

# Rapid Adaptation of a Polygenic Trait After a Sudden Environmental Shift

Kavita Jain<sup>\*.1</sup> and Wolfgang Stephan<sup>†.‡</sup>

<sup>\*</sup>Theoretical Sciences Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560064, India, <sup>†</sup>Natural History Museum Berlin, 10115, Germany, and <sup>‡</sup>Biocenter, University of Munich, 82152 Planegg, Germany

**ABSTRACT** Although a number of studies have shown that natural and laboratory populations initially well adapted to their environment can evolve rapidly when conditions suddenly change, the dynamics of rapid adaptation are not well understood. Here a population genetic model of polygenic selection is analyzed to describe the short-term response of a quantitative trait after a sudden shift of the phenotypic optimum. We provide explicit analytical expressions for the timescales over which the trait mean approaches the new optimum. We find that when the effect sizes are small relative to a scaled mutation rate, small to moderate allele frequency changes occur in the short-term phase in a synergistic fashion. In contrast, selective sweeps, *i.e.*, dramatic changes in the allele frequency, may occur provided the size of the effect is sufficiently large. Applications of our theoretical results to the relationship between QTL and selective sweep mapping and to tests of fast polygenic adaptation are discussed.

**KEYWORDS** polygenic selection; unequal effects; rapid adaptation

IN “What Darwin got wrong,” Losos (2014) persuasively argues that we can observe evolution in action and, in particular, that evolution can be so rapid that evolutionary and ecological timescales are confluent. The examples range from the peppered moth (Cook *et al.* 2012), insecticide resistance in *Drosophila* (Ffrench-Constant *et al.* 2002), color of field mice (Vignieri *et al.* 2010), beak size in Darwin’s finches (Grant and Grant 2008), guppies in Trinidad (Reznick 2009), and *Anolis* lizards (Losos 2009) to name but a few. The rapid changes are responses to natural and human-induced shifts in the environment. The genetic architecture underlying these traits ranges from few genes of major effect to highly polygenic systems (van’t Hof *et al.* 2011; Lamichhaney *et al.* 2012, 2015; Linnen *et al.* 2013). In this article, we study a model that encompasses a wide range of genetic architectures. Our aim is to understand the genomic signatures of positive selection in these systems that occur after environmental changes, with an emphasis on rapid adaptation.

There is a growing body of literature on the detection of adaptive signatures in molecular population genetics. Following the pioneering work of Maynard Smith and Haigh (1974), the impact of positive selection on neutral DNA variability (selective sweeps) has attracted much interest. This theory has been applied to huge data sets that emerge from modern high-throughput sequencing. A large number of statistical tests have been developed to detect sweep signals and estimate the frequency and strength of selection (Kim and Stephan 2002; Nielsen *et al.* 2005; Pavlidis *et al.* 2010). To find sweep signatures in the genome, strong positive selection is required (with  $N_e s \gg 1$ , where  $N_e$  is the effective population size and  $s$  is the selective advantage of a beneficial allele) (Kaplan *et al.* 1989; Stephan *et al.* 1992). Thus, sweeps are characteristic signals of fast adaptation. They are expected to be found at individual genes or at major loci if a trait is controlled by multiple genes.

This raises the question whether fast evolution is also possible in highly polygenic systems and, if so, which genomic signatures arise. A number of genome-wide association studies (GWAS) have shown that quantitative traits are typically highly polygenic (*e.g.*, Turchin *et al.* 2012) and there is growing evidence that the molecular scenario of sweeps only covers part of the adaptive process and needs to be revised to include polygenic selection (Pritchard and Di Rienzo 2010). As GWAS yield information about the distribution of single

Copyright © 2017 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.116.196972>

Manuscript received October 18, 2016; accepted for publication March 7, 2017; published Early Online March 24, 2017.

Supplemental material is available online at [www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.196972/-/DC1](http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.196972/-/DC1).

<sup>1</sup>Corresponding author Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore, 560064, India. E-mail: [jain@jncasr.ac.in](mailto:jain@jncasr.ac.in)

nucleotide polymorphisms (SNPs) relevant to quantitative traits (Visscher *et al.* 2012), it is important to understand the models of polygenic selection in terms of the frequency changes of molecular variants, *i.e.*, with reference to population genetics (Bürger 2000).

Although the equilibrium structure of the allele frequencies and the stationary variance have been the subject of research in a large number of such studies (Bürger 2000), the dynamics have received relatively little attention. Perhaps the most well-studied model in this context is Lande's model (Lande 1983) in which the changes in the allele frequency at a major locus in the background of a large number of minor loci are considered (Chevin and Hospital 2008; Gomulkiewicz *et al.* 2010). However, in Lande's model, the background is not explicitly modeled and it is assumed that the background variance does not evolve. Pavlidis *et al.* (2012) have studied a more detailed model but their analysis is restricted to a very small number of loci.

Following our preliminary study (Jain and Stephan 2015), here we perform a population genetic analysis of the dynamics of a polygenic trait after a sudden environmental shift of the phenotypic optimum. We consider a quantitative trait that is determined additively by a large number of diallelic loci of unequal effects. The population is assumed to be infinitely large and to evolve under stabilizing selection and mutation. This model was first proposed by Wright (1935) and more recently revisited by Barton (Barton 1986; de Vladar and Barton 2014). We carry out a detailed study of this model to understand the dynamics of allele frequencies as well as of the trait mean and variance. In particular, to describe rapid adaptive evolution, we concentrate on the short-term period after the environmental change, which may be defined as the time until the phenotypic mean reaches a value close to the new optimum.

We find that the short-time dynamics are well described by a *directional selection model* which is analytically tractable. Working within the framework of this simpler model, we reproduce some results obtained using different models or assumptions (Lande 1983; Chevin and Hospital 2008; Jain and Stephan 2015) when most effects are small relative to a scaled mutation rate (de Vladar and Barton 2014). In addition, we obtain several new results, in particular when most effects are large.

## Model

We consider an infinitely large population of diploids in which a trait  $z$  is determined by  $\ell$  diallelic loci. At the  $i$ th locus, let the + allele have an effect  $\gamma_i/2$  and frequency  $p_i$  and, correspondingly, the - allele have an effect  $-\gamma_i/2$  and frequency  $q_i = 1 - p_i$  (de Vladar and Barton 2014; Jain and Stephan 2015). The effect at each locus is assumed to be exponentially distributed with mean  $\bar{\gamma}$ , as is often the case in quantitative genetic studies (Mackay 2004; Goddard and Hayes 2009). Then, on neglecting dominance and epistasis, the trait  $z$  is determined additively and its mean  $c_1$  can be written as

$$c_1 = \sum_{i=1}^{\ell} (\gamma_i p_i - \gamma_i q_i) = \sum_{i=1}^{\ell} \gamma_i (2p_i - 1). \quad (1)$$

For a phenotypic trait under stabilizing selection, assuming that the deviation of the fitness from its optimum  $z'$  is quadratic (Bürger 2000), we can write the phenotypic fitness as  $w(z) = 1 - (s/2)(z - z')^2$ , where  $s$  measures the strength of stabilizing selection on the trait. Averaging over the phenotypic-trait distribution (which is not necessarily Gaussian), one finds the average fitness to be

$$\bar{w} = 1 - (s/2) \overline{(z^2 - c_1^2) + (c_1^2 - 2zz' + z'^2)}, \quad (2)$$

$$\approx \exp \left\{ -(s/2) \left[ c_2 + (c_1 - z')^2 \right] \right\}, \quad (3)$$

where  $c_2$  is the variance of the trait and is given by (see Appendix A)

$$c_2 = 2 \sum_{i=1}^{\ell} \gamma_i^2 p_i q_i. \quad (4)$$

When selection is weak and recombination rate is high [linkage equilibrium (LE); see also Appendix B], and mutations are absent, the allele frequency  $p_i$  evolves according to Wright's equation as (Bürger 2000, Chap. 6)

$$\dot{p}_i = \frac{p_i(1 - p_i)}{2\bar{w}} \frac{\partial \bar{w}}{\partial p_i}, \quad (5)$$

where the dot denotes the derivative with respect to time. For the model described above, we then have (de Vladar and Barton 2014; Jain and Stephan 2015)

$$\begin{aligned} \dot{p}_i = & -s\gamma_i p_i q_i (c_1 - z') - \frac{s\gamma_i^2}{2} p_i q_i (q_i - p_i) + \mu(q_i - p_i), \\ & i = 1, \dots, \ell. \end{aligned} \quad (6)$$

In the above equation, the first two terms on the right-hand side (RHS) follow from (3) and (5). The first term tends to stabilize the phenotypic mean to the optimum trait value, and the second term results in the fixation of one of the alleles, thus depleting genetic variation. The last term on the RHS models the mutation process restoring genetic variance (Barton 1986), and describes changes between the + and - allele at locus  $i$  with locus-independent, symmetric rate  $\mu$ . In the following, we refer to the model defined by (6) as the *full model*.

In this article, we are interested in the response of the phenotypic mean and variance when the phenotypic optimum is suddenly shifted to a new value  $z_f$  in a population initially equilibrated to  $z_0 = 0$ . A preliminary investigation of such a situation has been carried out numerically by de Vladar and Barton (2014), and our goal here is to provide an analytical understanding of the dynamics. We will thus focus on the

dynamics of the allele frequency that evolves according to (6) when the phenotypic optimum  $z' = z_f$ , starting from the stationary state frequency  $p_i^*$  of the population that is equilibrated to a phenotypic optimum at zero.

Unless stated otherwise, in this article, we assume that  $z_f > 0$  so that the initial mean deviation from the phenotypic optimum,  $\Delta c_1(0) = -z_f$ , is negative. We also restrict the magnitude of the shift such that  $z_f < \ell\bar{\gamma}$ ; this is because the first term on the RHS of (6) acts to minimize the deviation of the trait mean from its optimum. But, from (1), the magnitude of the mean cannot exceed  $\sum_{i=1}^{\ell} \gamma_i \approx \ell\bar{\gamma}$ , and therefore a shift beyond the total effect of all loci is not within evolutionary reach.

## Results

### Stationary state of the full model

Before proceeding further, we briefly review the known results for the stationary state that are pertinent to the later discussion. In the stationary state where the allele frequencies are independent of time, setting the left-hand side (LHS) of (6) to zero and performing some simple algebra, we find that the equilibrium allele frequency  $p_i^*$  is a solution of the following equation,

$$p_i^{*3} - p_i^{*2} \left( \frac{3}{2} + \frac{\Delta c_1^*}{\gamma_i} \right) + p_i^* \left[ \frac{1}{2} + \frac{\Delta c_1^*}{\gamma_i} + \frac{1}{4} \left( \frac{\hat{\gamma}}{\gamma_i} \right)^2 \right] - \frac{1}{8} \left( \frac{\hat{\gamma}}{\gamma_i} \right)^2 = 0, \quad (7)$$

where  $\hat{\gamma} = 2\sqrt{2\mu/s}$  and  $\Delta c_1^*$  is the deviation of the mean from the optimum in equilibrium.

When the stationary mean deviation  $\Delta c_1^*$  is zero, Equation (7) for the equilibrium allele frequency has three solutions that are given by de Vladar and Barton (2014):

$$p_i^* = \begin{cases} \frac{1}{2}, & \gamma_i < \hat{\gamma} \\ \frac{1}{2} \pm \frac{1}{2} \sqrt{1 - \frac{\hat{\gamma}^2}{\gamma_i^2}}, & \gamma_i > \hat{\gamma}. \end{cases} \quad (8)$$

While the solution  $p_i^* = 1/2$  is stable when the effect is smaller than the threshold effect  $\hat{\gamma}$ , the two latter ones hold when the effect is larger than  $\hat{\gamma}$ . The threshold  $\hat{\gamma}$  between large- and small-effect alleles arises because of mutation. When selection is much weaker than mutation, the equilibrium frequency is one half; *i.e.*, when the stationary trait mean is at the optimum, mutation and selection balance each other at an intermediate allele frequency. The value of one half arises because we assumed that mutation in both directions is symmetric (see above). Relaxing this symmetry assumption would lead to different equilibrium frequencies. However, since we focus in the following on the short-term behavior of the dynamics, detailed mutation models are of relatively little importance compared to selection and will not be considered here. For exponentially distributed effects, as

the fraction of loci with large effects is given by  $\bar{\gamma}^{-1} \int_{\hat{\gamma}}^{\infty} dx e^{-x/\bar{\gamma}} = e^{-\hat{\gamma}/\bar{\gamma}}$ , most effects are large when  $\hat{\gamma} \ll \bar{\gamma}$ , while most effects are small in the opposite parameter regime.

Furthermore, on separating the contribution from loci with small and large effects and using the stationary state frequency (8), we obtain the stationary state variance to be (de Vladar and Barton 2014; Jain and Stephan 2015)

$$c_2^* = \ell\bar{\gamma}^2 \left[ 1 - \left( 1 + \frac{\hat{\gamma}}{\bar{\gamma}} \right) e^{-\hat{\gamma}/\bar{\gamma}} \right]. \quad (9)$$

The above result shows that when most effects are small,  $c_2^* \approx \ell\bar{\gamma}^2$ ; while in the opposite case, it is well approximated by  $\ell\hat{\gamma}^2$ .

The correction to the equilibrium allele frequency in (8), when the stationary mean deviation is nonzero but small, can be found in Appendix B of de Vladar and Barton (2014). For future reference, we note that these corrections are negligible except when the effect is close to the threshold  $\hat{\gamma}$  (see Figure 2 below).

### Dynamics when most effects are small

We now study the dynamics of the allele frequency when most effects are small ( $\bar{\gamma} \ll \hat{\gamma}$ ). Since the population is initially equilibrated to the phenotypic optimum at zero, the initial frequency at the  $i$ th locus is well approximated by (8), and close to one half when most effects are small. Then, as  $p_i(t) \approx q_i(t)$  at short times, we can neglect the last two terms on the RHS of (6) compared to the first term to get (Jain and Stephan 2015)

$$\dot{p}_i = S_i p_i q_i, \quad (10)$$

where  $S_i = -s\gamma_i\Delta c_1$  and  $\Delta c_1 = c_1 - z_f$ . When the final mean deviation is negligible, the final allele frequency is also close to one half, and we expect that (10) can describe the bulk of the allele frequency dynamics as is indeed supported by Figure 1 and Figure 2.

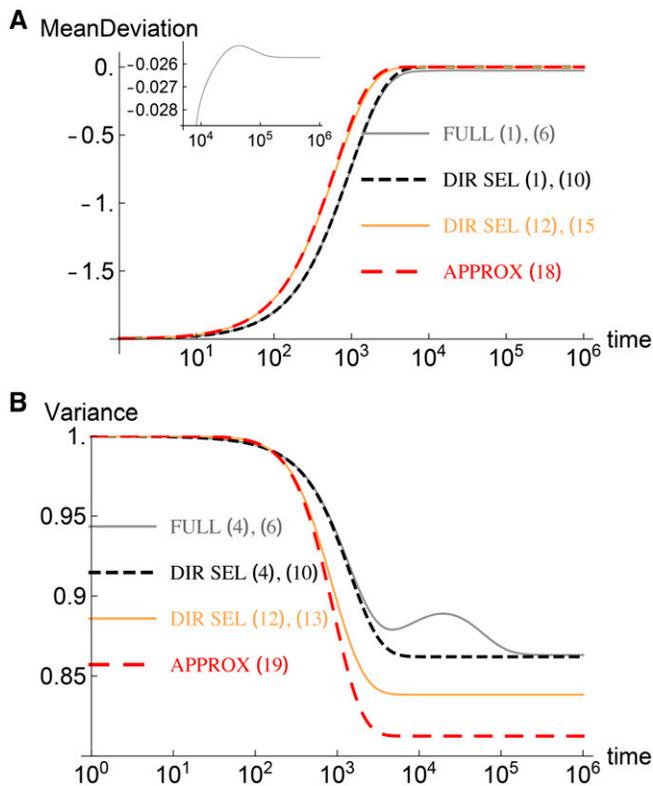
**Directional selection model:** In the following, we will refer to the model defined by (10) as the directional selection model. Appendix C details the prescription for calculating relevant quantities analytically in this model. As shown in Appendix C, the trait mean can be written as

$$c_1(t) = \sum_{i=1}^{\ell} \gamma_i - \sum_{i=1}^{\ell} \frac{2\gamma_i}{1 + \frac{p_i(0)}{q_i(0)} e^{\beta\gamma_i/\bar{\gamma}}}, \quad (11)$$

where the parameter  $\beta$  defined in (C6) is proportional to the logarithm of the ratio of allele frequencies and evolves in time as

$$\dot{\beta} = -s\bar{\gamma}\Delta c_1. \quad (12)$$

Thus a closed equation for  $\beta(t)$  can be obtained by eliminating the mean using (11), and on plugging the result for  $\beta(t)$  back



**Figure 1** Most effects are small: Dynamics of the (A) mean deviation and (B) variance (relative to its initial value) for  $\hat{\gamma} \approx 3.2\bar{\gamma} = 0.128$  and  $z_f = \ell\bar{\gamma}/4 = 2$ . The other parameters are  $s = 5 \times 10^{-3}$ ,  $\mu = 10^{-5}$ , and  $\bar{\gamma} = 0.04$ . The inset in (A) shows the mean deviation in the full model at large times. The gray and black curves are obtained for a single realization of effects in which 197 out of  $\ell = 200$  loci have effects smaller than the mean. The orange and red curves, on the other hand, are averaged over the distribution of effects. The results from the two procedures match when the number of loci is very large as explained in *Discussion*.

in (11), the mean can be found. Furthermore, it is shown in Appendix C that in the directional selection model, the  $n$ th cumulant of the effect evolves according to (Bürger 1991)

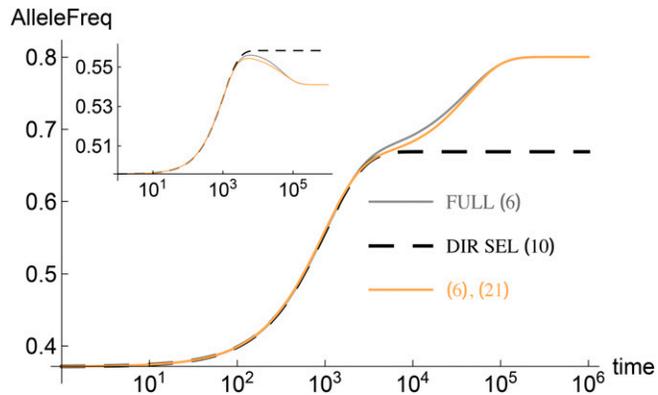
$$\dot{c}_n = -s\Delta c_1 c_{n+1}, n \geq 1, \quad (13)$$

which shows that the  $n$ th cumulant  $c_n$ ,  $n \geq 2$ , can be found once  $c_{n-1}$  is known. Thus to describe the short-time dynamics, we will focus on the key Equations 11 and 12.

We now return to (11) for the mean  $c_1$  and analyze it when the number of loci is large. When most effects are small ( $\hat{\gamma} \gg \bar{\gamma}$ ), as mentioned above, the initial allele frequency is one half. Furthermore, as explained in the *Discussion*, the sum over the contribution from individual loci in (11) amounts to averaging over the distribution of effects when the number of loci is large. For large  $\ell$ , on approximating the sum in (11) by an integral, we thus get

$$c_1(t) = \ell\bar{\gamma} - \frac{\ell}{\bar{\gamma}} \int_0^{\hat{\gamma}} d\gamma \frac{2\gamma e^{-\gamma/\bar{\gamma}}}{1 + e^{\beta\gamma/\bar{\gamma}}}. \quad (14)$$

Writing  $x = \gamma/\bar{\gamma}$  in the above integral, we get



**Figure 2** Most effects are small: Dynamics of the allele frequency for locus with effect  $\gamma_i = 0.121$  (main) and  $\gamma_i = 0.024$  (inset). All the parameters are the same as in Figure 1. Note that while the initial allele frequency in the inset is closely approximated by (8), the allele trajectory starts at a value much lower than one half in the main panel as the effect size is close to  $\hat{\gamma} \approx 0.126$ . In the latter case, the corrections to (8) are substantial and given by (B2) of de Vladar and Barton (2014).

$$c_1(t) = \ell\bar{\gamma}[1 - 2I_s(\beta)], \quad (15)$$

where

$$I_s(\beta) = \int_0^{\hat{\gamma}/\bar{\gamma}} dx \frac{x e^{-x}}{1 + e^{\beta x}}. \quad (16)$$

Our main task is to find the time dependence of  $\beta$  using (12) and (15), which is dealt with in Appendix D.

**Qualitative patterns of the cumulant dynamics:** Before turning to explicit expressions, we first make some general remarks on the behavior of the cumulants in the directional selection model. Since the variance is always nonnegative, it follows from (13) that the magnitude of mean deviation decreases with time. As we are considering the scenario where  $\Delta c_1(0) < 0$ , the mean deviation  $\Delta c_1(t)$  increases *monotonically* with time toward zero, as shown in Figure 1A. Due to (12), this also means that  $\beta(t)$  initially increases and then saturates to a constant.

Furthermore, due to (12) and (13), the variance can be written as

$$c_2 = -2\ell\bar{\gamma}^2 I_s'(\beta), \quad (17)$$

where prime denotes the derivative with respect to  $\beta$ . To arrive at the above expression, we have used that  $\dot{c}_1 = -2\ell\bar{\gamma} I_s'(\beta)\dot{\beta}$  due to (15). Equation 17 gives the rate of change of variance as  $\dot{c}_2 = -2\ell\bar{\gamma}^2 \dot{\beta} I_s''(\beta)$ . It is easy to check that the integrand of  $I_s''$  is always positive from which it follows that  $I_s''(\beta) > 0$ . Therefore the directional selection model predicts that when most effects are small, the variance *decreases* with time as indeed verified in Figure 1B. From this result and (13), it immediately follows that the skewness  $c_3$  defined in (A6) is always negative.

**Quantitative dynamics of the mean and higher cumulants:** As shown in Appendix D, the mean deviation vanishes exponentially fast,

$$\Delta c_1(t) = \Delta c_1(0)e^{-sc_2(0)t}, \quad (18)$$

where  $c_2(0) \approx \ell\bar{\gamma}^2$  is the initial variance given by (9) when most effects are small. The above result has been previously obtained by assuming that the variance and higher cumulants do not evolve in time (Lande 1983; Chevin and Hospital 2008; Jain and Stephan 2015). The basis of this assumption lies in phenotypic data (Lande 1976) or is motivated by mathematical tractability (Chevin and Hospital 2008; Gomulkiewicz *et al.* 2010; Jain and Stephan 2015). Here, in contrast, we have obtained (18) without any additional assumptions.

Using the solution (18) in (13) for cumulant dynamics, we find that the variance stays at its initial value  $c_2(0)$  and the higher cumulants vanish. The corrections to these behaviors at late times are given by (D13) and (D14), and we find that the variance remains a constant at short times  $\ll [sc_2(0)]^{-1}$  but decreases thereafter to a smaller value:

$$\frac{c_2(t)}{c_2(0)} \approx 1 - 3 \left( \frac{z_f}{\ell\bar{\gamma}} \right)^2 \left[ 1 - e^{-sc_2(0)t} \right]^2. \quad (19)$$

The above expression shows that the drop in the variance is larger for large shifts in the optimum and can be quite significant, as shown in Figure 1B, for a moderate shift (relative to the maximum mean). A comparison between the full model, directional selection model, and the analytical results is also shown in Figure 1 for mean and variance. The reason for the difference observed in Figure 1 between the cumulants obtained using (10) and the large  $\ell$  approximation (15) is explained in the *Discussion*.

The main panel of Figure 1A shows that the mean is very close to the stationary state at time  $t \approx 5000$ . However, as shown in the inset, the mean continues to change, albeit very slowly, until  $t \approx 2 \times 10^5$  when the true stationary state is reached. A similar pattern is seen for variance (and allele frequency shown below in Figure 2). These observations suggest that the dynamics of the mean can be divided into a short-term phase, in which the mean approaches a value close to the new optimum; and a long-term phase, in which it reaches the exact stationary state. To explore fast evolution, we will concentrate in the following on the short-term phase.

**Dynamics of the allele frequencies:** From (10) for allele frequency dynamics, we first note that for  $\Delta c_1(0) < 0$ , all the + allele frequencies increase in the short-term phase [note that so far, we have argued that (10) holds for small-effect loci only, but in the next section we will also find this to be true for large-effect loci]. This simple result has potentially useful application as explained in the *Discussion*.

Using (18) in (10), the time dependence of the allele frequency for small deviations in the phenotypic optimum

can be obtained analytically (Chevin and Hospital 2008; Jain and Stephan 2015):

$$p_i(t) = \frac{p_i(0)}{p_i(0) + q_i(0) \exp \left[ -\frac{z_f \gamma_i (1 - e^{-s\ell\bar{\gamma}^2 t})}{\ell\bar{\gamma}^2} \right]}. \quad (20)$$

Thus in the directional selection model, the allele frequency reaches the stationary state over a time period  $\sim (s\ell\bar{\gamma}^2)^{-1}$ . These shifts in allele frequency are small to moderate, as verbally predicted by several authors (*e.g.*, Pritchard and Di Rienzo 2010). However, this directional behavior may change in the long-term period, as we will discuss next.

The allele frequency dynamics in the full model and directional selection model are compared for some loci in Figure 2 and we see that while the two match well at short times, in the full model, the allele frequency at large times can vary in a *nonmonotonic* fashion (see inset of Figure 2) and approaches a stationary state value given by the solution of (7). The long-time dynamics of the allele frequency can be understood by solving (6) with zero mean deviation and the initial condition given by the stationary state solution of the directional selection model (Jain and Stephan 2015). A succinct way of expressing such a prescription is to write

$$\Delta c_1(t) = \Delta c_1^* + \Delta c_1(0)e^{-sc_2(0)t} \quad (21)$$

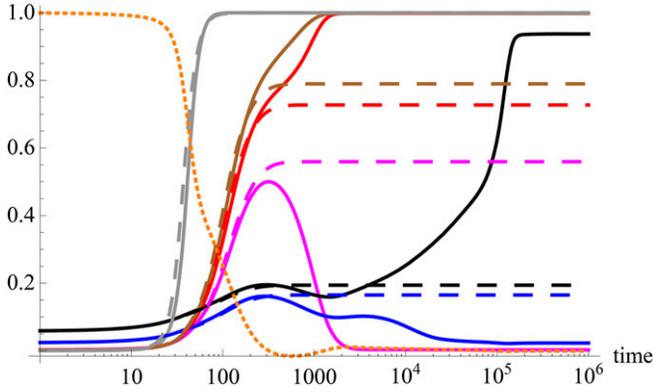
in (6). Numerical integration of the resulting equation matches well with the full model as shown in Figure 2. An analytical understanding of the long-time dynamics can also be obtained as described in Appendix E.

### Dynamics when most effects are large

We now turn to the situation when most effects are large ( $\bar{\gamma} \gg \hat{\gamma}$ ). Due to (8), the initial allele frequencies are close to either zero or one. When the final mean deviation is close to zero, either the final allele frequency stays close to the initial one or sweeps to fixation as shown in Figure 3. Thus, unlike in the case where most effects are small, in the full model the factor  $q_i - p_i$  in the last two terms on the RHS of (6) is not negligible. However, the directional selection model defined by (10) describes the short-time dynamics when most effects are large, as explained below.

Since we are dealing with the case where  $\hat{\gamma} \ll \gamma_i$ , we can ignore the effect of mutations. Furthermore, as  $|q_i - p_i|$  cannot exceed one, (6) shows that as long as  $2|c_1(t) - z_f| \equiv 2|\Delta c_1(t)| \gg \gamma_i$ , the second term on the RHS can be neglected to yield (10) for the allele frequency at the  $i$ th locus. Therefore we expect the directional selection model to apply at short times. But as time progresses, the mean deviation decreases in magnitude, and the above condition for the validity of the directional selection model fails.

Using the initial frequency given by (8) for  $\gamma_i \gg \hat{\gamma}$ , Equation 11 for the mean yields



**Figure 3** Selective sweeps when most effects are large: Dynamics of the scaled absolute mean deviation  $|\Delta c_1(t)|/z_f$  (dotted, orange) and allele frequency (solid and dashed curves) for some loci that satisfy the necessary condition  $\gamma_i < -2\Delta c_1(0)$ , where  $\Delta c_1(0) = -2.99$  and  $z_f = 3$ . The numerical solution of the full model (6) (solid) and directional selection model (10) (dashed) are shown for the effect size  $\gamma_{i=1,\dots,6} = 0.776$  (gray), 0.340 (brown), 0.319 (red), 0.272 (magenta), 0.092 (blue), and 0.060 (black). The other parameters are  $s = 0.1$ ,  $\mu = 10^{-4}$ ,  $\bar{\gamma} = 0.2$ , and  $\ell = 20$ . Except for the locus with effect  $\gamma_4$ , the allele frequency for the directional selection model exceeds one half at loci whose allele frequency for the full model sweeps. The allele frequency for the locus with effect  $\gamma_6$  also increases nearly to fixation, but at very long times where the directional selection model is not valid. Because this frequency increase is relatively slow, a genomic signature similar to a selective sweep cannot be expected.

$$c_1(t) = \sum_{i=1}^{\ell} \gamma_i - \sum_{i \in P^+} \frac{2\gamma_i}{1 + \frac{p_i^+(0)}{q_i^+(0)} e^{\beta\gamma_i/\bar{\gamma}}} - \sum_{i \in P^-} \frac{2\gamma_i}{1 + \frac{p_i^-(0)}{q_i^-(0)} e^{\beta\gamma_i/\bar{\gamma}}}, \quad (22)$$

where  $P^\pm$  refers to the set of loci with initial frequencies  $p_i^\pm(0)$  given in (8). From symmetry considerations, we expect the number of loci with initial frequency above and below one half to be equal. Furthermore, for  $\gamma_i \gg \hat{\gamma}$ , the initial allele frequency given by (8) can be approximated as  $p_i^-(0) \approx (\hat{\gamma}/2\gamma_i)^2$ ,  $p_i^+(0) = 1 - p_i^-(0)$ . We thus arrive at

$$c_1(t) \approx \sum_{i=1}^{\ell} \gamma_i - \sum_{i \in P^+} \frac{2\gamma_i}{1 + \left(\frac{2\gamma_i}{\hat{\gamma}}\right)^2 e^{\beta\gamma_i/\bar{\gamma}}} - \sum_{i \in P^-} \frac{2\gamma_i}{1 + \left(\frac{\hat{\gamma}}{2\gamma_i}\right)^2 e^{\beta\gamma_i/\bar{\gamma}}}. \quad (23)$$

Then, as in the last section, approximating the sums in (23) by integrals for large  $\ell$ , we get

$$c_1(t) = \ell\bar{\gamma}(1 - I_l^+ - I_l^-), \quad (24)$$

where

$$I_l^+ = \frac{1}{2\bar{\gamma}} \int_{\hat{\gamma}}^{\infty} d\gamma \frac{2\gamma e^{-\gamma/\bar{\gamma}}}{1 + \left(\frac{2\gamma}{\hat{\gamma}}\right)^2 e^{\beta\gamma/\bar{\gamma}}} \quad (25)$$

$$= \int_{2/\alpha}^{\infty} dx \frac{x e^{-x}}{1 + (\alpha x)^2 e^{\beta x}}, \quad (26)$$

and

$$I_l^- = \frac{1}{2\bar{\gamma}} \int_{\hat{\gamma}}^{\infty} d\gamma \frac{2\gamma e^{-\gamma/\bar{\gamma}}}{1 + \left(\frac{\hat{\gamma}}{2\gamma}\right)^2 e^{\beta\gamma/\bar{\gamma}}} \quad (27)$$

$$= \int_{2/\alpha}^{\infty} dx \frac{x e^{-x}}{1 + (\alpha x)^{-2} e^{\beta x}}. \quad (28)$$

In the above equations, the parameter  $\alpha = 2\bar{\gamma}/\hat{\gamma}$ .

**Qualitative patterns:** In the directional selection model, as (12) and (13) hold irrespective of whether most effects are small or large, the mean deviation  $\Delta c_1(t)$  and the allele-frequency-dependent variable  $\beta$  increase with time as discussed in the last section, and the variance is given by  $c_2 = -\ell\bar{\gamma}^2(I_l^+ + I_l^-)$ . However, unlike when most effects are small, here the rate of change of variance,  $\dot{c}_2 = s\ell\bar{\gamma}^3\Delta c_1(I_l^+ + I_l^-)$ , is not a monotonic function of time; this is because the sum  $I_l^+ + I_l^-$  is negative when  $\beta$  is small and positive for larger  $\beta$ . Thus the directional selection model predicts that initially the variance *increases* with time, and then decreases toward the stationary state at the new phenotypic optimum.

**Quantitative dynamics:** Equation 24, along with (12), forms a closed system of equations and allows us to find the mean. The integrals  $I_l^\pm$  are estimated in Appendix F; we find that the integral  $I_l^+$  does not contribute substantially to the mean since it includes the loci with initial frequency close to one and we are dealing with the case when the initial mean deviation is negative [see also (32) below], and  $I_l^-$  is well approximated by  $J_l$  given by (F8). When the initial deviation from the phenotypic optimum is small, using (F14) and (F21), we find that at large times the mean deviation decreases exponentially fast,

$$\Delta c_1(t) \sim e^{-s\ell\bar{\gamma}^2(\ln \alpha)^{-1}\rho(\ln \rho)^2 t} \quad (29)$$

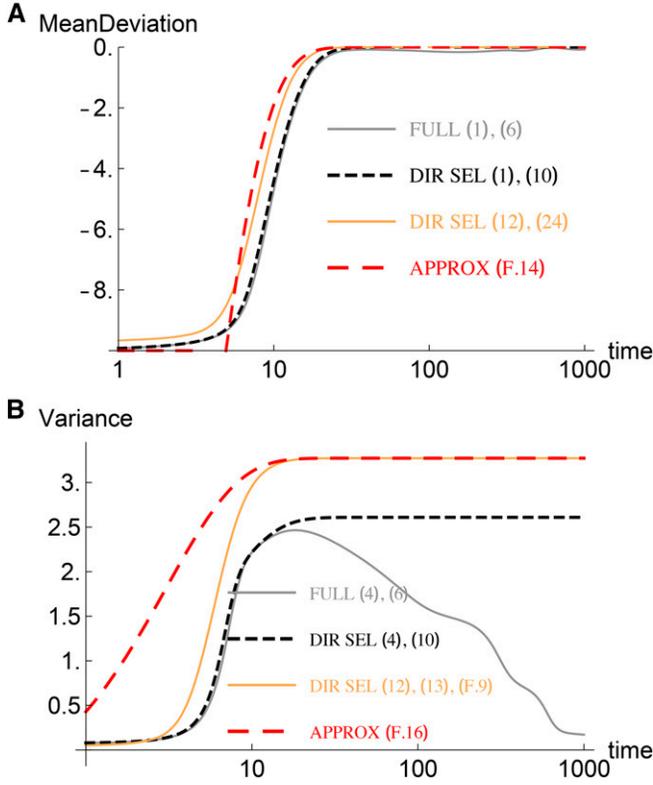
$$= e^{-s z_f \bar{\gamma} (\ln \alpha)^{-1} (\ln \rho)^2 t}, \quad (30)$$

where  $\rho = z_f/(\ell\bar{\gamma})$  is the phenotypic shift relative to the maximum mean magnitude. The above result shows that the timescale over which the mean approaches the stationary state depends weakly on the number of loci, unlike the case when most effects are small [see (18)]. However as the mean of the distribution is large here, the relaxation in the directional selection model occurs faster than in the small-effects case.

When most effects are large, the variance changes considerably, even when the relative phenotypic shift is quite small as Figure 4B shows. The peak variance  $c_2^{(\max)}$  can be estimated by the stationary state variance in the directional selection model. At large times, using (F16) and (F21), we find that

$$c_2^{(\max)} = \ell\bar{\gamma}^2(\ln \alpha)^{-1}\rho(\ln \rho)^2 \quad (31)$$

and therefore  $c_2^{(\max)}/c_2(0) \sim \alpha^2 \gg 1$ . This result can also be seen by noting that the allele frequencies at loci with an initial



**Figure 4** Most effects are large: Dynamics of the (A) mean deviation and (B) variance for  $\hat{\gamma} \approx 0.028$ ,  $\bar{\gamma} = 10\hat{\gamma}$ , and  $z_f = 10 \approx 0.18\ell\bar{\gamma}$ . For the realization of effects used in this figure, 183 out of  $\ell = 200$  loci have effects larger than the mean. The other parameters are  $s = 0.1$  and  $\mu = 10^{-5}$ .

frequency close to zero shift to intermediate values at short times. Then, from (4), the maximum variance is obtained when the allele frequency is close to one half leading to  $c_2^{(\max)} \sim \ell\bar{\gamma}^2$ .

The dynamics of the allele frequency at long times are discussed in Appendix E and at short times in the next section.

#### When do selective sweeps occur?

Below we obtain a criterion on the size of effects for which the allele frequency at a major locus can sweep. When  $p_i(0) \approx 0$ , at short enough times, keeping only the lowest order terms in  $p_i$  on the RHS of the full model (6) and neglecting the effect of mutations, we get

$$\dot{p}_i \approx -\frac{s\gamma_i p_i}{2} [2\Delta c_1(0) + \gamma_i]. \quad (32)$$

When the initial mean deviation  $\Delta c_1(0) < 0$ , the above equation shows that only the loci with effect  $\gamma_i < 2|\Delta c_1(0)|$  can potentially sweep, since their frequency increases with time. On repeating the above analysis for loci with initial frequency close to one, we find that the frequency at such loci does not sweep to zero. Thus, when  $\Delta c_1(0) < 0$ , only the loci with small initial frequency and effect  $< 2|\Delta c_1(0)|$  are likely to undergo large changes in frequency. These criteria are necessary, but as Figure 3 shows, they are not sufficient for a selective sweep to occur.

For the major loci that meet the above necessary conditions for selective sweeps, the second and the third term on the RHS of the full model (6) can be neglected in comparison to the first term, thereby reducing (6) to the directional selection model (10). For the rest of the (major) loci, the allele frequency changes are not appreciable and the frequencies remain close to zero or one, a solution satisfied by (10). Let  $\tilde{t}$  denote the time when the allele frequencies evolving according to the directional selection model equilibrate. Figure 3 shows that the mean deviation obtained using the full model is close to zero at time  $\tilde{t}$ . Then for subsequent times  $t > \tilde{t}$ , we can ignore mutations and set  $\Delta c_1 = 0$  in the full model (6) to obtain (Jain and Stephan 2015)

$$\dot{p}_i = \frac{s\gamma_i^2}{2} p_i q_i (2p_i - 1), t > \tilde{t}. \quad (33)$$

The above differential equation is subject to the initial condition  $p_i(\tilde{t})$ , which is the allele frequency in the steady state of the directional selection model.

From the definition (C6) of  $\beta$ , we have

$$\frac{p_i(\tilde{t})}{1 - p_i(\tilde{t})} = \frac{p_i(0)}{q_i(0)} \exp\left(\frac{\gamma_i \tilde{\beta}}{\bar{\gamma}}\right) \quad (34)$$

which simplifies to give the frequency  $p_i(\tilde{t})$  as

$$p_i(\tilde{t}) = \left[ 1 + \frac{q_i(0)}{p_i(0)} \exp\left(-\frac{\gamma_i \tilde{\beta}}{\bar{\gamma}}\right) \right]^{-1}, \quad (35)$$

where  $\tilde{\beta} \equiv \beta(\tilde{t})$ . It is clear from (33) that if the frequency  $p_i(\tilde{t})$  is greater than one half, the allele frequency at the  $i$ th locus will increase monotonically toward one. Thus we predict that the allele frequency at a locus would sweep when  $p_i(\tilde{t}) > 1/2$ . Combining this criterion with the necessary condition mentioned above, we find that a sweep can occur when the effects lie in the following range:

$$\frac{\bar{\gamma}}{\tilde{\beta}} \ln \left[ \frac{q_i^-(0)}{p_i^-(0)} \right] < \gamma_i < 2z_f. \quad (36)$$

This expectation is indeed seen to be consistent with the numerical solution of the full model as shown in Figure 3, except for one case.

**When most effects are small:** While considering the dynamics of mean and variance when most effects are small in an earlier section, we ignored the contribution of (few) loci that have large effect. However, the results obtained earlier can be used to understand the dynamics of the allele frequency at major loci too. As (D11) shows, the auxiliary parameter in the stationary state of the directional selection model is  $\tilde{\beta} = z_f / (\ell\bar{\gamma})$ . Using this in (36), we find that a selective sweep can occur at a major locus if its effect

$$2z_f > \gamma_i > \frac{\ell\bar{\gamma}^2}{z_f} \ln \left[ \frac{q_i^-(0)}{p_i^-(0)} \right], \quad (37)$$

**Table 1** Summary of the results for the dynamics of phenotypic mean, variance, and allele frequency when the phenotypic optimum is suddenly shifted

	Mean dynamics determined by	Variance dynamics characterized by	Selective sweeps at major loci
Small effects	Initial variance (18)	Small variations (19)	Unlikely (39)
Large effects	Effect size (30)	Large variations (31)	Probable (41)

$$= \frac{c_2(0)}{|\Delta c_1(0)|} \ln \left[ \frac{q_i^-(0)}{p_i^-(0)} \right], \quad (38)$$

which matches the criterion (26b) of Chevin and Hospital (2008). Furthermore, using (8), we find that

$$\frac{2c_2(0)}{|\Delta c_1(0)|} \ln \left( \frac{2\gamma_i}{\hat{\gamma}} \right) < \gamma_i < 2z_f. \quad (39)$$

For the parameters in Figure 1 where the average effect size  $\bar{\gamma} = 0.04$  and the threshold effect  $\hat{\gamma} = 0.128$ , the above criterion requires that the effect size at a major locus be  $> 0.81$ . For exponentially distributed effects, as is assumed in this article, the probability for the effect size to exceed this criterion is extremely low ( $\sim 10^{-9}$  for the parameters in Figure 1) and therefore selective sweeps at major loci are prevented when most effects are small (Chevin and Hospital 2008; Pavlidis *et al.* 2012).

**When most effects are large:** The criterion for short-time sweeps is more involved in this case; using (8) in (36), we obtain

$$2z_f > \gamma_i > \frac{2\bar{\gamma}}{\tilde{\beta}} \ln \left( \frac{2\gamma_i}{\hat{\gamma}} \right), \quad (40)$$

$$\approx \frac{2\bar{\gamma}}{\ln \left( \frac{2\bar{\gamma}}{\hat{\gamma}} \right)} \ln \left( \frac{\ell \bar{\gamma}}{z_f} \right) \ln \left( \frac{2\gamma_i}{\hat{\gamma}} \right), \quad (41)$$

where we have used (F20) for  $\tilde{\beta}$  in the last expression. For the parameters in Figure 4, using  $\tilde{\beta} \approx 2.32$  in (40), we find that for a selective sweep to occur,  $\gamma_i > 1.04$  is required whose probability is  $\approx 0.024$ . In other words, for these parameter values, we expect as many as  $0.024 \times 200 \approx 5$  sweeps to occur over the timescale the mean deviation approaches zero.

#### Data availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

## Discussion

### Dynamic regimes in the full model

Figure 3 shows an example of the dynamics of the mean and allele frequencies when an infinitely large population in LE and equilibrated to a phenotypic optimum is suddenly shifted

to a new optimum. We observe a *short-term phase* ( $t \lesssim 10^3$ ), where the mean deviation increases quickly to a value close to zero; followed by a *long-term phase* ( $10^3 < t < 10^5$ ), where the mean deviation undergoes minor changes at a slow pace. The former phase where the adaptation process is rapid is the focus of this article since (i) much of the action happens in this early phase (for example, in Figure 3, the allele frequencies at 13 out of 20 loci were close to the stationary state at the end of the short-term phase), (ii) these timescales are experimentally observable, and very importantly, (iii) all + alleles increase in frequency in a synergistic manner.

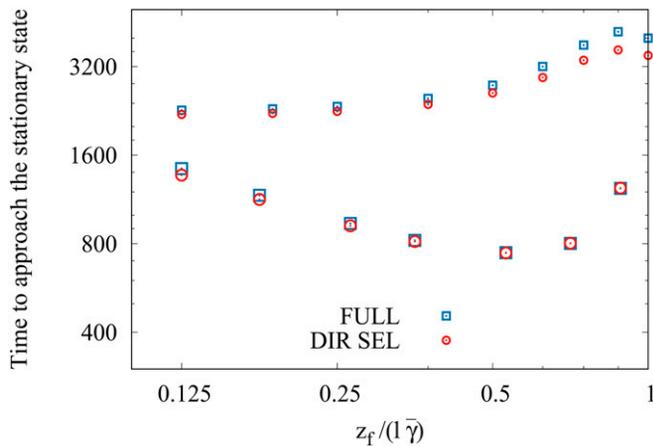
### Directional selection model

Although the full model defined by (6) can be investigated numerically (de Vladar and Barton 2014), as the allele frequency dynamics at a single locus are related to all the other  $\ell - 1$  loci via the phenotypic mean [first term on the RHS of (6)], it is difficult to obtain analytical results. Moreover, the full model results in equations for the cumulants that do not close, and therefore one resorts to an often unjustified and uncontrolled approximation of setting all the cumulants above a certain number to zero (see Barton and de Vladar 2009 for a discussion of this approach). This approximation, however, works quite well when most effects are small (Jain and Stephan 2015) but fails completely when most effects are large (see Supplemental Material, File S1 for a detailed explanation).

Here we have devised an approximate model (directional selection model) defined by (10) in which the allele frequencies are coupled but, nonetheless, one can make analytical progress without truncating the cumulants arbitrarily. The central equations in the directional selection model are (11) and (12) that relate the mean deviation and an auxiliary parameter  $\beta$  which is a function of a single allele frequency. As explained in Appendix D and Appendix F, it is possible to obtain explicit expressions for the time dependence of  $\beta$  and thereby that of the mean. Higher cumulants such as the variance can be then found using (13).

### Distribution of effects

All the numerical examples shown in the figures use a single realization of the effect at a locus, *i.e.*, one set of effects  $\{\gamma_i\}, i = 1, \dots, \ell$ , are chosen from an exponential distribution. For additive quantities, such as mean and variance, that are obtained by summing the contribution over a number of loci, we expect that a single realization of effects is a good representative when the number of loci are sufficiently large (in physics literature, this concept is known as self-averaging; Castellani and Cavagna 2005). A similar idea has been used



**Figure 5** Small vs. large effects when the number of loci is the same: Time at which the mean reaches 0.9 of the exact stationary state mean  $c_1^*$  for various values of  $z_f$  when most effects are small (small symbols) and large (large symbols). For the former (latter) case, the rest of the parameters are the same as in Figure 1 (Figure 4). To show the data in the two cases on the same scale, the time for the large-effect case has been multiplied by 70.

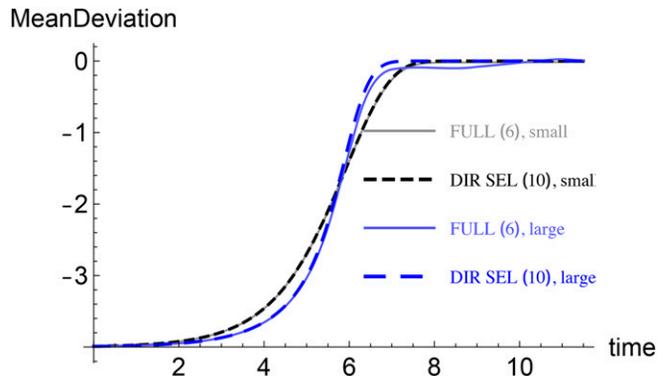
recently in a model of stabilizing selection with the same effects at all loci in which the phenotypic mean for one set of initial allele frequencies is approximated by the corresponding average (Charlesworth 2013). Figure S1 in File S1 shows that the quantitative agreement between the results for the mean deviation obtained for a single realization of phenotypic effects for  $\ell = 50, 200, 400$ , and the analytical result for large  $\ell$  gets better as the number of loci increases.

### Main results

The key conclusions of this article are discussed below and summarized in Table 1.

**Response of the mean to a sudden shift of the phenotypic optimum:** How does the timescale over which the mean reaches the stationary state depend on various factors such as the number of loci, initial phenotypic mean deviation, and size of the effects? We find that the mean deviation becomes negligible on a timescale that is inversely proportional to the number of loci ( $\ell$ ) when most effects are small, but depends weakly on  $\ell$  when most effects are large [cf. (18) and (30)]. This difference in the behavior may be understood as follows: when most effects are small, there is sufficient genetic variation initially for selection to act on because a large number of loci are initially polymorphic. Therefore, with increasing  $\ell$ , the equilibration time decreases. In contrast, when most effects are large, the initial genetic variation is negligible and the adaptation dynamics are determined by the size of the effects. As the availability of large-effect loci depends on the mean of the effect distribution, it follows that the larger the mean, the faster the adaptation dynamics.

Figure 5 shows that the relationship between the equilibration time and the initial mean deviation depends on the size of effects (see also Gomulkiewicz *et al.* 2010). When



**Figure 6** Many loci with small effects vs. few loci with large effects: Mean deviation as a function of time when most effects are small ( $\bar{\gamma} = 0.04$ ,  $\ell = 400$ ) and large ( $\bar{\gamma} = 0.32$ ,  $\ell = 50$ ). In both cases,  $\hat{\gamma} \approx 0.126$  ( $s = 5 \times 10^{-3}$ ,  $\mu = 10^{-5}$ ) and the new phenotypic optimum is shifted to  $z_f = \ell \bar{\gamma} / 4 = 4$ .

most effects are large, selective sweeps occur at some loci that result in drastic changes in the allele frequencies, thus accelerating the approach to the phenotypic optimum. But when most effects are small, the equilibration time remains roughly constant as selective sweeps do not occur, but small changes in the allele frequencies at a large number of loci drive the adaptation process. Figure 5 also shows that when the number of loci contributing to a trait are the same, adaptation is faster when the effect size is larger, which makes intuitive sense.

However, one may also compare situations in which a large number of small effects control a phenotypic trait with the one in which few loci with large effects contribute to a trait. An example shown in Figure 6 suggests that the adaptation at short times is faster when effects are small, but at longer times, larger effects lead to rapid adaptation. This conclusion is borne out by our analysis as well: on comparing (18) and (30), we find that for identical  $z_f$  and  $\ell \bar{\gamma}$ , the timescales are inversely proportional to  $\bar{\gamma}$  and therefore it takes a shorter time to equilibrate in the large-effect case.

An important omission in the above discussion is random genetic drift as we have considered infinitely large populations. Some progress in this regard has been made recently (Matuszewski *et al.* 2015; Bod'ová *et al.* 2016); however, a detailed analysis is currently not available and would be interesting to consider in the future.

**Dynamics of the variance and allele frequencies:** Besides the dynamics of the mean, we also studied the time dependence of the variance. We find that at short times, the variance does not vary much and decreases monotonically with time when most effects are small. This behavior can be explained by the fact that the allele frequencies show small to moderate changes. In the opposite case, the variance changes considerably and its variation is nonmonotonic in time: it increases at short times and then decreases to the stationary state value. The selective sweeps at major loci are responsible for this effect. The rate of change of average fitness given by (C9)

receives contributions from the square of the mean deviation and the rate of change of variance. Thus when most effects are large, at short times, the average fitness increases more slowly in the short-term phase than at longer times (Gomulkiewicz *et al.* 2010).

The directional selection model describes the bulk of the mean dynamics in the short-term phase. Therefore, at late times where the mean deviation is close to zero, the allele frequency at each locus evolves independently [the first term on the RHS of (6) is zero] and the dynamics are much slower, as described in Appendix E. At short times, the allele frequency dynamics involve the same timescales as for the mean and variance.

### Applications

The analysis of our model relates to various important questions that are currently discussed in evolutionary genetics.

**Rapid adaptation:** The approximations we have presented here describe the short-term evolution of a phenotypic trait after a sudden environmental change very well. We have shown that the mean of a phenotypic trait may respond very quickly after an environmental shift. In the case that most effects are small, this is possible because the time to the new optimum is inversely proportional to the number of loci controlling the trait. In the opposite case of mostly large effects, rapid fixations (leading to selective sweeps) may produce fast phenotypic responses. In the examples of fast adaptation mentioned in the Introduction, both of these extremes and combinations thereof have occurred.

**QTL and selective sweep mapping:** Selective sweep mapping has been used to dissect QTL (Svetec *et al.* 2011; Rubin *et al.* 2012; Axelsson *et al.* 2013; Wilches *et al.* 2014). However, the success of this approach was varied. Nonetheless, there seems to be a tendency that it worked better in domesticated than in natural populations (Rubin *et al.* 2012; Axelsson *et al.* 2013), probably due to the action of artificial selection during domestication. In artificial selection, the shift to the new optimum  $z_f$  may be larger than under natural conditions; our criterion (36) thus indicates an enhancement of sweeps in domesticated populations.

More generally, our analysis also provides some new insights into the question whether selective fixations (and thus sweeps) occur at QTL. While Chevin and Hospital (2008) predicted that the probability of selective sweeps is extremely low at QTL (based on a model with one major locus and infinitely many minor loci), others have found sweeps at appreciable frequencies using simulations of various multi-locus models (Pavlidis *et al.* 2012; Wollstein and Stephan 2014). The prediction of Chevin and Hospital is consistent with our study for mostly small-effect loci and small shifts in the phenotypic optimum.

**Testing for fast polygenic adaptation:** Using standardized frequencies, one can construct tests to detect SNPs that deviate strongly from a neutral population structure (e.g., Coop

*et al.* 2010). However, this approach only works if there are relatively large extended gradients of ecological variables, which may not be the case in rapid adaptation. Shortly after a population occupies a new habitat, we expect that the allele frequency shifts between the parental and derived populations are relatively small. This also means that available software, such as BayeScan-like methods (Foll and Gaggiotti 2008; Riebler *et al.* 2008), is not able to detect significant frequency shifts for individual SNPs between populations. Therefore, for detecting small allele frequency shifts after environmental changes in fast adapting populations, it may be best to consider the frequency shifts of alleles simultaneously at all loci involved (instead of individual SNPs). Our results suggest that this may be a promising approach, as all + alleles shift their frequencies in the same direction in the short-term phase, which should increase the power of the test. In human population genetics, this approach has been used to analyze recent polygenic adaptation of height, for instance (Turchin *et al.* 2012).

### Acknowledgments

We thank the International Centre for Theoretical Sciences, Bangalore, for their hospitality during the Second Bangalore School on Population Genetics and Evolution (ICTS/Progpopgen/2016/01) that facilitated our discussions. Furthermore, we are grateful to Graham Coop and two anonymous reviewers who made valuable suggestions to improve the article. The research of W.S. was supported by the Deutsche Forschungsgemeinschaft (grant STE 325/17-1 from the Priority Program 1819).

### Literature Cited

- Abramowitz, M., and I. A. Stegun (Editors), 1964 *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.
- Arfken, G., 1985 *Mathematical Methods for Physicists*. Academic Press, New York.
- Axelsson, E., A. Ratnakumar, M. L. Arendt, K. Maqbool, M. T. Webster *et al.*, 2013 The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495: 360–364.
- Barton, N. H., 1986 The maintenance of polygenic variation through a balance between mutation and stabilizing selection. *Genet. Res.* 47: 209–216.
- Barton, N. H., and H. P. de Vladar, 2009 Statistical mechanics and the evolution of polygenic quantitative traits. *Genetics* 181: 997–1011.
- Barton, N. H., and M. Turelli, 1991 Natural and sexual selection on many loci. *Genetics* 127: 229–255.
- Boďová, K., G. Tkacik, and N. H. Barton, 2016 A general approximation for the dynamics of quantitative traits. *Genetics* 202: 1523–1548.
- Bürger, R., 1991 Moments, cumulants, and polygenic dynamics. *J. Math. Biol.* 30: 199–213.
- Bürger, R., 2000 *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Chichester, United Kingdom.
- Castellani, T., and A. Cavagna, 2005 Spin-glass theory for pedestrians. *J. Stat. Mech. Theor. Exp.* 2005: P05012.

- Charlesworth, B., 2013 Stabilizing selection, purifying selection, and mutational bias in finite populations. *Genetics* 194: 955–971.
- Charlesworth, B., and D. Charlesworth, 2010 *Elements of Evolutionary Genetics*. Roberts and Company Publishers, Greenwood Village, CO.
- 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.
- Cook, L. M., B. S. Grant, I. J. Saccheri, and J. Mallet, 2012 Selective bird predation on the peppered moth: the last experiment of Michael Majerus. *Biol. Lett.* 8: 609–612.
- Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard, 2010 Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185: 1411–1423.
- de Vladar, H. P., and N. H. Barton, 2014 Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics* 197: 749–767.
- French-Constant, R. H., M. Bogwitz, P. Daborn, and J. Yen, 2002 A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* 27: 2253–2256.
- Foll, M., and O. Gaggiotti, 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180: 977–993.
- Goddard, M. E., and B. J. Hayes, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10: 381–391.
- Gomulkiewicz, R., R. D. Holt, M. Barfield, and S. L. Nuismer, 2010 Genetics, adaptation, and invasion in harsh environments. *Evol. Appl.* 3: 97–108.
- Grant, P. R., and B. R. Grant, 2008 *How and Why Species Multiply: The Radiation of Darwin's Finches*. Princeton University Press, Princeton, NJ.
- Jain, K., and W. Stephan, 2015 Response of polygenic traits under stabilising selection and mutation when loci have unequal effects. *G3 (Bethesda)* 5: 1065–1074.
- Kaplan, N. L., R. R. Hudson, and C. H. Langley, 1989 The “hitchhiking effect” revisited. *Genetics* 123: 887–899.
- Kim, Y., and W. Stephan, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160: 765–777.
- Lamichhaney, S., A. Martinez Barrio, N. Rafati, G. Sundström, C. J. Rubin *et al.*, 2012 Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proc. Natl. Acad. Sci. USA* 109: 19345–19350.
- Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr *et al.*, 2015 Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518: 371–375.
- Lande, R., 1976 Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30: 314–334.
- Lande, R., 1983 The response to selection on major and minor mutations affecting a metrical trait. *Heredity* 50: 47–65.
- Linnen, C. R., Y. P. Poh, B. K. Peterson, R. D. Barrett, J. G. Larson *et al.*, 2013 Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* 339: 1312–1316.
- Losos, J. B., 2009 *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles*. University of California Press, San Francisco, CA.
- Losos, J. B., 2014 What Darwin got wrong. *Chron. High. Educ.* B13–B15.
- Mackay, T. F. C., 2004 The genetic architecture of quantitative traits: lessons from *Drosophila*. *Curr. Opin. Genet. Dev.* 14: 253–257.
- Matuszewski, S., J. Hermisson, and M. Kopp, 2015 Catch me if you can: adaptation from standing genetic variation to a moving phenotypic optimum. *Genetics* 200: 1255–1274.
- Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. *Genet. Res.* 23: 23–35.
- 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.
- Pavlidis, P., J. D. Jensen, and W. Stephan, 2010 Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics* 185: 907–922.
- Pavlidis, P., D. Metzler, and W. Stephan, 2012 Selective sweeps in multilocus models of quantitative traits. *Genetics* 192: 225–239.
- Pritchard, J. K., and A. Di Rienzo, 2010 Adaptation - not by sweeps alone. *Nat. Rev. Genet.* 11: 665–667.
- Reznick, D. N., 2009 *The Origin Then and Now: An Interpretive Guide to the Origin of Species*. Princeton University Press, Princeton, NJ.
- Riebler, A., L. Held, and W. Stephan, 2008 Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics* 178: 1817–1829.
- Rubin, C. J., H. J. Megens, A. Martinez Barrio, K. Maqbool, S. Sayyab *et al.*, 2012 Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. USA* 109: 19529–19536.
- Sornette, D., 2000 *Critical Phenomena in Natural Sciences*. Springer, Berlin.
- 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.
- Svetec, N., A. Werzner, R. Wilches, P. Pavlidis, J. M. Alvarez-Castro *et al.*, 2011 Identification of X-linked quantitative trait loci affecting cold tolerance in *Drosophila melanogaster* and fine mapping by selective sweep analysis. *Mol. Ecol.* 20: 530–544.
- Turchin, M. C., C. W. K. Chiang, C. D. Palmer, S. Sankaraman, D. Reich *et al.*, 2012 Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* 44: 1015–1019.
- van't Hof, A. E., N. Edmonds, M. Daliková, F. Marec, and I. J. Saccheri, 2011 Industrial melanism in British peppered moths has a singular and recent mutational origin. *Science* 332: 958–960.
- Vignieri, S. N., J. G. Larson, and H. E. Hoekstra, 2010 The selective advantage of crypsis in mice. *Evolution* 64: 2153–2158.
- Visscher, P. M., M. A. Brown, M. I. McCarthy, and J. Yang, 2012 Five years of GWAS discovery. *Am. J. Hum. Genet.* 90: 7–24.
- Wilches, R., S. Voigt, P. Duchon, S. Laurent, and W. Stephan, 2014 Fine-mapping and selective sweep analysis of QTL for cold tolerance in *Drosophila melanogaster*. *G3 (Bethesda)* 4: 1635–1645.
- Wollstein, A., and W. Stephan, 2014 Adaptive fixation in two-locus models of stabilizing selection and genetic drift. *Genetics* 198: 685–697.
- Wright, S., 1935 Evolution in populations in approximate equilibrium. *J. Genet.* 30: 257–266.

Communicating editor: G. Coop

## Appendix A: Cumulants of the Effects and Their Dynamics

The effect  $e_i$  at the locus  $i$  is distributed according to a Bernoulli distribution

$$p(e_i) = p_i \delta_{e_i, \frac{\gamma_i}{2}} + q_i \delta_{e_i, -\frac{\gamma_i}{2}}. \quad (\text{A1})$$

The generating function of the distribution of the effects at the  $i$ th locus is given by

$$F_i(\xi) = \sum_{e_i} p(e_i) e^{-e_i \xi} = p_i e^{-\frac{\gamma_i}{2} \xi} + q_i e^{\frac{\gamma_i}{2} \xi}. \quad (\text{A2})$$

The logarithm of the generating function defines the  $n$ th cumulant  $c_n^{(i)}$  at locus  $i$  as (Sornette 2000)

$$\ln F_i(\xi) = \sum_{n=1}^{\infty} c_n^{(i)} \frac{(-\xi)^n}{n!} = \frac{\xi \gamma_i}{2} + \ln(q_i + p_i e^{-\gamma_i \xi}). \quad (\text{A3})$$

The  $n$ th cumulant is then obtained as

$$c_n^{(i)} = (-1)^n \left. \frac{d^n \ln F_i(\xi)}{d\xi^n} \right|_{\xi=0}. \quad (\text{A4})$$

In LE, since the cumulants for the  $n$ -locus problem are additive, we get

$$c_n = 2 \sum_{i=1}^{\ell} c_n^{(i)}, \quad (\text{A5})$$

where the factor of two on the RHS takes care of diploidy. Using the last three equations above, we find that the mean and variance are given by (1) and (4), respectively. The skewness  $c_3$ , which is a measure of asymmetry of a distribution, can also be found and given by

$$c_3 = 2 \sum_{i=1}^{\ell} \gamma_i^3 (q_i - p_i) p_i q_i. \quad (\text{A6})$$

## Appendix B: Quasi-LE Approximation

For the additive genotype-phenotype map, the trait  $z = \sum_{i=1}^{\ell} \gamma_i (X_i + X_i^* - 1)$  where  $X_i, X_i^* = 1(0)$  if the  $+( -)$  allele is present at the locus  $i$ . Following Barton and Turelli (1991), we first find the association coefficients by writing  $X_i = \langle X_i \rangle + \delta_i, X_i^* = \langle X_i^* \rangle + \delta_i^*$ , where  $\langle X_i \rangle = \langle X_i^* \rangle = p_i$  and expanding the relative fitness to lowest orders in  $s$ ,

$$\begin{aligned} \frac{w(\{X, X^*\})}{\bar{w}} &= \frac{1 - (s/2)(z - z')^2}{\bar{w}} \quad (\text{B1}) \\ &\approx 1 + \sum_{i=1}^{\ell} a_{i,\phi} (\delta_i - \langle \delta_i \rangle) + \sum_{i=1}^{\ell} a_{ii,\phi} (\delta_i^2 - \langle \delta_i^2 \rangle) \\ &\quad + \sum_{i=1}^{\ell} a_{i,\phi}^* (\delta_i^* - \langle \delta_i^* \rangle) + \sum_{i=1}^{\ell} a_{ii,\phi}^* (\delta_i^{*2} - \langle \delta_i^{*2} \rangle) \\ &\quad + \sum_{i < j} a_{ij,\phi} (\delta_i \delta_j - \langle \delta_i \delta_j \rangle) + \sum_{i < j} a_{ij,\phi}^* (\delta_i^* \delta_j^* - \langle \delta_i^* \delta_j^* \rangle) \end{aligned}$$

$$+ \sum_{i,j=1}^{\ell} a_{ij} \left( \delta_i \delta_j^* - \langle \delta_i \delta_j^* \rangle \right). \quad (\text{B2})$$

In the above equation, the average is taken with respect to the joint distribution of the frequencies at all the loci and the coefficients

$$a_{i,\phi} = -s(c_1 - z') \gamma_i, a_{ii,\phi} = -(s/2) \gamma_i^2, a_{ij,\phi} = a_{ij} = -s \gamma_i \gamma_j. \quad (\text{B3})$$

The average allele frequency change after selection (and recombination) is given by [see (10a) in Barton and Turelli 1991]

$$\dot{p}_i = a_{i,\phi} \langle \delta_i^2 \rangle + a_{ii,\phi} \langle \delta_i^3 \rangle + \sum_{j \neq i} a_{j,\phi} \langle \delta_i \delta_j \rangle + \sum_{j \neq i} a_{ji,\phi} \langle \delta_i \delta_j^2 \rangle + \sum_{j \neq k} a_{kj,\phi} \langle \delta_i \delta_j \delta_k \rangle. \quad (\text{B4})$$

All but the first two terms on the RHS vanish in LE; here we focus on the lowest order corrections to LE which are contained in the third and fourth terms. Using  $X_i^2 = X_i$  and  $\langle \delta_i \rangle = 0$ , it is easy to show that  $\langle \delta_i^2 \rangle = p_i q_i$ ,  $\langle \delta_i \delta_j^2 \rangle = (1 - 2p_j) \langle \delta_i \delta_j \rangle$ . Inserting these expressions in (B4), we verify that the first two terms on the RHS match the corresponding ones in (6). The two-locus linkage disequilibrium evolves according to [see (12) in Barton and Turelli 1991]

$$\langle \delta_i \delta_j \rangle = -r_{ij} \langle \delta_i \delta_j \rangle + 2a_{ij,\phi} \langle \delta_i^2 \rangle \langle \delta_j^2 \rangle + \text{higher order terms}, \quad i \neq j, \quad (\text{B5})$$

where  $r_{ij}$  denotes the recombination rate between the loci  $i$  and  $j$  and we have dropped terms that are nonlinear in the parameters  $s$  and  $r$  on the RHS of the above equation. Then at quasi-LE (QLE) where the LHS is of  $\mathcal{O}(s^2)$  or higher, we have

$$\langle \delta_i \delta_j \rangle_{\text{QLE}} \approx \frac{2a_{ij,\phi}}{r_{ij}} \langle \delta_i^2 \rangle_{\text{LE}} \langle \delta_j^2 \rangle_{\text{LE}} \quad (\text{B6})$$

$$= -\frac{2s \gamma_i \gamma_j}{r_{ij}} (p_i q_i p_j q_j) \Big|_{\text{LE}}, \quad (\text{B7})$$

where the allele frequencies in the last equation are in LE.

Using the above result in (B4) and assuming that all the recombination rates are equal to  $r$ , we find that within QLE, the allele frequency obeys the following equation:

$$\dot{p}_i \approx -s \gamma_i p_i q_i (c_1 - z') \left( 1 - \frac{s}{r} c_{2,\text{LE}} \right) - \frac{s \gamma_i^2}{2} p_i q_i (q_i - p_i) \left( 1 - \frac{s}{r} \frac{c_{3,\text{LE}}}{\gamma_i} \right), \quad (\text{B8})$$

where  $c_{2,\text{LE}}$  and  $c_{3,\text{LE}}$  are the variance and skewness, respectively, in LE. Thus, as expected, the corrections to the model defined by (6) are of order  $s/r$  which can be neglected when selection and linkage are weak.

### Appendix C: Directional Selection Model

The dynamics of the allele frequency  $p_i$  are described by (10), where the effective selection coefficient  $S_i = -s \gamma_i \Delta c_1$  depends on all the allele frequencies through the mean  $c_1$ . This property makes it difficult to solve for the allele frequencies; however, we note that for two arbitrary loci  $i$  and  $m$ , (10) gives

$$\frac{\dot{p}_i}{\dot{p}_m} = \frac{dp_i}{dp_m} = \frac{S_i p_i q_i}{S_m p_m q_m}, \quad (\text{C1})$$

which, on using the identity  $d[\ln(p/q)] = dp/(pq)$ , simplifies to

$$\frac{d \ln \varrho_i}{d \ln \varrho_m} = \frac{\gamma_i}{\gamma_m}, i \neq m, \quad (C2)$$

where  $\varrho_i(t) = p_i(t)/q_i(t)$ . As the above equation does not involve the phenotypic mean, it can be easily solved to give

$$\varrho_i(t) = \varrho_i(0) \left[ \frac{\varrho_m(t)}{\varrho_m(0)} \right]^{\frac{\gamma_i}{\gamma_m}}. \quad (C3)$$

We note that the simple idea used in (C1) to eliminate the global variable  $\Delta c_1$  is quite similar to that employed in single-locus directional selection models to get rid of the average fitness (see, e.g., Charlesworth and Charlesworth 2010, p. 73).

Equation C3 is useful since it allows us to express the mean  $c_1(t)$  in terms of a single allele frequency, say, at locus  $m$ . Using (C3) in (1), we can write the mean as

$$c_1(t) = \sum_{i=1}^{\ell} \gamma_i \left( \frac{\varrho_i - 1}{\varrho_i + 1} \right) \quad (C4)$$

$$= \sum_{i=1}^{\ell} \gamma_i - \sum_{i=1}^{\ell} \frac{2\gamma_i}{1 + \frac{p_i(0)}{q_i(0)} e^{\beta\gamma_i/\bar{\gamma}}}, \quad (C5)$$

which simplifies to give (11). In the above equation,

$$\beta(t) = \frac{\bar{\gamma}}{\gamma_m} \ln \left[ \frac{\varrho_m(t)}{\varrho_m(0)} \right] = \frac{\bar{\gamma}}{\gamma_m} \ln \left[ \frac{p_m(t) q_m(0)}{p_m(0) q_m(t)} \right]. \quad (C6)$$

Note that in the expression (C5), although the dynamics of the mean deviation are determined by that of the allele frequency  $p_m$ , the rest of the  $\ell - 1$  frequencies are not redundant as the mean also depends on their initial state.

Furthermore, from the evolution Equation 10 for the allele frequency, we also have  $\dot{\beta} = -s\bar{\gamma}\Delta c_1$ , where the mean is given by (C5); we have thus obtained a *closed* equation for  $\beta(t)$ . Using the result for  $\beta(t)$  in (C5), the dynamics of the phenotypic mean can be calculated. This knowledge then allows us to find higher cumulants using (13) which is derived below.

To obtain the evolution Equation 13 for the  $n$ th cumulant, we take the derivative with respect to time in (A3) and use (10) for the allele frequency dynamics to obtain (Bürger 1991)

$$\sum_{n=1}^{\infty} \dot{c}_n^{(i)} \frac{(-\xi)^n}{n!} = \frac{s\gamma_i p_i q_i \Delta c_1 (1 - e^{-\gamma_i \xi})}{q_i + p_i e^{-\gamma_i \xi}}. \quad (C7)$$

But, on taking a derivative of (A3) with respect to  $\xi$ , after some simple algebra, we also have

$$\sum_{n=1}^{\infty} c_{n+1}^{(i)} \frac{(-\xi)^n}{n!} = -\frac{\gamma_i p_i q_i (1 - e^{-\gamma_i \xi})}{q_i + p_i e^{-\gamma_i \xi}}. \quad (C8)$$

On comparing the last two equations and using (A5), we arrive at the promised result (13). The rate of change of the logarithm of the average fitness can be found using the above results and given by

$$\frac{d \ln \bar{w}}{dt} = \frac{\dot{\bar{w}}}{\bar{w}} = -\frac{s}{2} \dot{c}_2 + c_2 (s\Delta c_1)^2. \quad (C9)$$

## Appendix D: Cumulant Dynamics When Most Effects Are Small

Here we study the differential Equation 12 obeyed by  $\beta$ ,

$$\dot{\beta} = -s\ell\bar{\gamma}^2[1 - \rho - 2I_s(\beta)], \quad (\text{D1})$$

where  $\rho = z_f/\ell\bar{\gamma}$  and the integral

$$I_s = \int_0^{\hat{\gamma}/\bar{\gamma}} dx \frac{xe^{-x}}{1 + e^{\beta x}} \quad (\text{D2})$$

$$= \int_0^\infty dx \frac{xe^{-x}}{1 + e^{\beta x}} - \int_{\hat{\gamma}/\bar{\gamma}}^\infty dx \frac{xe^{-x}}{1 + e^{\beta x}}. \quad (\text{D3})$$

The first integral on the RHS is exactly solvable in terms of special functions and we have

$$I_s^{(1)} = \int_0^\infty dx \frac{xe^{-x}}{1 + e^{\beta x}} \quad (\text{D4})$$

$$= \Omega^2 \left[ \psi^{(1)}\left(\Omega + \frac{1}{2}\right) - \psi^{(1)}(\Omega + 1) \right] \quad (\text{D5})$$

$$= 1 + \frac{2\psi^{(1)}(2\Omega) - \psi^{(1)}(\Omega)}{2\beta^2}, \quad (\text{D6})$$

where  $\Omega = (2\beta)^{-1}$  and  $\psi^{(1)}(z)$  are the second derivatives of the logarithm of the gamma function [(6.4.1) in Abramowitz and Stegun 1964]. Since  $\bar{\gamma} \ll \hat{\gamma}$  when most effects are small, the second integral on the RHS of (D3) can be estimated by carrying out an integration by parts and we have

$$I_s^{(2)} \approx \frac{\tilde{x}e^{-\tilde{x}}}{1 + e^{\beta\tilde{x}}}, \quad (\text{D7})$$

where  $\tilde{x} = \hat{\gamma}/\bar{\gamma}$ . The above integral is thus exponentially small in  $\tilde{x}$  and can be neglected in comparison to  $I_s^{(1)}$ . Using the series expansion for the polygamma function  $\psi^{(1)}(z)$  for small arguments [(6.4.10) and (6.4.11) in Abramowitz and Stegun 1964], we obtain

$$I_s \approx \begin{cases} \frac{1 - \beta + \beta^3}{2} + \dots, & \beta \ll 1 \quad (\text{D8}) \\ \frac{\pi^2}{12\beta^2} + \mathcal{O}(\beta^{-3}), & \beta \gg 1. \quad (\text{D9}) \end{cases}$$

As both the initial and stationary state frequencies are close to one half, the function  $\beta(t)$  given by (C6) stays close to zero. Using the result (D8) in the Equation D1 for  $\beta(t)$ , we get

$$\frac{d\beta}{d\tau} \approx -\beta + \beta^3 + \rho, \quad (\text{D10})$$

where  $\tau = s\ell\bar{\gamma}^2 t$ . As  $\beta \ll 1$ , ignoring the term  $\beta^3$  on the RHS of the above equation and integrating, we obtain the zeroth order solution  $\beta_0$  given by

$$\beta_0(\tau) = \rho(1 - e^{-\tau}), \quad (\text{D11})$$

where we have used that  $\beta(0) = 0$ . An approximate solution of (D10) can be found iteratively by writing  $\beta = \beta_0 + \beta_1$  in (D10) and retaining the leading order term in  $\beta_1$  to obtain

$$\frac{d\beta_1}{d\tau} + \beta_1 \approx \beta_0^3, \quad (\text{D12})$$

which yields  $\beta \approx \beta_0 + \rho^{-1}\beta_0^4$ . Using this result in (D8), we get  $2I_s \approx 1 - \beta_0 + \beta_0^3 e^{-\tau}$ . Equations 15 and 13 then yield the first two cumulants to be

$$\Delta c_1(\tau) = -z_f e^{-\tau} (1 + \rho^{-1}\beta_0^3) \quad (\text{D13})$$

and

$$c_2(\tau) = \tau \bar{\gamma}^2 (1 - 3\beta_0^2). \quad (\text{D14})$$

### Appendix E: Long-Time Dynamics of the Allele Frequency

When most effects are small, as shown in Figure 1A, the mean deviation is close to zero when  $t \approx 10^4$ . But the exact stationary state for the mean and the allele frequency (shown in Figure 2) is reached much later at  $t \approx 10^5$ . A similar qualitative behavior of the allele frequencies when most effects are large is seen in Figure 3. To understand the slow dynamics at long times when the effect size can be small or large, here we study (6) for the allele frequency dynamics in the full model approximating the mean deviation by its stationary state value  $\Delta c_1^*$ . We thus have

$$\dot{p}_i \approx -s\gamma_i p_i q_i \Delta c_1^* - \frac{s\gamma_i^2}{2} p_i q_i (q_i - p_i) + \mu(q_i - p_i). \quad (\text{E1})$$

An exact solution of the above equation requires solving a cubic equation which is possible to do but not particularly enlightening.

To find the behavior of the allele frequency close to the stationary state, we write  $p_i(t) = p_i^* + y_i^{-1}(t)$ ,  $y_i \gg 1$  in the above equation and obtain

$$-\dot{y}_i = -s\gamma_i^2 y_i^{-1}(t) + D_1 - D_2 y_i(t), \quad (\text{E2})$$

where  $p_i^*$  is the exact solution of (7) with nonzero  $\Delta c_1^*$ , and the constants  $D_1$  and  $D_2$  are functions of  $\gamma_i, s, \mu, \Delta c_1^*$ . As we are interested in large times where the allele frequency is close to the stationary state, the first term on the RHS of the above equation can be neglected and we arrive at a first-order differential equation for  $y_i$  which can be easily solved. For large times, we then obtain  $p_i(t) - p_i^* \sim e^{-D_2 t}$ , where

$$D_2 = s\gamma_i^2 \left[ \frac{1}{2} + \left( \frac{\hat{\gamma}}{2\gamma_i} \right)^2 - 3p_i^* (1 - p_i^*) + \frac{(1 - 2p_i^*)\Delta c_1^*}{\gamma_i} \right]. \quad (\text{E3})$$

Thus after the directional selection phase, the allele frequency at the locus with effect  $\gamma_i$  reaches a stationary state over a timescale  $\sim D_2^{-1}$ . For small mean deviation from the optimum, which results in negligible  $\Delta c_1^*$ , we can use (8) to obtain

$$D_2 \propto \begin{cases} s\hat{\gamma}^2, & \gamma_i \ll \hat{\gamma} \\ s\gamma_i^2, & \gamma_i \gg \hat{\gamma}. \end{cases} \quad (\text{E4})$$

$$(\text{E5})$$

The above equation shows that when most effects are small, the dynamics at long times are driven by mutations (recall that  $\hat{\gamma}^2 = 8\mu/s$ ) while in the opposite case, the effect size continues to play a role even at long times.

### Appendix F: Cumulant Dynamics When Most Effects Are Large

Consider first the integral  $I_i^+$  defined by (26) that includes the contribution from loci with initial frequency close to one and given by

$$I_l^+ = \int_{2/\alpha}^{\infty} dx \frac{xe^{-x}}{1 + (\alpha x)^2 e^{\beta x}} \quad (\text{F1})$$

$$\approx \frac{1}{\alpha^2} \int_{2/\alpha}^{\infty} dx \frac{e^{-(1+\beta)x}}{x}, \quad (\text{F2})$$

where  $\alpha = 2\bar{\gamma}/\hat{\gamma}$  is large. The last integral can be written as an exponential integral  $E_1(z) = \int_1^{\infty} dt t^{-1}e^{-zt}$  and we have

$$I_l^+ = \frac{1}{\alpha^2} E_1 \left[ \frac{2}{\alpha} (1 + \beta) \right] \quad (\text{F3})$$

$$\approx -\frac{1}{\alpha^2} \ln \left[ \frac{2}{\alpha} (1 + \beta) \right], \quad (\text{F4})$$

on using that the argument of  $E_1(z)$  is small (Abramowitz and Stegun 1964). We next consider the integral  $I_l^-$  defined by (28) that contains the contribution from loci with small initial frequencies,

$$I_l^- = \int_{2/\alpha}^{\infty} dx \frac{xe^{-x}}{1 + (\alpha x)^{-2} e^{\beta x}} \quad (\text{F5})$$

$$= \int_{2/\alpha}^{\infty} dx xe^{-x} + \int_0^{2/\alpha} dx \frac{xe^{-x}}{1 + (\alpha x)^2 e^{-\beta x}} - J_l \quad (\text{F6})$$

$$\approx 1 - \left( \frac{2}{\alpha} \right)^2 - J_l, \quad (\text{F7})$$

where we have used that  $\alpha \gg 1$  to obtain the last expression and defined

$$J_l = \int_0^{\infty} dx \frac{xe^{-x}}{1 + (\alpha x)^2 e^{-\beta x}}. \quad (\text{F8})$$

As shown in Table S1 in File S1, the integral  $J_l$  decreases as  $1/\alpha$  or slower. Thus, for large  $\alpha$ , Equation (24) for the mean gives

$$c_1(t) = \ell \bar{\gamma} (1 - I_l^+ - I_l^-) \approx \ell \bar{\gamma} J_l(\beta). \quad (\text{F9})$$

In the above equation, the integral  $I_l^+$  does not contribute to the mean since, as discussed after (32), the allele frequency at loci with high initial frequencies does not change significantly while the mean is evolving.

The time dependence of  $\beta$  is obtained from Equation 12,

$$\frac{d\beta}{d\tau} \approx \rho - J_l(\beta). \quad (\text{F10})$$

The function  $\beta$  increases with time and saturates to its steady state value  $\bar{\beta}$  where  $\dot{\beta} = 0$ . Correspondingly, as Table S1 in File S1 indicates, the integral  $J_l(\beta)$  is close to zero when  $\beta \ll 1$  and approaches  $\rho$  as  $\beta \rightarrow \bar{\beta}$ . Thus at short times, we can neglect the term  $J_l$  on the RHS of (F10), and at large times, we can expand  $J_l$  about  $\rho$  to write (F10) as  $\dot{\beta} \approx -s\ell\bar{\gamma}^2 J_l'(\bar{\beta})(\beta - \bar{\beta})$ . Denoting the time below and above which these two solutions are valid by  $t_{\times}$  and setting  $\beta(t_{\times}) = 1$ , we obtain

$$\beta \approx \begin{cases} s\ell\bar{\gamma}^2\rho t, & t \ll t_\times \quad (\text{F11}) \\ \tilde{\beta} + (1 - \tilde{\beta})e^{-\omega(t-t_\times)}, & t \gg t_\times \quad (\text{F12}) \end{cases}$$

where  $t_\times = (s\ell\bar{\gamma}^2\rho)^{-1}$ ,  $J_l(\tilde{\beta}) = \rho$ , and  $\omega = s\ell\bar{\gamma}^2 J_l'(\tilde{\beta})$ . Using this result in (F9), we find the mean deviation to be

$$-\Delta c_1(t) \approx \begin{cases} z_f, & t \ll t_\times \quad (\text{F13}) \\ \ell\bar{\gamma} J_l'(\tilde{\beta}) (\tilde{\beta} - 1) e^{-s\ell\bar{\gamma}^2 J_l'(\tilde{\beta})(t-t_\times)}, & t \gg t_\times. \quad (\text{F14}) \end{cases}$$

Using the above expressions in (13), we find that the variance is given by

$$c_2(t) \approx \begin{cases} 0, & t \ll t_\times \quad (\text{F15}) \\ \ell\bar{\gamma}^2 \left[ J_l'(\tilde{\beta}) - (\tilde{\beta} - 1) J_l''(\tilde{\beta}) e^{-\omega(t-t_\times)} \right], & t \gg t_\times. \quad (\text{F16}) \end{cases}$$

To express the mean and the variance in terms of the model parameters, we first need to determine  $\tilde{\beta}$  which is a solution of the equation  $J_l(\tilde{\beta}) = \rho$ . The integral  $J_l(\beta)$  is not exactly solvable but we can estimate it by a saddle-point method [see (F17) below] (Arfken 1985). Denoting the integrand of  $J_l$  in (F8) by  $j(x)$  [i.e.,  $J_l = \int_0^\infty j(x) dx$ ] and the maximum of  $j(x)$  by  $x_0$ , we obtain

$$J_l \approx j(x_0) \left[ \frac{-d^2 \ln j(x)}{dx^2} \Big|_{x_0} \right]^{-1}, \quad (\text{F17})$$

where  $x_0$  is a solution of the equation  $dj/dx|_{x_0} = 0$ ,

$$e^{\tilde{\beta}x_0} (\tilde{\beta}x_0 - \tilde{\beta}) + \frac{\alpha^2}{\tilde{\beta}} (\tilde{\beta}x_0)^2 (1 + x_0 - \tilde{\beta}x_0) = 0. \quad (\text{F18})$$

As our numerics indicate that  $\tilde{\beta}x_0$  is large and constant for a given  $\alpha$ , using the above equation, we obtain  $\tilde{\beta}x_0 \approx 2\ln\alpha$  for large  $\alpha$ . Using this solution in (F17), after some simplifications, we find that

$$J_l \propto x_0 e^{-x_0}. \quad (\text{F19})$$

Setting  $J_l(\tilde{\beta}) = \rho$  in the above expression yields

$$\tilde{\beta} \approx \frac{\ln \alpha}{\ln \rho^{-1}}. \quad (\text{F20})$$

Taking the derivative with respect to  $\beta$  in (F19), we arrive at

$$J_l'(\tilde{\beta}) = \frac{\rho(x_0 - 1)}{\tilde{\beta}} \sim \frac{\rho(\ln \rho)^2}{\ln \alpha}. \quad (\text{F21})$$

Our numerics are consistent with the above functional form and a numerical fit shows  $J_l'(\tilde{\beta}) = (0.69/\ln\alpha)\rho [\ln(1.7/\rho)]^2$ .