), $\pi = 1$ (triangles); The number of loci are $n = 1000$ (blue), $n = 500$ (black), n=100 (red).

4 Effective number of loci

The response of the trait is given by $d\bar{z}/dt = 2\nu S\Delta\Omega$. If n_e effective loci contribute to the variance, then by assuming the HoC we get that

$$\frac{d\bar{z}}{dt} = 4n_e\mu\Delta\Omega \ . \tag{19}$$

We saw that the alleles that first respond to a sudden shift in the optimum are those of effect close to $\hat{\gamma}$. Assuming no initial deviations from the optimum, and using Eq. 12 of the main text we get

$$\frac{d\bar{z}}{dt} = \Delta\Omega\left(4\mu n_f + S\sum_{k\in\mathbb{S}}\gamma_k^2\right). \tag{20}$$

Hence, in as long as there are some fixed alleles ($n_f > 0$), the two expressions of $d\bar{z}/dt$ equate to give

$$n_e = n_f\left(1 + \frac{\frac{1}{2}\sum_{k\in\mathbb{S}}\gamma_k^2}{2\mu n_f/S}\right). \tag{21}$$

In essence, this amounts to the variance of the HoC model but using an effective number of loci n_e.

A similar approximation using effective number of loci for the initial response of the genetic variance, indicates that ν remains roughly constant. Therefore the approximations can be regarded in a "breeder's equation regime", where the response to selection is sustained with constant genetic variance.

5 Moving optimum

5.1 Oscillating moving optimum

In this section, we study how a trait responds to an optimum that oscillates deterministically as:

$$z_\circ = \Omega \sin\left(2\pi \frac{t}{\tau}\right). \tag{22}$$

In Figures S5-S8 we show how the amplitude and the period of the fluctuations affect the evolution of the trait. We find that the trait remains very close to the optimum, and therefore the lag is small. An interesting feature is that the stationary lag shows fluctuations of different frequencies, which are more erratic as amplitude and frequency are increased.

5.2 Quasiperiodicity

Figure S9 shows that this is a pervasive behavior which seems to occur when the deviation from the optimum is large enough. In these plots, the allelic effects are the same but the initial peaks are slightly different. Although the trajectories seem to converge after a few thousand generations, they briefly separate, and, surprisingly they enter a phase of fluctuations, but which are entirely deterministic, since all initial conditions converge to that path (Fig. S9). Hence, we rule out chaotic behavior, because, if the trajectories were chaotic, they would diverge from each other, resembling random fluctuations that are uncorrelated across runs.

The fluctuations are entirely deterministic, and are composed by oscillations of many different frequencies. This is known as quasiperiodicity [1, Ch. 6]. Quasiperiodicity is associated with unpredictable physical systems such as weather and turbulence, and appears when there is an external forcing, as happens with a moving optimum in a population. However, although quasiperiodicity imply unpredictability of the evolutionary outcome, on the scale of response of the trait, these are minor fluctuations ($\sim .1 - .01\%$ of the value of the mean trait).

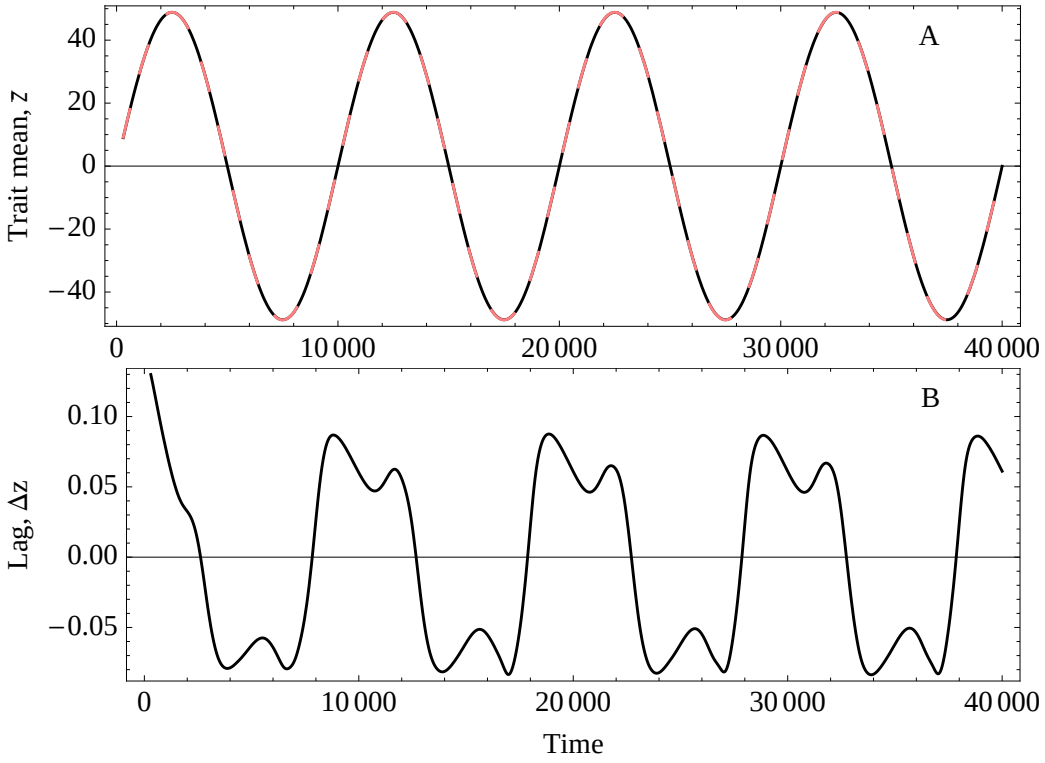

Figure S5: Response of a trait to an oscillatory moving optimum. (A) Moving optimum (red, dashed), trait mean (black). (B) Lag of the trait mean from the optimum. The optimum fluctuates according to Eq. 22 with amplitude $\Omega = 0.5 z_x$, and period $\tau = 10^4$. The trait is composed by $n = 1000$ loci with exponentially distributed effects with mean 0.1, and $z_x \simeq 100$. $S = 0.1, \mu = 10^{-4}$

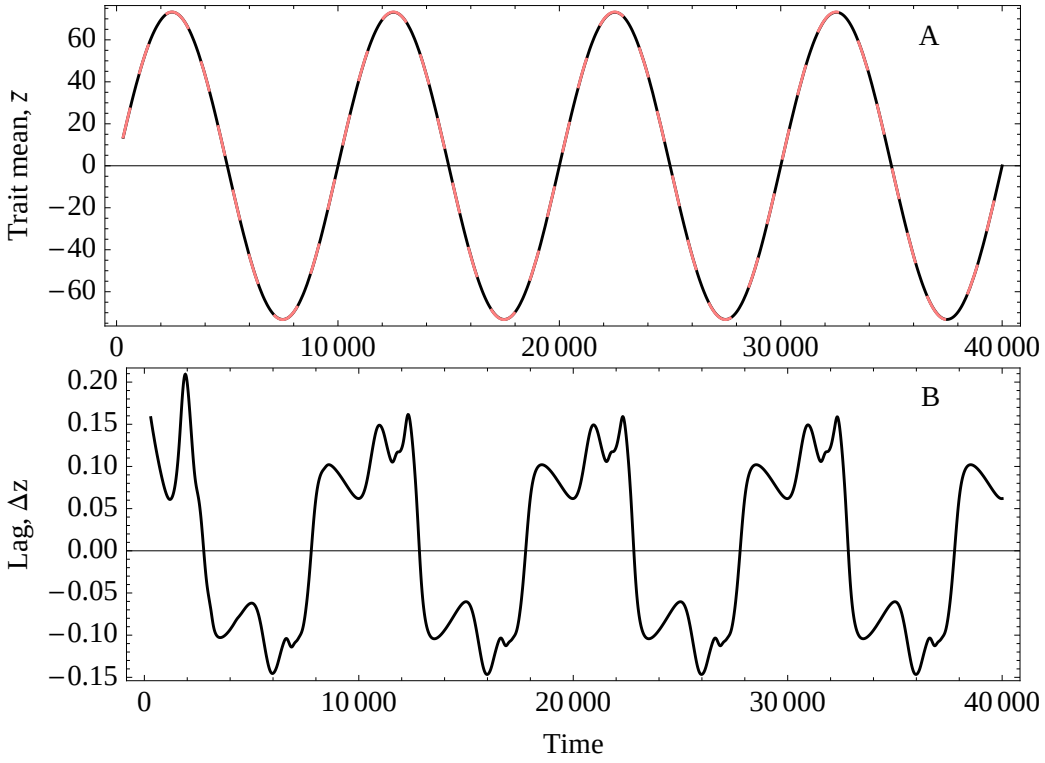

Figure S6: Response of a trait to an oscillatory moving optimum. The optimum fluctuates with amplitude $\Omega = 0.75 z_x$, and period $\tau = 10^4$. Otherwise as Fig. S5

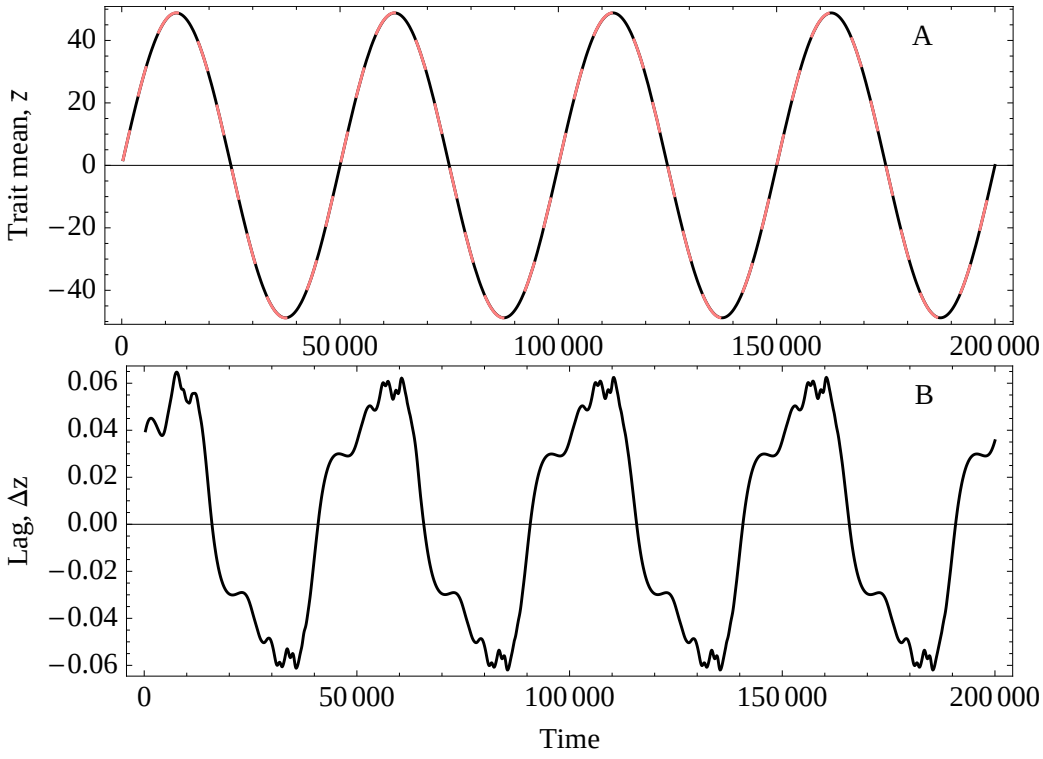

Figure S7: Response of a trait to an oscillatory moving optimum. The optimum fluctuates with amplitude $\Omega = 0.5 z_x$, and period $\tau = 5 \times 10^4$. Otherwise as Fig. S5

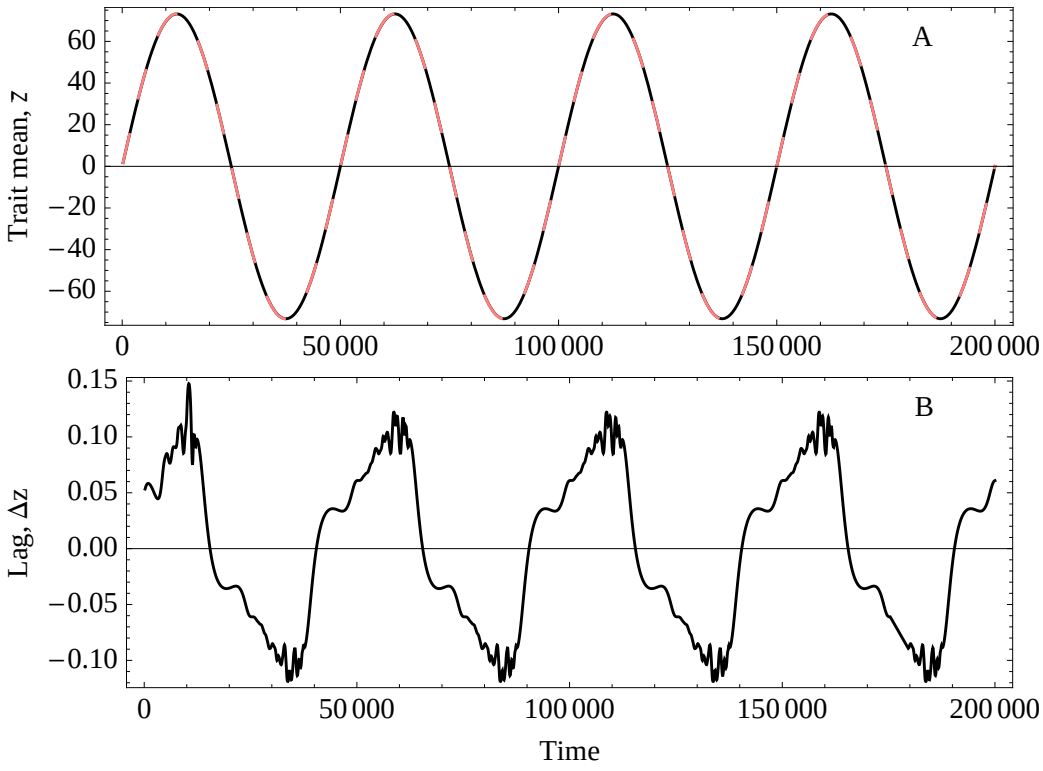

Figure S8: Response of a trait to an oscillatory moving optimum. The optimum fluctuates with amplitude $\Omega = 0.75 z_x$, and period $\tau = 5 \times 10^4$. Otherwise as Fig. S5

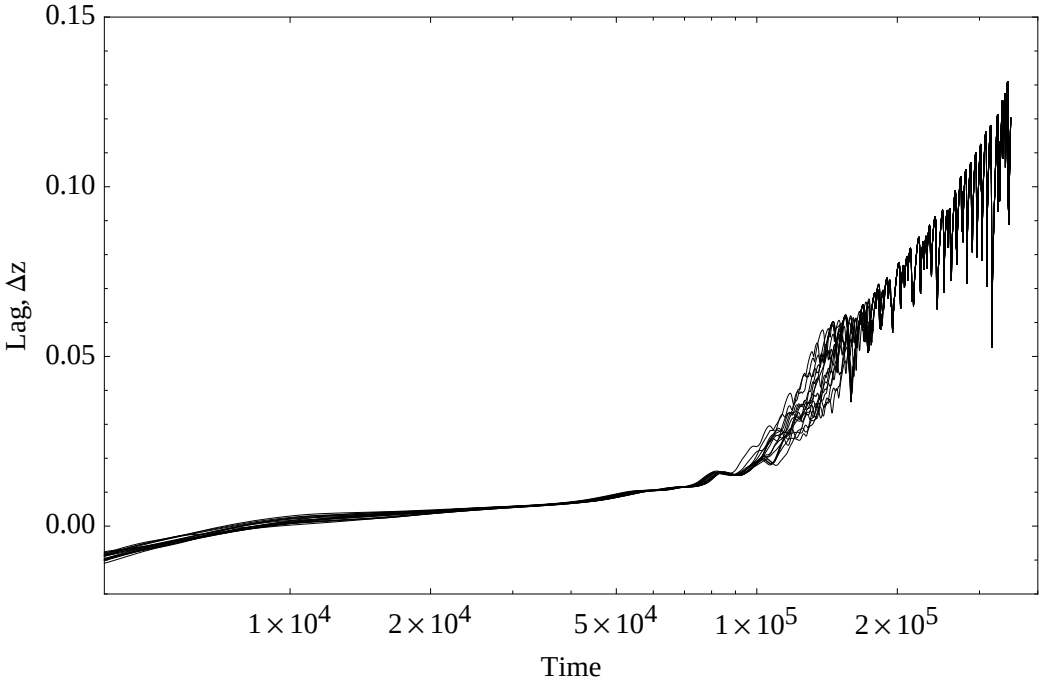

Figure S9: Long term response of a trait of unequal effects responding to a moving optimum. Sixteen different trajectories start at nearby adaptive peaks that are close to the initial optimum $\Omega = -23$. The trait is composed by $n = 500$ loci with unequal effects (exponentially distributed with mean 0.1), and in this case $n_s = 214$. $S = 0.1, \mu = 10^{-4}$.

5.3 Damped oscillating moving optimum

Now we consider an optimum that oscillates but its amplitude gradually dampens. That is

$$z_\circ = z_\circ = \Omega \sin\left(2\pi\frac{t}{\tau} + \phi\right) e^{-t/\tau}. \tag{23}$$

When $t >> \tau$ the optimum is a t zero and oscillations come to an end. Figure S10 shows an example of a population that evolves under a damped oscillating optimum. We also compare it to a linearly moving optimum that starts at the same state, and stops at a similar time. Fig. S10B shows that once equilibrium is achieved the oscillating optimum leads to better adaptation than the linearly moving optimum. By the moment when the optimum stops fluctuating or moving, the population responding to the oscillatory optimum ends up with more genetic variance than the one evolving under the linearly moving optimum. This is consistent with the final lag, since the population that has more genetic variation is able to find a peak with higher fitness (or lower load) than the population that has lower genetic variance and that cannot escape from a local optimum.

In Fig. S11 we present the end-point of simulations similar as the one in Fig. S10, but where we randomize the initial conditions, the effects, and vary the number of loci that compose the trait. As in the particular example of Fig. S10 we find that the linear moving optimum leads populations to larger deviations (up to an order of magnitude) than damped oscillating moving optimum. We also find that the final lag is smaller with less loci. The genetic variance, is also smaller for the linear moving optimum, although only slightly.

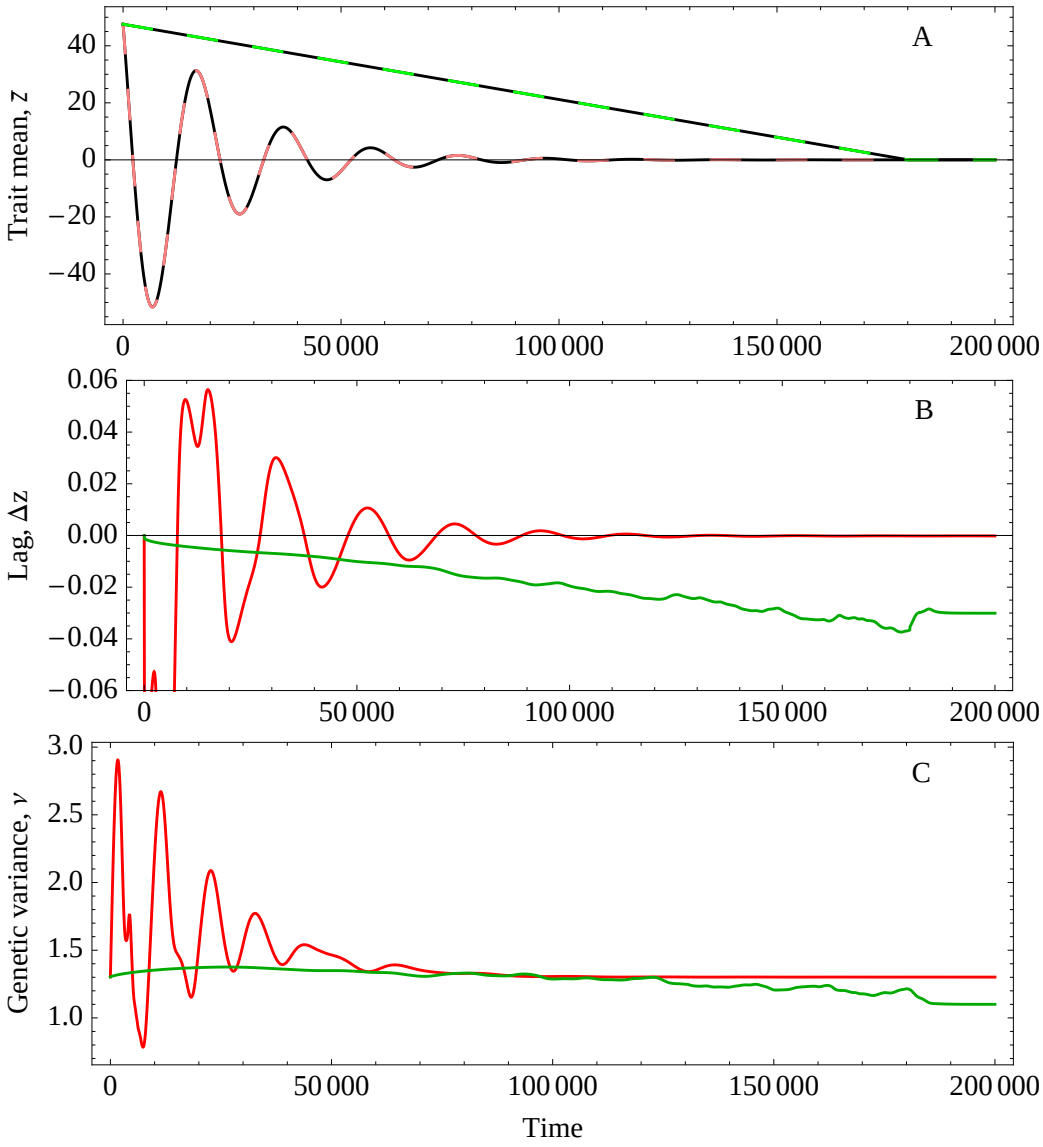

Figure S10: Polygenic trait following a moving optimum with damped oscillations (red) compared to a linearly moving optimum (green). (A) Response of a polygenic trait. (B) Stationary lag. (C) Response of the genetic variance. The oscillatory optimum has a period of $\tau = 2 \times 10^4$, and an amplitude of $\Omega = 3/4 z_x$. The phase is set so that the population initially stands at the optimum peak with no deviations from the optimum. The linearly moving optimum starts at the same point and moves to $z_\circ = 0$ in $T = 1.8 \times 10^5$ time units. Otherwise as Fig. S5.

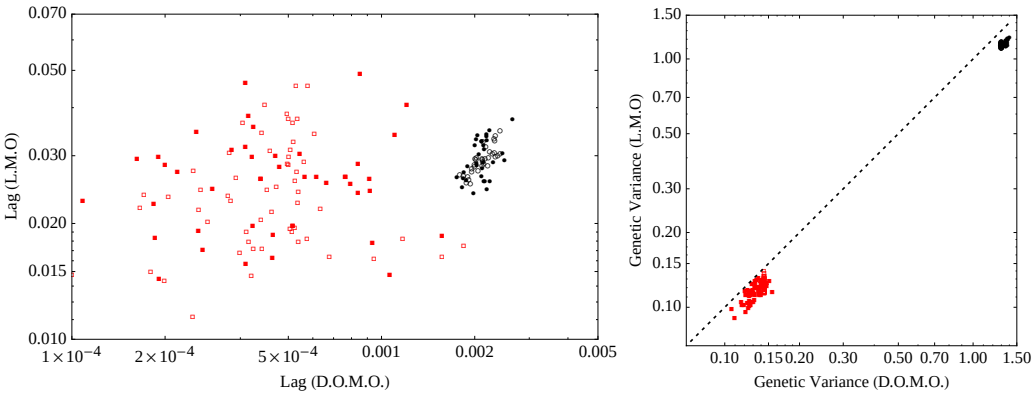

Figure S11: End point of polygenic trait following a moving optimum with damped oscillations (D.O.M.O.)compared to a linearly moving optimum (L.M.O). Each point compares (A) the final lag and (B) the genetic variance of populations that differ only on the way the get to equilibrium. The dotted line is the identity (for visual reference). Circles $n = 1000$, squares $n = 100$. Open symbols represent randomization of the initial peaks but same realization of effects. Solid symbols randomize both initial conditions and effect. The realization of effects used for $n = 1000$ is the same as in Figs S5-S8. All effects are sampled from an exponential distribution with mean 0.1. Otherwise as Fig. S10.

6 Admixed populations and genetic incompatibilities.

Assume that two populations that are genetically distinct, although with the same mean trait and genetic variance come into contact. The new, admixed, frequencies at each locus i are $p_i^{ad} = \alpha p_i^{(1)} + (1-\alpha)p_i^{(2)}$, where the superscripts denote the source population, and α is the degree of admixture. Since the trait is additive, the admixed mean trait is the weighted average $\bar{z}^{ad} = \alpha \bar{z}^{(1)} + (1-\alpha)\bar{z}^{(2)}$. The genetic variance, however, is not additive, thus:

$$\nu^{ad} = \alpha^2 \nu^{(1)} + (1-\alpha)^2 \nu^{(2)} + 2\alpha(1-\alpha) \sum_{i=1}^{n} \gamma_i^2 \left(p_i^{(1)} q_i^{(2)} + p_i^{(2)} q_i^{(1)} \right). \quad (24)$$

Necessarily, n_s and n_f are the same in both populations, and thus have the same genetic variance, $\nu^{(1)} = \nu^{(2)} = \nu$. By distinguishing those alleles that are at the same frequency in both populations, and those that have different frequencies across populations we get:

$$\nu^{ad} = \nu + 2\alpha(1-\alpha) \left(\sum_{i \in \mathbb{I}} \gamma_i^2 - 2\tilde{n} \frac{\mu}{S} \right) \quad (25)$$

where the set $\mathbb{I} = \left\{ i : p_i^{(1)} = q_i^{(2)} \right\}$ has $\tilde{n}$ loci. Notice that $\tilde{n}$ is the number of loci that have contrary alleles close to fixation in each population, i.e. $p_i^{(1)} = q_i^{(2)}$. Since $\gamma_i^2 > 2\mu/S$, then $\nu^{ad} > \nu$ the admixed population, has larger genetic variance, and is unfit with respect to any of the source populations (independently of the degree of admixture).

References

[1] Ott, E., 1993 *Chaos in dynamical systems*. Cambridge University Press.