

# Selective Sweep at a Quantitative Trait Locus in the Presence of Background Genetic Variation

Luis-Miguel Chevin<sup>\*,†,1</sup> and Frédéric Hospital<sup>‡</sup>

<sup>\*</sup>UMR de Génétique Végétale, Ferme du Moulon, 91190 Gif sur Yvette, France, <sup>†</sup>Ecologie, Systématique et Evolution, UMR 8079, Université Paris-Sud, 91405 Orsay Cedex, France and <sup>‡</sup>INRA, UMR1236 Génétique et Diversité Animales, 78352 Jouy-en-Josas, France

Manuscript received July 4, 2008

Accepted for publication September 17, 2008

## ABSTRACT

We model selection at a locus affecting a quantitative trait (QTL) in the presence of genetic variance due to other loci. The dynamics at the QTL are related to the initial genotypic value and to the background genetic variance of the trait, assuming that background genetic values are normally distributed, under three different forms of selection on the trait. Approximate dynamics are derived under the assumption of small mutation effect. For similar strengths of selection on the trait (*i.e.*, gradient of directional selection  $\beta$ ) the way background variation affects the dynamics at the QTL critically depends on the shape of the fitness function. It generally causes the strength of selection on the QTL to decrease with time. The resulting neutral heterozygosity pattern resembles that of a selective sweep with a constant selection coefficient corresponding to the early conditions. The signature of selection may also be blurred by mutation and recombination in the later part of the sweep. We also study the race between the QTL and its genetic background toward a new optimum and find the conditions for a complete sweep. Overall, our results suggest that phenotypic traits exhibiting clear-cut molecular signatures of selection may represent a biased subset of all adaptive traits.

THE recent improvements in methods to detect positive selection from its molecular signature on neutral polymorphism (NIELSEN 2005) and the great amount of information thus generated provide an unprecedented opportunity for evolutionary biologists to improve their understanding of the genetics of adaptation and of the recent adaptive history of species, in particular in humans (NIELSEN *et al.* 2007).

On the one hand, genome scans search for recent (GLINKA *et al.* 2003; AKEY *et al.* 2004; NIELSEN *et al.* 2005; WILLIAMSON *et al.* 2007) or ongoing (VOIGHT *et al.* 2006; TANG *et al.* 2007) selection by genotyping many markers distributed throughout the genome and using the properties of the polymorphism pattern (heterozygosity, frequency spectrum) to reject neutrality either through a parametric model-based approach or by just picking outliers in the distribution (for the caveats of the latter method, see TESHIMA *et al.* 2006). This kind of study typically reveals many positive results. The next step is then to search databases for the functions of the identified genes (VOIGHT *et al.* 2006; TANG *et al.* 2007), to go from the genotype to the phenotype, *i.e.*, to answer the question “Which phenotypic traits were recently adaptive in the lineage leading to the species or population under study?” Yet, the relationship between the strength of selection on a gene and the selection on the trait is

most often not explicitly defined in those studies. Clearly, a gene that was under strong selection must have affected a trait that was itself strongly selected. Nevertheless, not all traits that were recently under strong directional selection necessarily show strong signatures of selection at the gene level. If there were some systematic biases, it would be useful to identify and quantify them, to be able to interpret molecular signatures of selection in terms of phenotypic selection.

On the other hand, hitchhiking mapping methods focus on smaller candidate regions (SCHLOTTERER 2003) that were previously identified either by functional genetics or by genome scans (THORNTON *et al.* 2007). The aim is here to confirm recent positive selection in the region, as well as to localize more precisely the target of selection (KIM and STEPHAN 2002). The selection coefficient may also be estimated on the basis of the pattern of either the frequency spectrum (KIM and STEPHAN 2002) or the heterozygosity (WANG *et al.* 1999; SCHLENKE and BEGUN 2004; OLSEN *et al.* 2006) in the genomic region. The selection coefficient thus estimated is a summary of the dynamics of the allele under selection, but it has no clear biological meaning; it measures only the propensity of the allele to grow in frequency. Yet in biological reality a mutation affects a phenotypic trait, and the trait is subject to selection. To better understand the genetics of adaptation, we would need to be able to interpret the selection coefficient estimated at the gene level in terms of parameters of selection at the trait level, especially in the context of

<sup>1</sup>Corresponding author: UMR 8079, Bât. 360, Université Paris-Sud, 91405 Orsay Cedex, France. E-mail: luis-miguel.chevin@u-psud.fr

complex quantitative traits (FALCONER and MACKAY 1996; LYNCH and WALSH 1998).

There are two main views of how adaptation occurs as a genetic mechanism. The first one, periodic selection (ATWOOD *et al.* 1951), considers that adaptation proceeds by the successive fixations of beneficial mutations that sweep through the population one after the other. This view is supported by examples of experimental evolution with microbes (ELENA *et al.* 1996; ELENA and LENSKI 2003). In microbes, the large population sizes make purifying selection very efficient, which strongly limits the amount of slightly deleterious polymorphisms [the so-called nearly neutral mutations (OHTA 1992)] that can be maintained by drift. As a consequence, phenotypic traits have little genetic variance in such organisms, except the variance generated anew by mutation. Moreover, asexuality induces clonal interference, which is expected to prevent simultaneous segregation at several loci (GERRISH and LENSKI 1998). Note, however, that the process of clonal interference is partly challenged by recent theories of “travelling waves” that take into account the possibility that several beneficial mutations occur in the same lineage (DESAI and FISHER 2007; DESAI *et al.* 2007), so periodic selection may not apply to asexual populations under a high beneficial mutation rate.

The second view of adaptation is that of quantitative genetics, in which there is simultaneous selection at many loci that contribute to adaptive traits (polygenic selection). Such a vision is based on studies of the response to selection (FALCONER and MACKAY 1996) and of the maintenance of genetic variation (BARTON and KEIGHTLEY 2002). It is more relevant when linkage is loose, such that selective interferences between loci are low—allowing for simultaneous sweeps at many loci—and when selection is not very efficient, so that phenotypic traits can accumulate substantial genetic variation. Therefore it is intended for sexual organisms of reasonably low population sizes. At its most extreme, this view leads to the infinitesimal model designed by FISHER (1918), in which many loci of small effect contribute to a trait, such that the genetic values are normally distributed, an approach that has led to many fascinating developments in evolutionary quantitative genetics (WALSH 2007). In practice, the periodic selection and quantitative genetics views are probably two extremes of a continuum, and their main interest is that they provide fairly good approximations of more complex real genetic systems under certain conditions.

It is striking that the theory behind signatures of positive selection, namely genetic hitchhiking (MAYNARD SMITH and HAIGH 1974), is clearly one of periodic selection, whereas such signatures are searched in sexual species (since hitchhiking mapping is dependent on the recombination rate) that often exhibit substantial genetic variation for many traits. For such species (essentially plants or animals), the quantitative genetic

approach is widely used and recognized as efficient to predict the response to selection, at least in the short term, for cultivated as well as natural populations (FALCONER and MACKAY 1996).

Here, our aim is to extend the theory of hitchhiking to the context of a locus affecting a quantitative trait that also harbors background genetic variation due to other loci. Our hope is to help draw clearer conclusions regarding selection on adaptive traits based on molecular signatures of selection. For instance, is it just the strength of phenotypic selection that matters or are there other parameters that affect the outcome of a selective sweep? If so, then what types of traits are more likely to exhibit molecular signatures of selection? When the selection coefficient of a gene can be estimated, what can be inferred about the strength of selection on the trait affected by that gene?

First, we recall the classic theory of hitchhiking and the way that selection affects the neutral polymorphism in such a context. Then, we use a model of selection on a quantitative trait controlled by one focal locus and a distribution of background genetic values—contributed to by many other loci—to calculate the trajectory of a beneficial mutation in such a case. We show that the strength of selection on a quantitative trait locus and its signature on neutral polymorphism can be strongly affected by the background genetics of the trait and that the outcome critically depends on the shape of the fitness function.

## MODEL

**The classical hitchhiking model:** We first summarize the classical model of genetic hitchhiking as initially developed by MAYNARD SMITH and HAIGH (1974) and further studied by STEPHAN *et al.* (1992) and BARTON (1998). Consider a diploid, randomly mating population and two biallelic loci  $A$  and  $B$  with recombination rate  $r$  between them. The first locus, with alleles  $A_1$  and  $A_2$  in respective frequencies  $p$  and  $q = 1 - p$ , is under additive positive selection, such that the relative fitnesses of the genotypes  $A_2A_2$ ,  $A_2A_1$ , and  $A_1A_1$  are 1,  $1 + s$ , and  $1 + 2s$ , respectively. The other locus is neutral, with two alleles  $B_1$  and  $B_2$  in frequencies  $u$  and  $1 - u$ , respectively. In all the following, the subscript 0 denotes initial conditions.

At the selected locus, the change in the frequency  $p$  of allele  $A_1$  in one generation is

$$\Delta p = \frac{spq}{1 + 2ps}. \quad (1)$$

Since this expression is dependent on the mean fitness of the population  $\bar{W} = 1 + 2ps$ , it is more convenient to write the recursion using the ratio  $\rho = p/q$  of allelic frequencies, which yields

$$\rho' = \rho \left( 1 + s - s^2 \frac{p}{1 + ps} \right), \tag{2}$$

where the prime denotes the value in the next generation and, for small  $s$ ,

$$\rho \approx \rho_0(1 + s)^t \approx \rho_0 e^{st}. \tag{3}$$

Hence, the selection model used in hitchhiking theory (as well as in many standard population genetic models) considers an exponential increase of the ratio of allelic frequencies at the selected locus. This is also equivalent to a logistic growth of the allele frequency  $p$ .

It can be shown easily that the change in the frequency of allele  $B_I$  at the neutral locus is

$$\Delta u = \frac{sD}{1 + 2ps} = \frac{D}{pq} \Delta p, \tag{4}$$

where  $D$  is the linkage disequilibrium between loci  $A$  and  $B$ , defined as  $f(A_I B_I) - pu$  [where  $f(A_I B_I)$  is the frequency of the  $A_I B_I$  haplotype]. The quantity  $D/pq$  is a measure of the statistic association between the loci, from the perspective of the selected locus. Using indicator variables to denote the presence of allele  $A_I$  at locus  $A$  and of allele  $B_I$  at locus  $B$ ,  $D$  is homologous to the covariance between loci  $A$  and  $B$ , and  $D/pq$  is homologous to the coefficient of regression (covariance divided by the variance of one of the variables) of locus  $B$  on locus  $A$ . It can be shown (see, e.g., BARTON 2000) that  $D/pq$  is also equal to the difference between the frequencies of allele  $B_I$  in the selected ( $A_I$ ) and in the unselected ( $A_2$ ) genetic backgrounds at locus  $A$  and that it changes only because of recombination, regardless of the strength of selection, such that

$$\frac{D}{pq} = \frac{D_0}{p_0 q_0} (1 - r)^t \approx \frac{D_0}{p_0 q_0} e^{-rt}. \tag{5}$$

Assuming  $s \ll 1$ , the changes in frequencies are well approximated by continuous-time processes, and the total change in frequency at the neutral locus is

$$\Delta u_{\text{tot}} = \frac{D_0}{p_0 q_0} \int_{p_0=\varepsilon}^{p_{\text{fix}}=1-\varepsilon} e^{-rt(p)} dp, \tag{6}$$

which is equivalent to Equation 1 (second line) in BARTON (2000). The function  $t(p)$  is the inverse dynamics of the mutation  $A_I$ , i.e., the time needed for this mutation to reach a given frequency  $p$ , starting from a threshold frequency  $\varepsilon$  defined below. Note that the argument leading to (6) is purely deterministic, in that it does not take into account changes in frequencies resulting from genetic drift. It is a well-known result of the hitchhiking literature that the trajectory of a beneficial mutation in a finite population is very well approx-

imated by its deterministic expectation in an infinite population, provided the frequency  $p$  is sufficiently distant from the absorbing edges 0 and 1; i.e.,  $p \in [\varepsilon, 1 - \varepsilon]$  (see, e.g., STEPHAN *et al.* 1992 or BARTON 1998). However, at very low ( $p < \varepsilon$ ) or very high ( $p > 1 - \varepsilon$ ) frequencies, the trajectory of a beneficial mutation is mainly controlled by genetic drift. The threshold frequency  $\varepsilon$  is thus defined as the frequency above which the deterministic effect of selection overwhelms the stochastic effect of drift on the trajectory of the beneficial mutation in time. Using  $\varepsilon$  instead of  $1/(2N_e)$  as the starting frequency of the mutation thus allows us to partially account for stochasticity in a deterministic framework: it acts as a “filter” to focus only on the mutations that can indeed reach high frequencies. In practice, simulation results show that a value of  $\varepsilon = 1/(4N_e s)$  performs well (KIM and STEPHAN 2002). Note also that conditional on final fixation, a beneficial mutation reaches  $\varepsilon$  rapidly, so it is a good approximation to assume that no recombination occurred in this time lapse (BARTON 1998).

The initial association  $D_0/p_0 q_0$  depends on the starting conditions. When the mutation  $A_I$  appears in a single copy, it can be associated either with  $B_I$  (with probability  $u_0$ ), leading to  $D_0/p_0 q_0 \approx 1 - u_0$ , or with  $B_2$  (with probability  $1 - u_0$ ), leading to  $D_0/p_0 q_0 \approx -u_0$ . Combining these two events, the expected total reduction of heterozygosity at the neutral locus is, from Equation 6,

$$R_{\text{HI}} = E \left( \frac{2(u_0 + \Delta u_{\text{tot}})(1 - (u_0 + \Delta u_{\text{tot}}))}{2u_0(1 - u_0)} \right) = 1 - \left( \int_{\varepsilon}^{1-\varepsilon} e^{-rt(p)} dp \right)^2, \tag{7}$$

where  $E$  denotes the expectation over all possible starting conditions. This equation can be understood from the standpoint of the set of haplotypes carrying the beneficial allele  $A_I$ . At the beginning,  $A_I$  starts in one copy, and the haplotype that carries it harbors no genetic diversity at other loci. During the selective sweep, diversity is introduced on these haplotypes through recombination, and the total amount of regained diversity depends on the trajectory in time of the selective sweep, that is, on the dynamics of  $A_I$ . Note that formulated this way, Equation 7 is quite general and applies whenever the dynamics of the beneficial mutation are sufficiently slow that it can be approximated by a continuous process. Equation 7 indicates that the full trajectory of a beneficial mutation is sufficient and necessary information for predicting its hitchhiking effect. In the simple case studied by MAYNARD SMITH and HAIGH (1974), the reverse dynamics is, from Equation 3,

$$t(p) = \frac{1}{s} \log \left( \frac{(1 - \varepsilon)p}{\varepsilon(1 - p)} \right), \tag{8}$$

which yields for the reduction of heterozygosity

$$\begin{aligned}
R_H &= 1 - \left( \int_{\varepsilon}^{1-\varepsilon} \left( \frac{(1-\varepsilon)p}{\varepsilon(1-p)} \right)^{-r/s} dp \right)^2 \\
&= 1 - \left( \frac{\varepsilon}{1-\varepsilon} \right)^{2r/s} \\
&\quad \times \left( B\left(1-\varepsilon; 1-\frac{r}{s}, 1+\frac{r}{s}\right) - B\left(\varepsilon; 1-\frac{r}{s}, 1+\frac{r}{s}\right) \right)^2,
\end{aligned} \tag{9}$$

where  $B$  is Euler's incomplete beta function. Assuming  $r \ll s$ ,  $s \leq 1$ , and  $\varepsilon \leq 1$ , the reduction of heterozygosity can be approximated by

$$R_H \approx 1 - \varepsilon^{2r/s}, \tag{10}$$

which corresponds the simplest approximation proposed in STEPHAN *et al.* (1992).

In this article, contrary to the usual approach for selective sweeps (MAYNARD SMITH and HAIGH 1974; STEPHAN *et al.* 1992), we do not define a constant selection coefficient *a priori* for  $A_I$ . Instead, we focus on a mutation affecting a quantitative trait and define selective pressure at the level of the trait only. Our aim is to describe the dynamics of the beneficial mutation in the presence of background genetic variation in a form similar to Equations 2 and 3. We thus wish to identify the factors affecting the dynamics of the mutation and to assess the range of parameter values for which our model converges to the classical hitchhiking model. We also wish to characterize the molecular signature left on neutral polymorphism by selection at a quantitative trait locus, to understand what information is provided by genome scans for traces of selection or hitchhiking mapping. We mainly use a deterministic argument as in MAYNARD SMITH and HAIGH (1974), since it allows us to characterize major trends analytically, but our results may be good approximations for stochastic populations when the frequency of the beneficial mutation is neither very low nor very high. We also discuss the qualitative and quantitative changes that may be introduced in a stochastic framework.

**Lande's model—simultaneous selection on a focal gene and on background variation:** We use the model proposed by LANDE (1983) to describe the effect of selection on a focal gene affecting a quantitative trait and on background genetic variation for this trait contributed by other loci. Our aim is to predict the complete trajectory of a beneficial mutation starting in one copy at the focal locus (*i.e.*, of a hard selective sweep) as a function of the genetic and selective parameters of the trait. For the focal locus we use the same notations as in the previous section, with the exception that allele  $A_I$  now has an additive effect  $a$  on the trait, and no selection coefficient is defined *a priori*. LANDE's (1983) model assumes normally distributed genetic values in the background, with mean  $m$  and variance  $\sigma^2$ , which is standard practice in quantitative genetics (FALCONER and MACKAY 1996). It also assumes a large population size (that is, the

model is deterministic), no linkage, and no interaction of the focal locus with the genetic background, such that the distribution of genetic values within each genotypic class at locus  $A$  is also normal, with the same variance  $\sigma^2$  as in the entire population, and the mean shifted by 0,  $a$ , or  $2a$  for genotypes  $A_2A_2$ ,  $A_1A_2$ , or  $A_1A_1$ , respectively. This latter assumption is a good approximation as long as the number of individuals in each genotypic class is not too small ( $>50$ , see DISCUSSION). The absolute fitness of a phenotype with value  $z$  is  $W(z)$ , and we neglect environmental variance for the sake of clarity, so the phenotypic and genotypic values are identical.

LANDE (1983) showed that in this context, since the distribution of genetic values is a linear combination of identical normal distributions, the change in the mean genetic background value  $m$  of the trait follows the same equation as in the case of a single normal distribution. Namely, the change in  $m$  in one generation is

$$\Delta m = \sigma^2 \beta, \tag{11}$$

where  $\beta = \partial \log(\bar{W}) / \partial m = (1/\bar{W})(\partial \bar{W} / \partial m)$  is the gradient of directional selection on the trait (LANDE 1976), which measures the slope of the log mean fitness landscape at the position of the population. Under random mating, the mean fitness of the population is simply

$$\bar{W} = p^2 \bar{W}_{A_1A_1} + 2pq \bar{W}_{A_1A_2} + q^2 \bar{W}_{A_2A_2}, \tag{12}$$

where  $\bar{W}_{A_iA_j}$  is the mean fitness across all genetic backgrounds of individuals with genotype  $A_iA_j$  at locus  $A$ . The change in frequency of  $A_I$  at the focal locus can be described using WRIGHT's (1969) equation:

$$\Delta p = \frac{pq}{2} \frac{1}{\bar{W}} \frac{\partial \bar{W}}{\partial p}. \tag{13}$$

Note that  $\partial \bar{W} / \partial p = 2s$  in the classical model described in the previous section, so Equation 1 is a particular case of Equation 13. LANDE's (1983) model allows describing the changes in frequencies at a gene in a polygenic context in a simple way (without the use of complex multigenic recursions), by combining results from quantitative and population genetics. Note that as long as the normal approximation holds, any locus in the background could be the focal locus. In our context, however, the focal locus is chosen to be one where a new mutation has been introduced in one copy very recently.

In this article, we want to study the dynamics at the quantitative trait locus  $A$  under contrasted forms of selection on the trait. Indeed, there is no *a priori* reason to focus on one specific type of function. Studies aiming at estimating the shapes of fitness functions showed that those can vary substantially between traits (SCHLUTER 1988). Moreover, simply assessing the relative prevalence of directional *vs.* stabilizing or disruptive selection in the wild is often difficult because of statistical issues

(KINGSOLVER *et al.* 2001). We thus chose to study three simple cases that could induce adaptation and to use them to compare the dynamics of selective sweeps under markedly distinct selective pressures on the trait. We used a linear directional fitness function ( $W_l$ ), an exponential fitness function ( $W_e$ ), and a Gaussian stabilizing selection ( $W_g$ ). In all cases, a parameter  $\omega > 0$  quantifies the strength of selection, such that

$$\begin{aligned} W_l(z) &= b + \omega z \text{ if } z \geq 0, & W_l(z) &= b \text{ if } z \leq 0 \\ W_e(z) &= e^{\omega z} \\ W_g(z) &= e^{-(z^2/2\omega^2)}. \end{aligned} \tag{14}$$

In the Gaussian fitness function,  $z$  is taken to be the distance to the genetic optimum, without loss of generality. Note that  $\omega$  measures the width of the bell for the Gaussian fitness function, so the intensity of selection increases with  $1/\omega$  instead of  $\omega$  for this fitness function (contrary to the other two). This notation was used for the sake of homogeneity with standard quantitative genetic literature. These three examples of fitness functions are not meant to be completely realistic, but rather to encompass three clearly different shapes, which together can well approximate many real fitness functions in the vicinity of the current state of the population (see DISCUSSION).

**A measure of selection on a gene based on its dynamics:** The quantity used to describe the dynamics of the mutation is the growth rate  $s$  of the ratio of allelic frequencies  $\rho = p/q$ ; that is,  $s = \rho'/\rho - 1$ . This approach stems from FISHER (1930), who underlined the interest of quantifying selection by measuring changes in the ratio of allelic frequencies rather than in the frequencies themselves, using an argument of geometric increase. Note that in the limit where selection can be approximated by a continuous process,  $s = (1/\rho)(dp/dt) = d \log(\rho)/dt$ , as in FISHER (1930, p. 34). The term  $s$  does not change with the mean fitness of the population, so it is often used in experimental evolution to measure selection on genes on the basis of allele frequency change (LENSKI *et al.* 1991; PERFEITO *et al.* 2007). In the simple case presented in the first section, it also equals  $\frac{1}{2}(\partial \bar{W}/\partial p) = s$ , but it is not always so. In the general case,  $s$  captures the influence of parameters on the dynamics of the beneficial mutation more accurately than  $\frac{1}{2}(\partial \bar{W}/\partial p)$ , as we will see later. We calculated the change in frequency  $\Delta p$  using Equation 13 and then we calculated  $s$  as

$$s = \frac{p + \Delta p}{1 - (p + \Delta p)} \frac{1 - p}{p} - 1. \tag{15}$$

The full expression of  $s$  depends on the details of the genetic architecture of the trait as well as on the type of selection acting on this trait. We then used  $s$  to derive the complete dynamics of beneficial mutations and then to calculate the expected signature of selection on

neutral polymorphism. We focused on selection on new mutations (*i.e.*, on “hard” selective sweeps) and not on “soft” sweeps where a beneficial mutation that was already segregating in the population starts being selected after a change in the environment (INNAN and KIM 2004; HERMISSON and PENNINGS 2005; PRZEWORSKI *et al.* 2005). Soft sweeps are expected to leave weaker and more complex signatures on neutral polymorphism.

## RESULTS

In the following, we derive the equations for the dynamics of a mutation affecting a quantitative trait under distinct types of selection, focusing on hard selective sweeps, *i.e.*, on beneficial mutations that start in one copy and eventually reach fixation. Subscripts l, e, and g are used throughout to refer to results obtained under linear, exponential, and Gaussian fitness functions, respectively. We also use an asterisk to denote approximated results under the assumption that the effect of the mutation is small relative to the mean value of the trait; that is, that  $|a| \ll |m|$ .

**Dynamics without background variation:** We first describe the dynamics of the mutation in the absence of background genetic variation ( $\sigma^2 = 0$ ). The general derivations are presented in the APPENDIX. The growth rate of the focal mutation in the absence of background genetic variation is approximately

$$\begin{aligned} s_{l,\sigma^2=0}^* &= \frac{a\omega}{b + m\omega} \\ s_{e,\sigma^2=0}^* &= s_{e,\sigma^2=0} \approx a\omega \\ s_{g,\sigma^2=0}^* &= e^{-(am/\omega^2)} - 1 \approx -\frac{am}{\omega^2}. \end{aligned} \tag{16}$$

It can already be seen from Equation 16 that the way selection on a trait translates into selection on a gene affecting this trait crucially depends on the shape of the fitness function, which was already discussed in KIMURA and CROW (1978). Under an exponential fitness function, the selective pressure on the mutation does not depend on the mean background genetic value  $m$ , which was emphasized in LANDE (1983). In contrast, under linear and Gaussian fitness functions, selection on the mutation depends not only on its genetic effect  $a$  and on parameters of the selection function ( $b, \omega$ ), but also on the present mean genetic state of the population ( $m$ ). Note that the influence of parameters on the dynamics of the quantitative trait locus is captured accurately by the term  $s$  that we use here, but not by other definitions of the selection coefficient. For instance, using  $\frac{1}{2}(\partial \bar{W}/\partial p)$  instead of  $s$  (by identifying Equations 13 and 1), one could think that the mean genetic value  $m$  influences selection at locus  $A$  even in the case of exponential fitness function, whereas recursions show that it is actually not the case (not shown).

Our focus here is on hard selective sweeps, *i.e.*, on new beneficial mutations that reach fixation. Under linear and exponential fitness functions, all mutations with  $a > 0$  can sweep to fixation. Under the Gaussian fitness function, adaptation obviously takes place only if the population is not at the optimum, that is, if  $m \neq 0$ . This can occur after a recent change in the environment, which rapidly shifted the optimal genetic value of the trait. If so, a mutation can reach fixation if it allows the population to get closer to the optimum, that is, if  $|m + 2a| < |m|$ . This is equivalent to stating that  $(m + 2a)^2 < m^2$ , which leads to the condition  $a(a + m) < 0$ . Note that if the mutation effect is small relative to the mean genetic value, the only condition is that  $a$  and  $m$  have opposite signs, as can be seen directly from  $s_{g,\sigma^2=0}^*$  (Equation 16).

In the absence of background genetic variation, and assuming  $|a| \ll |m|$ ,  $s_{\sigma^2=0}^*$  does not change in time. We can then write, in a form similar to Equation 3:

$$\rho = \rho_0(1 + s_{\sigma^2=0}^*)^t \approx \rho_0 e^{s_{\sigma^2=0}^* t}. \quad (17)$$

In this context, the parameter  $s_{\sigma^2=0}^*$  is equal to the  $s$  defined in the classical hitchhiking model and can be estimated by hitchhiking mapping. Moreover, using the definition of the directional selection gradient in Equation 11 and using the same formalism as previously, we note that

$$s_{\sigma^2=0}^* = \beta_{\sigma^2=0}^* a, \quad (18)$$

for all three fitness functions. This means that, for a given strength of selection (*i.e.*, gradient of directional selection) on an adaptive trait and in the absence of background genetic variance, the strength of selection on a weak-effect mutation affecting this trait is proportional to its genetic value for the trait.

**Dynamics with background variation:** *Changes in the frequency of the mutation and in the mean genetic background in one generation:* In the presence of background genetic variation for the trait ( $\sigma^2 > 0$ ), the expressions for the growth rate of  $\rho$  in one generation remain unchanged under linear and exponential selection. Under Gaussian stabilizing selection and assuming  $|a| \ll |m|$ , the expression for  $s_g$  with background genetic variation is approximately (introducing  $\eta = 1 + \sigma^2/\omega^2$ )

$$s_g^* = e^{-(am)/(\sigma^2 + \omega^2)} - 1 \approx -\frac{am}{\sigma^2 + \omega^2} = \frac{s_{g,\sigma^2=0}^*}{\eta}. \quad (19)$$

Hence, among our chosen fitness functions, background genetic variance has a direct effect on the change of frequency at the focal locus only in the case of Gaussian stabilizing selection. In this case, the growth rate of the ratio  $\rho$  of frequencies at the selected locus in one generation decreases with increasing  $\sigma^2$ . With either exponential or linear selection, background genetic variance has no direct effect on the change in frequency at

the focal locus. Nevertheless, the genetic variance does affect the dynamics of the mean genetic background value, which in turn influences selection at the focal locus under linear as well as under Gaussian selection. In contrast, selection at the focal locus is never affected by the genetic background of the trait under exponential selection.

The change in the mean value of the trait in one generation is given by Equation 11, where the exact expressions for the directional selection gradients are given in the APPENDIX. Under the small-effect approximation, these selection gradients are

$$\begin{aligned} \beta_1^* &= \frac{\omega}{b + m\omega} \\ \beta_e^* &= \beta_e = \omega \\ \beta_g^* &= -\frac{m}{\sigma^2 + \omega^2}. \end{aligned} \quad (20)$$

Note that under this small-effect assumption, we still find, for any generation, that  $s^* = \beta^* a$ , as in the absence of background genetic variation (Equation 18). Using a similar small-effect approximation, KIMURA and CROW (1978) derived a formula that is close to the latter, except that they defined the selection coefficients as  $\Delta p/p$  and the effect of the gene as  $(1 - p)a$  (after rearranging and reformulating with our notations). Therefore their prediction was that the change in the frequency of  $A_1$  is  $\Delta p = \beta^* a p q$ . This is not strictly equivalent to stating that  $s^* = \beta^* a$ , which proves more useful to calculate the complete trajectory of the mutation as we will see below. Note that their result can also be obtained in our case by noting that from Equation 13,  $\partial \bar{W}/\partial p = (\partial \bar{W}/\partial z)(\partial z/\partial p) \approx (\partial \bar{W}/\partial m)(\partial z/\partial p) = 2\beta a$ .

Importantly, the approximate directional selection gradients in Equation 20 are independent of the frequency of the mutation at the focal locus, contrary to the exact expressions presented in the APPENDIX. This uncouples the dynamics of the gene from those of the mean genetic background, which allows calculating first the trajectory in time of the mean genetic background  $m$  and then using it to derive the trajectory of the focal mutation, using Equations 16 and 19.

*Complete trajectory for a mutation of small effect:* To calculate the trajectory of the beneficial mutation  $A_1$ , we first need to know the complete dynamics of the trait. To obtain it, we assume that the background genetic variance remains constant over time, as in LANDE (1983). This assumption may not be very realistic on the longer term (TURELLI 1988). However, under the infinitesimal model (*i.e.*, for a background composed of many loci with small effects as assumed here) it remains a reasonable approximation over the shorter term (*i.e.*, in the beginning of the sweep), and it allows a treatment of the problem that may remain reasonably robust when  $\sigma^2$  changes through time. In any case, considering a constant variance may be a conservative assumption as

to the effects of background variation on the dynamics of the focal locus, as we shall see in the DISCUSSION. We also assume that  $|a| \ll |m|$  all along the selective sweep, even in the case of Gaussian stabilizing selection, although as the population approaches the optimum,  $|m|$  may ultimately decrease to zero. We assess the conditions that allow a selective sweep at the focal locus to complete under stabilizing selection in the next section.

Using Equations 11 and 20 and under the assumptions above, the dynamics of the mean genetic background for all three fitness functions are

$$\begin{aligned} m_l^*(t) &= \frac{\sqrt{(b + m_0\omega)^2 + 2\sigma^2\omega^2 t} - b}{\omega} \\ m_e^*(t) &= m_0 + \omega\sigma^2 t \\ m_g^*(t) &= m_0\eta^{-t}, \end{aligned} \tag{21}$$

where time 0 is taken when the mutation  $A_l$  is in the threshold frequency  $\varepsilon$  defined in Equation 6 and  $\eta = 1 + \sigma^2/\omega^2$  as previously. The results for Gaussian and exponential functions [ $m_e^*(t)$  and  $m_g^*(t)$ ] are exact, whereas the one for linear selection [ $m_l^*(t)$ ] is based on a continuous-time approximation. With a linear fitness function, the mean genetic value of the trait increases as a square-root function of time; that is, it increases more and more slowly with time. This can also be seen from the gradient of directional selection (Equation 20), which takes  $m$  in its denominator. This stems from a well-known property of linear fitness functions, which generate negative epistasis for fitness among mutations in the same direction (see, *e.g.*, TENAILLON *et al.* 2007); here, as  $m$  increases, there is less and less advantage in increasing it any further. Under exponential selection, the mean value of the trait increases linearly with time, as a consequence of the constant gradient of selection. Finally, under Gaussian stabilizing selection, the absolute distance to the optimum decreases exponentially with time, at rate  $\sigma^2/\omega^2$ .

Knowing  $m(t)$ , the growth rate of the ratio of allelic frequencies can be expressed as a function of time by combining Equations 20 and 21 and by recalling that  $s^* = \beta^*a$ , which leads to

$$\begin{aligned} S_l^*(t) &= \frac{a\omega}{\sqrt{(b + m_0\omega)^2 + 2\sigma^2\omega^2 t}} \\ S_e^*(t) &= a\omega \\ S_g^*(t) &= -\frac{am_0}{\omega^2} \left(1 + \frac{\sigma^2}{\omega^2}\right)^{-(t+1)}. \end{aligned} \tag{22}$$

This shows that the effect of background genetic variation on the dynamics of a locus affecting an adaptive quantitative trait crucially depends on the type of fitness function that governs selection on this trait. Under exponential selection, the growth rate of the mutation is constant and independent of background genetic variation. Under Gaussian stabilizing selection,

it decreases exponentially with time and with the genetic variance  $\sigma^2$ . Finally, if the fitness function is linear, it has a more complicated decreasing dynamics, proportional to  $1/\sqrt{t}$ .

In our model, as we assume no linkage and no epistasis, there is no covariance between the focal locus and the genetic background, so the part of variance explained by the locus is simply  $2pqa^2/(\sigma^2 + 2pqa^2)$ . This term cannot be easily related to the dynamics of the focal mutation; for instance, it does not affect it at all under exponential fitness function. Therefore, the weight of a given QTL for a selected trait (defined as the proportion of the total variance explained by the QTL) is not necessarily a strong determinant of the signature of selection that it will show at the molecular level, and the absolute genetic value  $a$  of the QTL may be more informative in that respect. Nevertheless, the weight of the QTL may also be correlated to some extent to the strength of the signature of selection since it is a growing function of  $a$ .

From Equation 22, the full trajectory of the mutation at the focal locus can be calculated using, for  $t > 0$ ,

$$\begin{aligned} \rho^*(t) &= \frac{\varepsilon}{1 - \varepsilon} \prod_{k=0}^{t-1} (1 + s^*(k)) \approx \frac{\varepsilon}{1 - \varepsilon} \exp(S(t)), \\ S(t) &= \sum_{k=0}^{t-1} s^*(k). \end{aligned} \tag{23}$$

The term  $S(t)$  is the cumulative growth rate of the mutation  $A_l$ , that is, the total amount of increase of  $\log(\rho)$  at time  $t$ , resulting from the action of selection over all generations since the frequency of the  $A_l$  allele became superior to  $\varepsilon$ . For each type of fitness function (and assuming that the dynamics for the linear fitness function can be approximated by a continuous-time process, *i.e.*, replacing the sum with an integral), the cumulated growth rate  $S(t)$  is

$$\begin{aligned} S_l(t) &= \frac{a}{\beta_{1,0}^*\sigma^2} \left( \sqrt{1 + 2\beta_{1,0}^{*2}\sigma^2(t-1)} - 1 \right) \\ S_e(t) &= a\omega(t-1) \\ S_g(t) &= -\frac{am_0}{\sigma^2} (1 - \eta^{-t}), \quad \sigma > 0, \end{aligned} \tag{24}$$

where  $\beta_{1,0}^* = \omega/(b + m_0\omega)$  is the initial gradient of directional selection under the linear fitness function. These expressions are compared to the product  $s(t-1)$  with a constant selection coefficient  $s$ . The dynamics in Equations 23 and 24 can then be easily translated into that of the frequency of the beneficial mutation by noting that  $p^*(t) = \rho^*(t)/(1 + \rho^*(t))$ , which leads to

$$p^*(t) = \frac{\varepsilon}{\varepsilon + (1 - \varepsilon)\exp(-S(t))}. \tag{25}$$

Figure 1 shows the approximate and exact dynamics of the beneficial mutation under a linear fitness

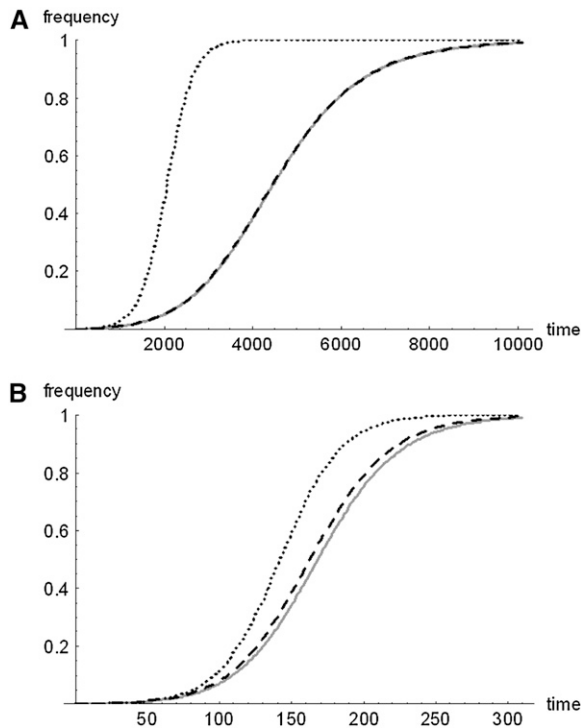


FIGURE 1.—Dynamics at the QTL under linear directional selection. Exact recursions of Equations 11 and 13 (shaded line) are compared to approximate dynamics using Equations 24 and 25 (dashed line) and to approximate dynamics without background genetic variance ( $\sigma^2 = 0$ , dotted line). The assumption  $|a| \ll |m|$  was either valid from the beginning (A) ( $m_0 = 10$ ,  $a = 0.1$ ,  $b = 1$ ,  $\omega = 0.05$ ,  $\sigma = 1$ ,  $\varepsilon = 0.001$ ) or initially violated (B) ( $m_0 = 0.1$ ,  $a = 1$ ,  $b = 1$ ,  $\omega = 0.05$ ,  $\sigma = 1$ ,  $\varepsilon = 0.001$ ). The measuring unit is arbitrary and is scaled to  $\omega$ .

function. The presence of standing variation can substantially affect these dynamics. For instance, in Figure 1A, the time to fixation (defined by the frequency  $1 - \varepsilon$ ) is tripled compared to the situation without background variation. Even if  $|a| > |m|$  at the start of the selective sweep,  $m$  increases with time so that the weak-effect approximation (assuming  $a \ll m$ ) gets better as the selective sweep proceeds. On the contrary, under stabilizing selection, the approximation performs well only when  $m_0$  is substantially greater (in absolute value) than the effect of the mutation,  $a$  (Figure 2A). Indeed, if  $|a|$  is initially close to  $|m|$ , the approximation will worsen as  $|m|$  decreases under selection for an optimum (Figure 2B). Note, however, that even in cases where the approximation performs poorly, it still describes the dynamics more accurately than assuming no background genetic variation at all.

*Conditions for a complete selective sweep under stabilizing selection:* The results above were obtained assuming  $|a| \ll |m|$  all along the selective sweep. However, under stabilizing selection,  $m$  tends to 0 as the trait approaches the optimum, so this assumption may be violated in the course of the sweep. Furthermore, if the mean genetic value approaches the optimum too quickly because of

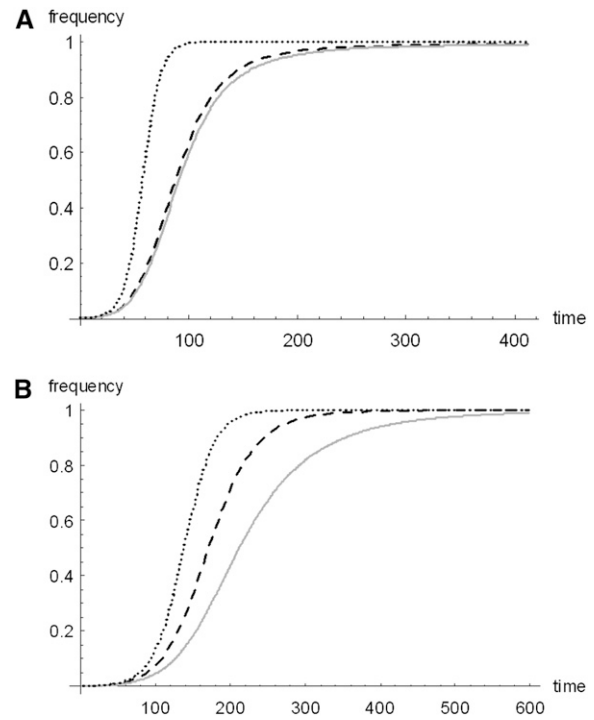


FIGURE 2.—Dynamics at the QTL under Gaussian stabilizing selection. Exact recursions of Equations 11 and 13 (shaded line) are shown together with approximate expected dynamics using Equations 24 and 25 (dashed line) and expected dynamics without background genetic variance (dotted line). The situation where the assumption that  $|a| \ll |m|$  is met from the beginning ( $m_0 = 2$ ,  $a = -0.06$ ,  $\omega = 1$ ,  $\sigma = 0.1$ ,  $\varepsilon = 0.001$ ) is shown in A, and that where it is initially violated ( $m_0 = 0.5$ ,  $a = -0.1$ ,  $\omega = 1$ ,  $\sigma = 0.05$ ,  $\varepsilon = 0.001$ ) is shown in B. Note that  $|a|$  cannot be very large relative to  $|m_0|$  in the presence of background genetic variation, not to overshoot the optimum.

the background genetic variation, the mutation at the focal locus will obviously become deleterious and eventually disappear from the population. To put it another way, mutations starting in higher frequencies (pooled here in the genetic background) will be more likely to reach fixation first—thus reducing the distance to phenotypic optimum—and to prevent the spread of new mutations (represented by the focal locus). Hence there are necessary conditions on the parameters of the system that allow the possibility of a complete hard sweep at the focal locus, that, is a beneficial mutation that starts in one copy and reaches fixation. Unfortunately, the range of parameters that defines these conditions is also the one for which the assumption  $|a| \ll |m|$  is not valid, so there is interdependency between the dynamics at the focal locus and those of the mean genetic background, which prevents us from finding an exact solution. Nevertheless, a criterion for fixation can be built from the approximated system of Equations 21 and 24, and its accuracy can be tested with numerical examples. This criterion needs to be conservative (in the sense that it must lead to a parameter range that



bounds the actual one) and informative enough (the obtained range must not be too large relative to the actual range). On the basis of numerical results under various parameter values, the criterion that we used was that the frequency  $p$  of the mutation reaches  $\frac{1}{2}$  before the population gets to a distance  $2a$  of the optimum, that is, before  $a(m + 2a)$  becomes positive. Using the simplified system in Equations 21 and 24 and assuming  $\varepsilon \ll 1$ , this leads to

$$\sigma^2 \leq -\frac{m_0^2}{8 \log(\varepsilon)} \tag{26a}$$

$$\begin{aligned} a < -\frac{m_0}{4} (1 - \sqrt{1 + 8(\sigma^2/m_0^2) \log(\varepsilon)}) & \text{ if } m_0 > 0 \\ a > -\frac{m_0}{4} (1 + \sqrt{1 + 8(\sigma^2/m_0^2) \log(\varepsilon)}) & \text{ if } m_0 < 0, \end{aligned} \tag{26b}$$

where  $m_0$  is the initial distance to the optimum. Hence, there is a maximum value of the background genetic variance  $\sigma^2$  above which the mutation cannot reach fixation, however strong its own effect may be, because the population reaches the optimum too quickly as a consequence of the response to selection by the background. Importantly, this maximum value does not depend on the intensity of stabilizing selection on the trait, quantified by  $1/\omega^2$ . This is because  $\omega$  affects both the dynamics of the QTL and those of the mean background genetic value in the same manner and hence does not influence their competition. The maximum variance that allows a complete selective sweep at the QTL depends only on the ratio of the squared initial distance  $m_0$  over the frequency  $\varepsilon$  that determines the beginning of the quasi-deterministic phase of the sweep. If  $\sigma^2$  is below this maximum value, the effect  $a$  of the focal mutation must still be above a given threshold (in absolute value) for the mutation to be able to spread to fixation in the population. This threshold depends on the ratio of the background variance over the squared initial distance to the optimum, as well as on  $\varepsilon$ , but is again independent of  $\omega$ . Obviously, the initial conditions must also verify the condition  $a(a + m_0) < 0$  mentioned earlier in the absence of background variation. Figure 3A shows an example of the threshold value for  $a$  as a function of  $\sigma$ . Note that this threshold tends to infinity as  $\sigma$  approaches its maximal value. A selective sweep at the focal mutation can be completed only if the background genetic variation is very low and if the mutation effect is of the same order of magnitude as the background genetic standard deviation, that is, if the mutation has a very strong effect on the trait compared to other loci. Since the chosen criterion was defined *ad hoc*, we used numerical recursions to test whether it was close to the actual conditions for a complete selective sweep in the context of our model. The conditions

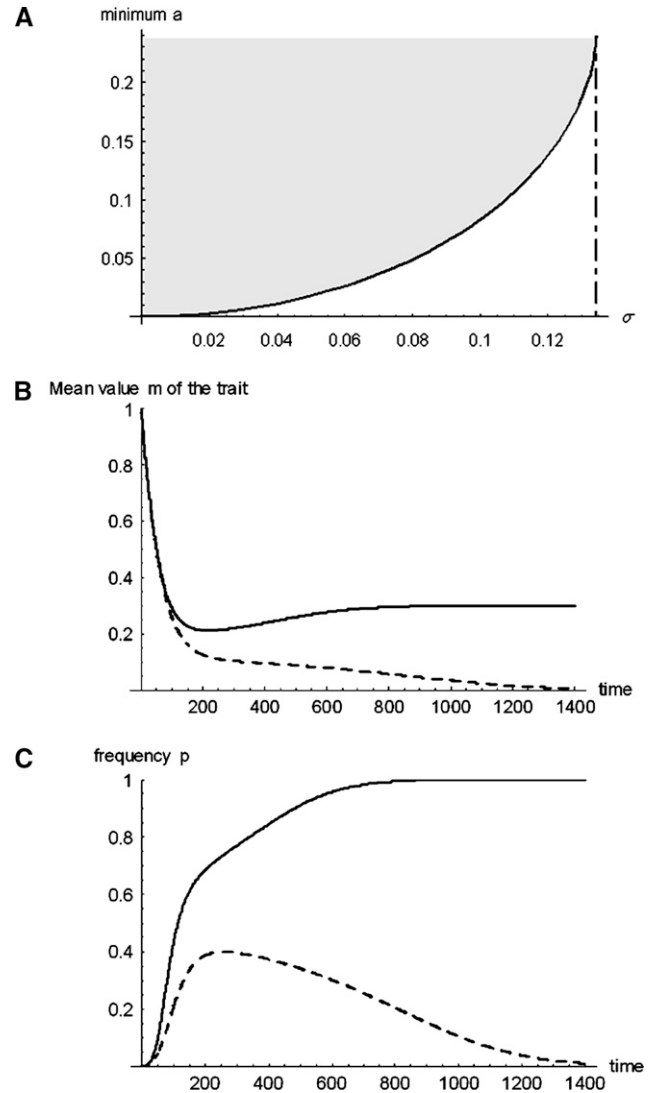


FIGURE 3.—Conditions for a complete “hard” sweep under Gaussian stabilizing selection. (A) Expected values of  $a$  (shaded area) that allow the focal mutation at the QTL to fix in the population as a function of the background standard deviation  $\sigma$  (parameters  $m_0 = 1$ ,  $\omega = 1$ ,  $\varepsilon = 0.001$ ). The dotted-dashed line is the maximum  $\sigma$  that allows fixation of the mutation, from Equation 26. Dynamics of (B) the mean value of the trait (arbitrary measuring unit, scaled to  $\omega$ ) and (C) the frequency  $p$  of the focal mutation  $A_I$  near the limits of the range of parameters defined in Equation 26 are shown. Parameters are as in A, with  $\sigma = 0.12$  and  $a = -0.15$  (solid line) or  $a = -0.12$  (dashed line). Expected threshold from Equation 26:  $a = -0.14$ .

obtained under our criterion were in good agreement with the actual behavior of the system, as shown in Figure 3, B and C. When  $|a|$  is below the threshold value defined in Equations 26, the frequency  $p$  of the focal mutation does not reach  $\frac{1}{2}$  and eventually decreases back to zero as the mean background genetic value reaches the optimum. In contrast, when the effect of the mutation is slightly above the threshold (in absolute value), the mutation manages to exceed a frequency of  $\frac{1}{2}$  and eventually reaches fixation (despite a slowing down of its

dynamics), whereas the mean genetic background goes back down, to the value  $m - 2a$ . Hence the background genetic variance of a trait has a strong negative impact on the probability of a complete selective sweep at a QTL under stabilizing selection, since it allows the population to adapt and reach its optimum without using any new mutation. Importantly, by assuming a constant background variance as we did in our simple model we likely underestimated the impact of background genetic variance on the focal locus (see DISCUSSION), so our results (Equations 26) are conservative in that respect.

Note also that these results are based on a deterministic argument. In a finite population, when the mutation  $A_I$  is in a very low number of copies, its frequency will also fluctuate stochastically, so it is very likely to be lost by genetic drift. This stochastic sieve has been studied thoroughly, and the probability of fixation of mutations can be calculated under various selective contexts (CROW and KIMURA 1970; EWENS 2004), including that where there is competition between beneficial mutations at several loci (BARTON 1995; GERRISH and LENSKI 1998). Here, we assume that the stochastic sieve has been passed successfully (*i.e.*,  $p > \varepsilon$ ), as mentioned earlier; hence this section focuses specifically on selective sweeps that are stopped because the optimum phenotype was reached ( $m = 0$ ) through selection at other loci, regardless of any stochastic effects. Stochastic effects should even worsen the situation for the focal locus, as we develop in the DISCUSSION.

**Signatures of selection:** As shown in Equation 7, the expected reduction of heterozygosity  $R_H$  for a neutral locus located at a recombination distance  $r$  from a locus under positive selection can be found analytically by calculating the integral  $I = \int_{\varepsilon}^{1-\varepsilon} e^{-rt(p)} dp$ . The obtained pattern of neutral polymorphism can then be compared to the one provoked by a selective sweep of constant  $s$ , thus mimicking the empirical approach where Equation 7 is fitted to an empirical polymorphism pattern to infer the selection coefficient  $s$  (assumed constant). This estimated  $s$  can be thought of as an “effective selection coefficient”  $s_e$  for the hitchhiking effect in a context of varying selection coefficient, *i.e.*, the constant value of  $s$  that would lead to the same signature of selection as the one observed. Using the approximated Equation 10, this leads to

$$s_e = r \frac{\log(\varepsilon)}{\log(I)}. \quad (27)$$

Unfortunately, when replacing  $t(p)$  with the inverse dynamics obtained from Equation 7, the resulting integral  $I$  cannot be solved. Alternatively, we can numerically compute the increase of polymorphism through recombination on the haplotype that carries the beneficial mutation. This is done by using the exact or approximated dynamics for the selected locus (Equations 24 and 25) and then converting them into the

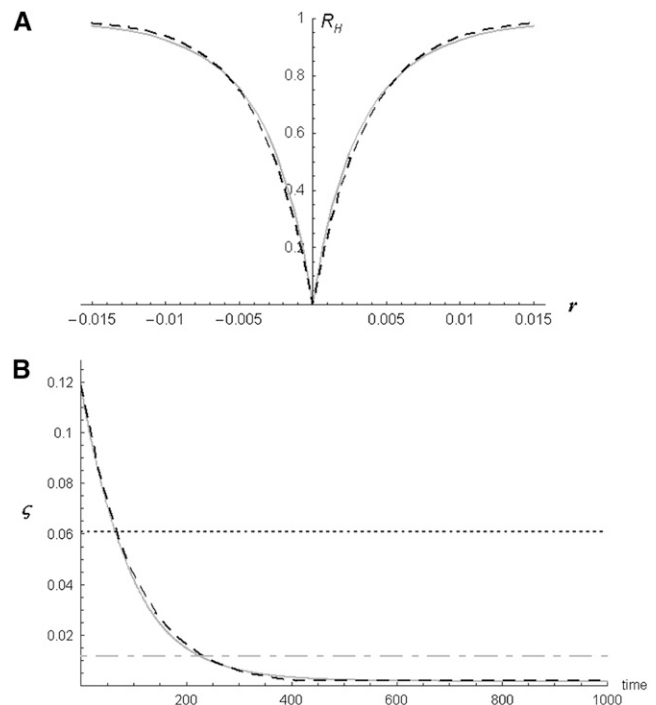


FIGURE 4.—Effective selection coefficient  $s_e$  for the hitchhiking effect, under Gaussian stabilizing selection. (A) Expected reduction of heterozygosity  $R_H$  plotted against the recombination rate with the locus under selection, in a selective sweep at a QTL with background genetic variation, under stabilizing selection. Expected pattern (shaded line) and fitted pattern under the assumption of a constant selection coefficient are shown. (B) Actual and estimated selection coefficients. Exact recursion (shaded line) and approximate value using Equation 22 (dashed line) for the growth rate  $\varsigma$  of the mutation are shown. The dotted line denotes the effective selection coefficient estimated by fitting the approximated formula in Equation 10 to the expected heterozygosity pattern. The dotted-dashed shaded line is the mean selection coefficient estimated by the composite-likelihood method of KIM and STEPHAN (2002) on 100 coalescence simulations of selective sweep with a decreasing selection coefficient following Equation 24. Parameters are the same as in Figure 2A, except  $\varepsilon = 1/(4N_e s_0)$  with  $N_e = 10,000$  and  $s_0 = -am_0/(\sigma^2 + \omega^2)$ .

change in frequency at the neutral locus using Equation 4. The final polymorphism pattern can then be compared to an expected pattern assuming a constant selection coefficient, for instance, by fitting Equation 10 to the heterozygosity pattern. Figure 4 shows an example of such an estimation of the effective selection coefficient under Gaussian stabilizing selection (based on Equation 27). The actual and fitted polymorphism curves are shown in Figure 4A. The pattern of reduction of heterozygosity  $R_H$  with a changing selection coefficient (implied by the Gaussian stabilizing selection with background genetic variance) is quite similar to that expected under a constant selection coefficient. As seen in Figure 4B, the actual selection coefficient decreases exponentially with time, as is also apparent from Equa-

tion 24. The selection coefficient  $s_e$  estimated through the reduction of heterozygosity corresponds to the  $s$  in the early phase of the selective sweep, as expected. Indeed, the hitchhiking effect is mostly concentrated in the beginning of a selective sweep, when the mutation is in low frequency (BARTON 1998). Yet this “effective” selection coefficient is substantially lower than the initial selection coefficient (about half of its value in our example).

Since the above results neglect the mutation events at the neutral locus during the selective sweep, we also ran coalescence simulations of a selective sweep with a decreasing selection coefficient. To do so, we modified the program “ssw” by Yuseob Kim (available online at <http://yuseobkim.net/YuseobPrograms.html>), by replacing the expected trajectory of the beneficial mutation with the approximate dynamics of Equations 24 and 25. We then applied KIM and STEPHAN’S (2002) composite-likelihood-ratio test to estimate the selection coefficient involved in the selective sweep. This method uses the site-frequency spectrum (SFS) of neutral mutations (*i.e.*, the proportion of mutations that are found at each frequency in a sample) to infer parameters of a selective sweep together with the (composite) likelihood of selection *vs.* neutrality. As seen in Figure 4B, the selection coefficient thus estimated is far below the  $s_e$  predicted through the reduction of heterozygosity and neglecting mutation during the selective sweep. This is because a decreasing selection coefficient induces the mutation  $A_1$  to remain at high frequencies for a longer time before it can fix. During this time lapse, (i) mutation restores some of the neutral genetic diversity in the population and (ii) genetic drift causes very frequent neutral variants to fix, thus decreasing their proportions in the SFS. Hence at the time of fixation of the beneficial mutation, the selective sweep looks older than a selective sweep with a constant  $s$ . We also used another method in which the likelihood-ratio test is calculated from the linkage disequilibria between neutral sites (KIM and NIELSEN 2004). Specifically, this method detects a peculiar pattern of linkage disequilibrium generated by complete selective sweeps, in which the linkage disequilibrium is strong between loci located on the same side of the selected locus, but is very low between loci located on each side of the selected locus (KIM and NIELSEN 2004; STEPHAN *et al.* 2006). However, this method failed to detect selection in our example, because the linkage disequilibrium was quickly broken down at the end of the selective sweep (not shown).

## DISCUSSION

We studied the dynamics of a beneficial mutation at a gene affecting a quantitative trait that also harbors background genetic variance contributed by other loci. To do so, we used a purely deterministic model in the range of allelic frequencies where stochastic effects can

be neglected. We showed that, if the effect of a mutation at the focal locus is small relative to the mean genetic background value of the trait, the full trajectory of the mutation (and of the mean background) can be derived under various fitness functions and related to the genetic and selective parameters of the trait. We also found conditions for a complete selective sweep at a quantitative trait locus under stabilizing selection. The selection coefficient of the beneficial mutation decreases in time because of background genetic variance under two of the three fitness functions studied (linear and Gaussian). The deterministic reduction of heterozygosity at a linked neutral locus is mostly influenced by the selection coefficient in the early generations of the selective sweep, as expected from BARTON (1998), but the effective selection coefficient for the hitchhiking effect can be much lower than the one at the starting conditions.

In the following, we first discuss the limits of the model that we used and the robustness of our results to departures from our assumptions, and then we discuss some possible improvements and applications.

**Potential limits of the model:** In this study, we neglected the possible changes in the variance of the trait for the sake of clarity. In fact, selection modifies the genetic variance of the population at each generation following  $\Delta\sigma^2 = 1/2\sigma^4(\gamma - \beta^2)$ , where the gradient of quadratic selection on the trait  $\gamma$  measures the mean curvature of the adaptive landscape experienced by the population (LANDE and ARNOLD 1983). Genetic drift decreases the genetic variance at rate  $(1 - 1/(2N_e))$  per generation, whereas mutation increases it by an amount  $V_m^2$ . The combined effect of all these factors on the change in variance depends on the parameter values and on the type of selection (fitness function) operating on the trait. However, assuming a constant variance may be a good approximation in the short term, *i.e.*, in the early generations of the selective sweep where the hitchhiking effect is strongest. Moreover, under stabilizing selection, and assuming that the population was at mutation–selection–drift equilibrium before the environmental change—that is, it remained at the optimum for many generations—BURGER and LYNCH (1995) showed that the environmental change actually increases the variance of the trait under selection (see their Figure 6). Hence, if the population was at equilibrium prior to the environmental change (which seems a reasonable assumption), our results obtained with a constant variance are conservative and actually underestimate the reduction of the selection coefficient of a new mutation affecting a quantitative trait due to selection on the genetic background of this trait.

As explained in the text, the fitness functions used here were chosen for illustrative purposes, but also because they can be good approximations to real fitness functions in the vicinity of the present state of a population. The linear and Gaussian fitness functions

are widely used to describe directional and stabilizing selective pressures on traits, respectively, because of their simplicity and because they are good approximates to more complicated functions under a wide range of parameter values (see LANDE 1976 for a discussion of the Gaussian approximation for general stabilizing selection). The exponential fitness function is less frequent in models of directional selection as it seems more extreme than the linear one (but see LANDE 1983). Its prevalence in natural populations is difficult to assess, since it has rarely been tested explicitly on empirical data. Nevertheless SCHLUTER (1988), using a nonparametric approach, demonstrated empirically that exponential-like fitness functions occur in natural populations. Moreover, these functions have the interesting property that they combine a strong positive slope (gradient of directional selection) with a positive curvature [positive quadratic selection gradient (LANDE and ARNOLD 1983)]. Since a positive curvature is in general interpreted as a sign of disruptive selection, finding this feature together with a strong directional gradient may seem like an empirical paradox. SCHLUTER (1988) emphasized that such a pattern may very well be caused by exponential-like fitness functions. This type of fitness function may thus be important in the study of evolutionary biology and is also quite illustrative, so we considered it as well.

Also, we assumed that the distribution of background genetic values was the same in every genetic class at the QTL. However, in a finite population, even a very large one, when the frequency of the beneficial mutation is very low (or very high), the number of individuals in a particular genotypic class at the focal locus is small, so this class may harbor a background genetic variance different from those of the other genotypic classes. This is an important stochastic factor, since a new mutation may be associated by chance to very beneficial or very deleterious alleles at other loci, instead of experiencing all possible genotypic values at other loci. This can be accounted for by noting that the background genetic variance in a genotypic class of  $n$  individuals is distributed like the sample variance for a sample of size  $n$  (since we assume no linkage and no epistasis). Its expectation is  $(n - 1)/n\sigma^2$ , which for large  $n$  tends to  $\sigma^2$ , the variance of the entire population, and its coefficient of variation (*i.e.*, the standard deviation divided by the mean) is  $\sqrt{2/(n - 1)}$ , a decreasing function of  $n$ . This coefficient of variation depends only on the number  $n$  of individuals in the genotypic class, regardless of the total population size. We can thus define  $n_\alpha = (2 + \alpha^2)/\alpha^2$  as the threshold number of individuals in a given genotypic class at locus  $A$ , above which the coefficient of variation of the background genetic variance is below a tolerance level  $\alpha$ . The definition of  $\varepsilon$  in Equation 7 can be extended to represent the frequency above which not only do the dynamics become quasi-deterministic without background variation (STEPHAN *et al.* 1992), but also all genotypes harbor background genetic variances close

to that in the entire population (that is  $\varepsilon > n_\alpha/N$ , where  $N$  is the population size).

Another stochastic effect that might matter in the competition between several loci affecting the trait is the reduction of the effective size of the population. Indeed, selection decreases the effective population size by creating variance in family size (SANTIAGO and CABALLERO 1995), thus reducing the efficacy of selection. Since the most frequent alleles contribute most to the genetic variance in fitness (all other things being equal), they also have the strongest effect in decreasing the effective population size for other loci. Hence they should experience a positive feedback that further increases their deterministic advantage of starting at higher frequencies. A comprehensive treatment of this stochastic competition between unlinked mutations is worth investigating, but it is beyond the scope of this article.

Finally, we have assumed that the focal locus is completely unlinked to the loci in the genetic background and that there is no epistasis between the locus and its genetic background for the trait considered. Yet, epistatic interactions for quantitative traits (defined as departures from additivity) have been repeatedly reported in QTL analyses (CARLBORG *et al.* 2003; KROYMANN and MITCHELL-OLDS 2005). Epistasis could affect our results by modifying the mean effect of the mutation over all the genetic backgrounds, thus changing  $a$  in our model. However, there is no reason to think that epistasis is systematically biased toward positive or negative values, such that the pooled effect of all loci in the background would result in a systematic decrease or increase of  $a$ . Hence as far as each genotype at the focal locus is represented by many individuals, and the genetic background consists of many loci, epistatic interactions should sum up to zero. Linkage is difficult to incorporate in the framework that we used; it would introduce partial covariance between the focal locus and the genetic background, resulting in an apparent bias for  $a$ . Contrary to epistasis, this bias may be sustained over several generations, which may influence the outcome of the selective sweep. In any case, the presence of a second selective sweep in the vicinity of the focal selective sweep may substantially affect the polymorphism pattern in the region as compared to a single selective sweep (CHEVIN *et al.* 2008). Therefore, the present model is *a priori* best suited for fully sexual species and traits for which QTL are spread over the genome.

**Potential developments and applications:** In practice, both the focal mutation and the genetic background may affect several traits under selection. Pleiotropy can alter predictions regarding selection at the focal locus based on a single character (OTTO 2004), and genetic covariances in the background can lead to changes of the mean genetic value of the focal trait as an indirect consequence of selection on other traits (LANDE 1979). Multivariate approaches are now a standard tool in evolutionary genetics (WALSH 2007) and have been applied recently

to relate the phenotypic effect of pleiotropic mutations to their fitness effects, in the absence of background genetic variation (MARTIN and LENORMAND 2006a). To understand how selection affects pleiotropic mutations in the presence of other polymorphic loci, our approach may be generalized by treating the focal mutation as a vector of effects on several traits, in the presence of a background genetic variance–covariance  $G$ -matrix, as in AGRAWAL *et al.* (2001). Even in this context, the simple results presented here may be good approximations in cases where the focal mutation has unbalanced effects and influences mainly one trait (low effective dimensionality) and where there is little background genetic covariance between this trait and other traits under selection.

Our results may have important implications regarding the methods used to detect signatures of selection. TESHIMA and PRZEWORSKI (2006) showed that dominance can modify the trajectory of a beneficial mutation (relative to the case of additive selective advantage) in a way that dramatically influences the signature that a selective sweep leaves on neutral variation. In our general context, background genetic variance can provoke a significant decrease in the selection coefficient over time. Note that a decrease of the selection coefficient, as described here under linear or Gaussian fitness function, may also occur under other types of fitness functions, for instance, those that increase until they reach a “plateau.” When we considered only the reduction of heterozygosity and neglected neutral mutations during the selective sweep, the two types of signatures (constant selection coefficient or decreasing selection coefficient) were barely distinguishable, but the effective selection coefficient in the case of a decreasing  $s$  was inferior to the initial  $s$ . When we included mutation during the selective sweep and looked at the frequency spectrum of mutations, the selection coefficient estimated was much lower, and we failed to detect selection through its effect on linkage disequilibrium. Hence at the time the beneficial mutation reaches fixation, a selective sweep with decreasing selection coefficient is very similar to an old selective sweep. In this context, it may thus be more efficient to use methods that search for an ongoing selective sweep (as in, *e.g.*, VOIGHT *et al.* 2006), since the methods that assume that the beneficial mutation is fixed may have a low power as a consequence of the decrease of the selection coefficient in time. It may also be possible to estimate the decrease of the selection coefficient of a mutation. This could be done, for instance, by modifying methods that jointly infer the selection coefficient and the age of a selected sweep, such as that of PRZEWORSKI (2003).

**Conclusion:** There has been marked interest recently in the population genetic theory of adaptation. Besides the theoretical developments stemming from Fisher’s geometrical model (ORR 2005), experimental evolution with microbes (ELENA and LENSKI 2003) provides exam-

ples of how adaptation proceeds in controlled laboratory conditions. Specifically, empirical estimates of the distribution of the fitness effects of mutations (reviewed in EYRE-WALKER and KEIGHTLEY 2007), and in particular of beneficial ones, aim at quantifying the raw material for adaptation. However, the applicability of fitness effects measured in the laboratory to natural populations has been questioned only in a few studies to date. It is not clear whether the selection coefficients estimated under specific controlled conditions can be directly translated into another context. Fundamentally, the underlying questions are as follows: Can we assign a constant selection coefficient to a given mutation? And if not, to what extent is this selection coefficient determined by other factors?

First, the environment that the population experiences affects the fitness effect of a mutation. Using a multivariate model of stabilizing selection and comparing it with data from the literature, MARTIN and LENORMAND (2006b) showed that the effect of a change of environment on the distribution of fitness effects of mutations follows a predictable trend. The selection coefficient of a mutation may also depend on the genetic environment in which it occurs. Most empirical studies of the distribution of the fitness effects of mutations focus on single mutations in a given reference background (EYRE-WALKER and KEIGHTLEY 2007), so that they are only informative about the process of adaptation from rare *de novo* mutations where only one mutation can sweep at a time. When several mutations segregate simultaneously at several loci, it is not clear what their selection coefficients will become, *i.e.*, what the effect of a variable genetic background is on allele frequency changes. Theoretical predictions (MARTIN *et al.* 2007), confirmed by experimental results (ELENA and LENSKI 1997; SANJUAN *et al.* 2004), indicate that epistasis for fitness between pairs of mutations can be substantial. In sexual species with smaller population sizes, such as higher eukaryotes, quantitative traits under selection can exhibit substantial variation caused by many loci (FALCONER and MACKAY 1996; BARTON and KEIGHTLEY 2002). In such a situation, the selection coefficient of an allele, defined either at the locus level only or as a departure from the mean fitness of the population (as in KIMURA and CROW 1978 or BARTON and TURELLI 1991), can be informative only about selection acting in one generation. The actual dynamics of a mutation all along its trajectory are more complex. As we show here with a simple model, the dynamics of a beneficial mutation affecting a quantitative trait under selection depend not only on its own effect, but also on the mean and variance of the genetic background for the trait and on the strength of selection on this trait. Moreover, the relative importance of each of these parameters crucially depends on the shape of the adaptive landscape. In any case, the selection coefficient that matters for the dynamics of a gene cannot be related

in a simple manner to the proportion of genetic variance explained by this gene [a quantity often used to quantify the QTL effect in empirical studies (LYNCH and WALSH 1998)].

To better understand the meaning of molecular signatures of selection in humans or in model species of higher eukaryotes (such as fruit flies, *Arabidopsis*, etc.), it is thus essential to empirically assess how adaptation proceeds in those species. If the model of periodic selection (ATWOOD *et al.* 1951; ELENA *et al.* 1996) applies, then theoretical results from the population genetics of adaptation and experimental evolution on microbes can be helpful in understanding adaptation in those species, too. In contrast, if the response to selection is essentially multigenic, selection at specific loci may strongly vary in time and depend on the background genetics of the trait. If so, the effective selection coefficient for molecular signatures of selection would be only partly informative about the actual advantage of the mutation while it segregated in the population. Moreover, some types of selection on traits (*i.e.*, shapes of the fitness function) would be over-represented in detectable selective sweeps, such that some categories of adaptive traits would be systematically missed by genome scans. On the other hand, specific models of phenotypic selection such as the ones proposed here provide alternative null models of variable selection coefficients that could be tested with molecular data.

We thank Emmanuelle Porcher, Guillaume Martin, Russell Lande, and three anonymous reviewers for helpful comments and criticisms on earlier versions of this manuscript. L.-M.C. is supported by a bourse de doctorat pour ingénieurs from the Centre National de la Recherche Scientifique.

#### LITERATURE CITED

- AGRAWAL, A. F., E. D. BRODIE and L. H. RIESEBERG, 2001 Possible consequences of genes of major effect: transient changes in the G-matrix. *Genetica* **112**: 33–43.
- AKEY, J. M., M. A. EBERLE, M. J. RIEDER, C. S. CARLSON, M. D. SHRIVER *et al.*, 2004 Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* **2**: e286.
- ATWOOD, K. C., L. K. SCHNEIDER and F. J. RYAN, 1951 Periodic selection in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **37**: 146–155.
- BARTON, N. H., 1995 Linkage and the limits to natural selection. *Genetics* **140**: 821–841.
- BARTON, N. H., 1998 The effect of hitch-hiking on neutral genealogies. *Genet. Res.* **72**: 123–133.
- BARTON, N. H., 2000 Genetic hitchhiking. *Philos. Trans. R. Soc. Lond.* **355**: 1553–1562.
- BARTON, N. H., and P. D. KEIGHTLEY, 2002 Understanding quantitative genetic variation. *Nat. Rev. Genet.* **3**: 11–21.
- BARTON, N. H., and M. TURELLI, 1991 Natural and sexual selection on many loci. *Genetics* **127**: 229–255.
- BURGER, R., and M. LYNCH, 1995 Evolution and extinction in a changing environment—a quantitative-genetic analysis. *Evolution* **49**: 151–163.
- CARLBORG, O., S. KERJE, K. SCHUTZ, L. JACOBSSON, P. JENSEN *et al.*, 2003 A global search reveals epistatic interaction between QTL for early growth in the chicken. *Genome Res.* **13**: 413–421.
- CHEVIN, L. M., S. BILLIARD and F. HOSPITAL, 2008 Hitchhiking both ways: effect of two interfering selective sweeps on linked neutral variation. *Genetics* **180**: 301–316.
- CROW, J. F., and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- DESAI, M. M., and D. S. FISHER, 2007 Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* **176**: 1759–1798.
- DESAI, M. M., D. S. FISHER and A. W. MURRAY, 2007 The speed of evolution and maintenance of variation in asexual populations. *Curr. Biol.* **17**: 385–394.
- ELENA, S. F., and R. E. LENSKI, 1997 Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**: 395–398.
- ELENA, S. F., and R. E. LENSKI, 2003 Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **4**: 457–469.
- ELENA, S. F., V. S. COOPER and R. E. LENSKI, 1996 Punctuated evolution caused by selection of rare beneficial mutations. *Science* **272**: 1802–1804.
- EWENS, W. J., 2004 *Mathematical Population Genetics—I. Theoretical Introduction*. Springer-Verlag, New York.
- EYRE-WALKER, A., and P. D. KEIGHTLEY, 2007 The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**: 610–618.
- FALCONER, D. S., and T. F. MACKAY, 1996 *Introduction to Quantitative Genetics*. Longman Group, Harlow, England.
- FISHER, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**: 35–42.
- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, London/New York/Oxford.
- GERRISH, P. J., and R. E. LENSKI, 1998 The fate of competing beneficial mutations in an asexual population. *Genetica* **102–103**: 127–144.
- GLINKA, S., L. OMETTO, S. MOUSSET, W. STEPHAN and D. DE LORENZO, 2003 Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics* **165**: 1269–1278.
- HERMISSE, J., and P. S. PENNING, 2005 Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**: 2335–2352.
- INNAN, H., and Y. KIM, 2004 Pattern of polymorphism after strong artificial selection in a domestication event. *Proc. Natl. Acad. Sci. USA* **101**: 10667–10672.
- KIM, Y., and R. NIELSEN, 2004 Linkage disequilibrium as a signature of selective sweeps. *Genetics* **167**: 1513–1524.
- KIM, Y., and W. STEPHAN, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**: 765–777.
- KIMURA, M., and J. F. CROW, 1978 Effect of overall phenotypic selection on genetic change at individual loci. *Proc. Natl. Acad. Sci. USA* **75**: 6168–6171.
- KINGSOLVER, J. G., H. E. HOEKSTRA, J. M. HOEKSTRA, D. BERRIGAN, S. N. VIGNIERI *et al.*, 2001 The strength of phenotypic selection in natural populations. *Am. Nat.* **157**: 245–261.
- KROYMANN, J., and T. MITCHELL-OLDS, 2005 Epistasis and balanced polymorphism influencing complex trait variation. *Nature* **435**: 95–98.
- LANDE, R., 1976 Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**: 314–334.
- LANDE, R., 1979 Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33**: 402–416.
- LANDE, R., 1983 The response to selection on major and minor mutations affecting a metrical trait. *Heredity* **50**: 47–65.
- LANDE, R., and S. J. ARNOLD, 1983 The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226.
- LENSKI, R. E., M. R. ROSE, S. C. SIMPSON and S. C. TADLER, 1991 Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**: 1315–1341.
- LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- MARTIN, G., and T. LENORMAND, 2006a A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* **60**: 893–907.

- MARTIN, G., and T. LENORMAND, 2006b The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution* **60**: 2413–2427.
- MARTIN, G., S. F. ELENA and T. LENORMAND, 2007 Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat. Genet.* **39**: 555–560.
- MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**: 23–35.
- NIELSEN, R., 2005 Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**: 197–218.
- NIELSEN, R., S. WILLIAMSON, Y. KIM, M. J. HUBISZ, A. G. CLARK *et al.*, 2005 Genomic scans for selective sweeps using SNP data. *Genome Res.* **15**: 1566–1575.
- NIELSEN, R., I. HELLMANN, M. HUBISZ, C. BUSTAMANTE and A. G. CLARK, 2007 Recent and ongoing selection in the human genome. *Nat. Rev. Genet.* **8**: 857–868.
- OHTA, T., 1992 The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23**: 263–286.
- OLSEN, K. M., A. L. CAICEDO, N. POLATO, A. MCCLUNG, S. MCCOUCH *et al.*, 2006 Selection under domestication: evidence for a sweep in the rice waxy genomic region. *Genetics* **173**: 975–983.
- ORR, H. A., 2005 The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**: 119–127.
- OTTO, S. P., 2004 Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc. Biol. Sci.* **271**: 705–714.
- PERFEITO, L., L. FERNANDES, C. MOTA and I. GORDO, 2007 Adaptive mutations in bacteria: high rate and small effects. *Science* **317**: 813–815.
- PRZEWORSKI, M., 2003 Estimating the time since the fixation of a beneficial allele. *Genetics* **164**: 1667–1676.
- PRZEWORSKI, M., G. COOP and J. D. WALL, 2005 The signature of positive selection on standing genetic variation. *Evol. Int. J. Org. Evol.* **59**: 2312–2323.
- SANJUAN, R., A. MOYA and S. F. ELENA, 2004 The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc. Natl. Acad. Sci. USA* **101**: 15376–15379.
- SANTIAGO, E., and A. CABALLERO, 1995 Effective size of populations under selection. *Genetics* **139**: 1013–1030.
- SCHLENKE, T. A., and D. J. BEGUN, 2004 Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* **101**: 1626–1631.
- SCHLOTTERER, C., 2003 Hitchhiking mapping: functional genomics from the population genetics perspective. *Trends Genet.* **19**: 32–38.
- SCHLUTER, D., 1988 Estimating the form of natural-selection on a quantitative trait. *Evolution* **42**: 849–861.
- STEPHAN, W., T. H. E. WIEHE and M. W. LENZ, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor. Popul. Biol.* **41**: 237–254.
- STEPHAN, W., Y. S. SONG and C. H. LANGLEY, 2006 The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics* **172**: 2647–2663.
- TANG, K., K. R. THORNTON and M. STONEKING, 2007 A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* **5**: 1587–1602.
- TENAILLON, O., O. K. SILANDER, J. P. UZAN and L. CHAO, 2007 Quantifying organismal complexity using a population genetic approach. *PLoS ONE* **2**: e217.
- TESHIMA, K. M., and M. PRZEWORSKI, 2006 Directional positive selection on an allele of arbitrary dominance. *Genetics* **172**: 713–718.
- TESHIMA, K. M., G. COOP and M. PRZEWORSKI, 2006 How reliable are empirical genomic scans for selective sweeps? *Genome Res.* **16**: 702–712.
- THORNTON, K. R., J. D. JENSEN, C. BECQUET and P. ANDOLFATTO, 2007 Progress and prospects in mapping recent selection in the genome. *Heredity* **98**: 340–348.
- TURELLI, M., 1988 Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution* **42**: 1342–1347.
- VOIGHT, B. F., S. KUDARAVALLI, X. WEN and J. K. PRITCHARD, 2006 A map of recent positive selection in the human genome. *PLoS Biol.* **4**: e72.
- WALSH, B., 2007 Evolutionary quantitative genetics, pp. 533–580 in *Handbook of Statistical Genetics*, Ed. 3, edited by D. J. BALDING, M. BISHOP and C. CANNINGS. John Wiley & Sons, Chichester, UK.
- WANG, R. L., A. STEC, J. HEY, L. LUKENS and J. DOEBLEY, 1999 The limits of selection during maize domestication. *Nature* **398**: 236–239.
- WILLIAMSON, S. H., M. J. HUBISZ, A. G. CLARK, B. A. PAYSEUR, C. D. BUSTAMANTE *et al.*, 2007 Localizing recent adaptive evolution in the human genome. *PLoS Genet.* **3**: e90.
- WRIGHT, S., 1969 *Evolution and the Genetics of Populations, Volume 2: The Theory of Gene Frequencies*. University of Chicago Press, Chicago.

Communicating editor: J. WAKELEY

## APPENDIX

In this APPENDIX, we derive the exact and approximate dynamics of the frequency  $p$  of the  $A_1$  mutation at the focal locus and of the mean genetic background value  $m$ .

We first calculate the mean fitness in the population under the three fitness functions. The mean fitness of each genotypic classe  $A_i A_j$  is

$$\bar{W}_{A_i A_j} = \int_{-\infty}^{\infty} W(z) f(z - (m + \alpha_{ij} a)) dz,$$

where  $f(\cdot)$  is the Gaussian distribution with mean 0 and variance  $\sigma^2$  and  $\alpha_{ij} = 2$  if  $\{i, j\} = \{1, 1\}$ ,  $\alpha_{ij} = 1$  if  $\{i, j\} = \{1, 2\}$  or  $\{1, 2\}$ , and  $\alpha_{ij} = 0$  if  $\{i, j\} = \{2, 2\}$ . Then the mean fitness in the population is, according to Equation 12,

$$\begin{aligned} \bar{W}_1 &= b + \omega(m + 2ap) \\ \bar{W}_c &= e^{m\omega + (\sigma^2\omega^2)/2} ((e^{a\omega} - 1)p + 1) \\ \bar{W}_g &= \frac{p^2 e^{-((m+2a)^2/2(\sigma^2+\omega^2))} + 2pqe^{-((m+a)^2/2(\sigma^2+\omega^2))} + q^2 e^{-((m)^2/2(\sigma^2+\omega^2))}}{\sqrt{\eta}}, \end{aligned} \tag{A1}$$

where  $q = 1 - p$  and  $\eta = 1 + (\sigma/\omega)^2$ . The change in frequency of the allele  $A_1$  can be calculated using Equation 13, which gives

$$\begin{aligned}
 \Delta p_l &= \frac{a\omega}{b + (m + 2ap)\omega} pq \\
 \Delta p_e &= \frac{e^{a\omega} - 1}{(e^{a\omega} - 1)p + 1} pq \\
 \Delta p_g &= \frac{p + (q - p)e^{a(3a+2m)/2(\sigma^2+\omega^2)} - qe^{2a(a+m)/(\sigma^2+\omega^2)}}{p^2 + 2pqe^{a(3a+2m)/2(\sigma^2+\omega^2)} + q^2e^{2a(a+m)/(\sigma^2+\omega^2)}} \frac{pq}{2}.
 \end{aligned}
 \tag{A2}$$

The growth rate of the ratio of allelic frequencies  $\rho = p/q$  can then be found following Equation 15, which, after some rearrangement, leads to

$$\begin{aligned}
 s_l &= \frac{a\omega}{b + (m + ap)\omega} \\
 s_e &= e^{a\omega} - 1 \approx a\omega \\
 s_g &= \frac{p + (1 - 2p)e^{a(3a+2m)/2(\sigma^2+\omega^2)} - (1 - p)e^{2a(a+m)/(\sigma^2+\omega^2)}}{pe^{a(3a+2m)/2(\sigma^2+\omega^2)} + (1 - p)e^{2a(a+m)/(\sigma^2+\omega^2)}}.
 \end{aligned}
 \tag{A3}$$

Under linear and Gaussian fitness functions, the dynamics are slightly frequency dependent, since  $p$  is present in the expression of  $s$ . Note that it is also the case in the classical selection model (see Equation 2). We now introduce an approximation for mutations of small effects (similar to considering  $s \ll 1$  in Equation 3). This consists in considering that the effect of the focal mutation on the trait,  $a$ , is small relative to the mean background genetic value of the trait (discounting the effect of the mutation),  $m$ . The results obtained under this approximation are denoted by an “\*” in the rest of this article. When  $|a| \ll |m|$  (where “|” denotes the absolute value), the growth rate of the focal mutation becomes

$$\begin{aligned}
 s_l^* &= \frac{a\omega}{b + m\omega} \\
 s_e^* &= s_{e,\sigma^2=0} \approx a\omega \\
 s_g^* &= e^{-(am/(\sigma^2+\omega^2))} - 1 \approx -\frac{am}{\sigma^2 + \omega^2},
 \end{aligned}
 \tag{A4}$$

which is now independent of the frequency  $p$  of the mutation. Note that only in the case of Gaussian selection does the selection coefficient in one generation depend on the background genetic variance  $\sigma^2$ .

The change in the mean background value is controlled by the gradient of directional selection on the trait, as shown in Equation 11. Using the expressions for the mean fitness in (A1), those gradients are

$$\begin{aligned}
 \beta_l &= \frac{\omega}{b + (m + ap)\omega} \\
 \beta_e &= \omega \\
 \beta_g &= -\frac{p^2(m + 2a)e^{-((m+2a)^2/2(\sigma^2+\omega^2))} + 2pq(m + a)e^{-((m+a)^2/2(\sigma^2+\omega^2))} + q^2me^{-(m^2/2(\sigma^2+\omega^2))}}{(q^2e^{-(m^2/2(\sigma^2+\omega^2))} + 2pqe^{-((m+a)^2/2(\sigma^2+\omega^2))} + p^2e^{-((m+2a)^2/2(\sigma^2+\omega^2))})(\sigma^2 + \omega^2)}.
 \end{aligned}
 \tag{A5}$$

In the case of exponential selection, the directional selection gradient is constant and depends neither on the frequency of the mutation at the focal locus nor on the mean genetic background or the amount of genetic variance for the trait. In contrast, for the linear and Gaussian fitness functions, the selection gradient depends on the frequency  $p$  of the focal mutation. There is thus complete interdependency between the dynamics of the focal mutation and that of the genetic background for the trait: (i) the focal mutation influences the mean fitness  $\bar{W}$ , which changes the selective pressure on the trait ( $\beta$ ), and (ii) the change in the mean background genetic value  $m$  changes the dynamics of the mutation, characterized by the relative growth rate  $s$ . Therefore, there is no simple solution to the full system of equations. Nevertheless, under the small-effect assumption ( $|a| \ll |m|$ ) the gradients of directional selection become

$$\begin{aligned}
 \beta_l^* &= \frac{\omega}{b + m\omega} \\
 \beta_e^* &= \beta_e = \omega \\
 \beta_g^* &= -\frac{m}{\sigma^2 + \omega^2},
 \end{aligned}
 \tag{A6}$$

which are all independent of the frequency of the  $A_I$  allele. This allows calculating the trajectory in time of the mean genetic background  $m$  first, and then using it to find the full dynamics of the beneficial mutation at the focal locus.