

# Selective Strolls: Fixation and Extinction in Diploids Are Slower for Weakly Selected Mutations Than for Neutral Ones

Fabrizio Mafessoni\*<sup>1</sup> and Michael Lachmann<sup>†</sup>

\*Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany 04103, and

<sup>†</sup>Santa Fe Institute, Santa Fe, New Mexico 87501

ORCID ID: 0000-0003-4319-2076 (F.M.)

**ABSTRACT** In finite populations, an allele disappears or reaches fixation due to two main forces, selection and drift. Selection is generally thought to accelerate the process: a selected mutation will reach fixation faster than a neutral one, and a disadvantageous one will quickly disappear from the population. We show that even in simple diploid populations, this is often not true. Dominance and recessivity unexpectedly slow down the evolutionary process for weakly selected alleles. In particular, slightly advantageous dominant and mildly deleterious recessive mutations reach fixation slightly more slowly than neutral ones (at most 5%). This phenomenon determines genetic signatures opposite to those expected under strong selection, such as increased instead of decreased genetic diversity around the selected site. Furthermore, we characterize a new phenomenon: mildly deleterious recessive alleles, thought to represent a wide fraction of newly arising mutations, on average survive in a population slightly longer than neutral ones, before getting lost. Consequently, these mutations are on average slightly older than neutral ones, in contrast with previous expectations. Furthermore, they slightly increase the amount of weakly deleterious polymorphisms, as a consequence of the longer unconditional sojourn times compared to neutral mutations.

**KEYWORDS** weak selection; diffusion approximation; dominance; recessive mutations

A new allele emerging in a finite population usually has two possible fates—extinction or fixation. Selection affects the probability with which these occur and how long it will take. Thus, an advantageous allele has an increased chance to fix, due to positive selection, while a deleterious mutation has an increased chance of extinction. In both cases, the time until fixation and extinction is commonly thought to decrease with the strength of selection (Kimura and Ohta 1969b; van Herwaarden and Van Der Wal 2002). Kimura and Ohta (1969a,b) and Ewens (2004) applied diffusion theory to model finite populations, obtaining keystone approximations for the neutral case, in the absence of selection, or when selection is fairly strong. Recently it has been pointed out

that in haploid models, in the presence of frequency-dependent fitness, the time to fixation of a positively selected allele can increase with the strength of selection (Altrock *et al.* 2010, 2012). In diploids, a newly arising mutation is usually expected to have a time to fixation longer than that of a neutral one in the case of overdominance, when the heterozygote has a higher fitness than the two homozygotes—a case usually referred to as heterozygote advantage or more generally balancing selection (Key *et al.* 2014). In the case of overdominance, both alleles would be maintained in an infinite population. In this article we unify these results, showing that in diploids, certain classes of mutations behave as *slow sweeps/selective strolls*: the time to fixation of a positively selected allele *A* can be slightly longer than in the neutral case, provided that selection is weak and requiring only incomplete dominance (*i.e.*, the heterozygote *AB* has higher fitness than the disfavored homozygote *BB* but lower or equal to the favored homozygote *AA*).

Maruyama (1974) has already shown that the time to fixation for deleterious alleles decreases as selection becomes more negative. For a simple haploid model with a selected

Copyright © 2015 by the Genetics Society of America

doi: 10.1534/genetics.115.178160

Manuscript received May 14, 2015; accepted for publication October 14, 2015; published Early Online October 22, 2015.

Supporting information is available online at [www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178160/-/DC1](http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178160/-/DC1).

<sup>1</sup>Corresponding author: Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig 04103, Germany. E-mail: [fabrizio.mafessoni@gmail.com](mailto:fabrizio.mafessoni@gmail.com)

allele of selection strength  $1 + s$  vs. the wild type with fitness 1, the time to fixation is equal for  $s = +|x|$  and  $s = -|x|$ . For diploids the same phenomenon occurs, provided that the dominance relationship between alleles is reversed (Ewens 2004). Hence, our findings extend to the case of mildly deleterious and recessive mutations.

We also investigate the trajectories conditional to extinction of newly arising deleterious mutations. Deleterious mutations are thought to be rapidly purged from a population. Combining our results, we demonstrate that mutations with these features, likely constituting a large fraction of newly arising ones, have not only slightly longer fixation times, but also longer extinction times. Hence in a population these mutations, despite being under negative selection, survive as polymorphisms slightly longer than neutral ones and are on average older.

When a selected mutation spreads in a population until fixation, surrounding sites are carried along, a phenomenon called genetic hitchhiking. Since a selected mutation usually spreads rapidly and less time is available for recombination, the genetic diversity around a positively selected site is commonly expected to decrease, a phenomenon called a *selective sweep* (Smith and Haigh 1974; Nielsen 2005). Most statistics used to detect genetic signatures of positive selection rely on this assumption (Tajima 1989; Fu and Akey 2013; Hider *et al.* 2013; Racimo *et al.* 2014). We show that these signatures are reversed in slow sweeps/selective strolls, with increased genetic diversity compared to that of a neutrally fixing allele. We then explore the question of how often such a slowdown would occur in a population, by using different assumptions about the distribution of selection and dominance effects in nature.

## Materials and Methods

### Diffusion approximation

We assume a Wright–Fisher population model in the limit of large population size and use diffusion theory as described in Ewens (2004). The density function  $\psi(t, p)$  of the time  $t$  until absorption, starting from frequency  $p$ , satisfies the backward Kolmogorov equation that can be used to determine the unconditional sojourn time between  $a$  and  $b$ , expressed as  $\int_a^b t(x, p) dx$ . When  $a$  and  $b$  correspond to the absorbing states 0 and 1, we obtain the average unconditional time until absorption

$$\bar{t}(p) = \int_0^1 t(x, p) dx, \quad (1)$$

where

$$t(x, p) = \frac{2P_0(p)}{x(1-x)\psi(x)} \int_0^x \psi(y) dy, \quad 0 \leq x \leq p, \quad (2)$$

$$t(x, p) = \frac{2P_1(p)}{x(1-x)\psi(x)} \int_x^1 \psi(y) dy, \quad p \leq x \leq 1, \quad (3)$$

$$\psi(y) = \exp\{2N_e s(2h - 1)y^2 - 2hy\}, \quad (4)$$

and the probability of fixation  $P_1(x)$  can be expressed as

$$P_1(x) = 1 - P_0(x) = \frac{\int_0^p \psi(y) dy}{\int_0^1 \psi(y) dy}. \quad (5)$$

The mean conditional time until fixation or extinction can be directly obtained by conditioning sojourn times to fixation or extinction, respectively, and then integrating over all population frequencies:

$$\bar{t}^1(p) = \frac{\int_0^1 t(x, p) P_1(x) dx}{P_1(p)}, \quad (6)$$

$$\bar{t}^0(p) = \frac{\int_0^1 t(x, p) P_0(x) dx}{P_0(p)}. \quad (7)$$

Following Maruyama and Kimura, the age of a mutant allele can be calculated by integrating over the density of sojourn times at  $z$ , given that the current frequency is  $x$ :

$$a(x) = \int_0^1 \frac{t(x, z) t(z, p)}{t(x, p)} dz. \quad (8)$$

### Moran process

We model a birth–death Moran process as in Altrock *et al.* 2012, for which fixation and extinction times can be calculated analytically. We consider a biallelic haploid well-mixed population of size  $2N$ . We denote the number of  $A$  alleles as  $i$  and  $B$  alleles as  $2N - i$ . To mimic the fitness structure of a diploid population we consider the birth and death probability of birth of type  $A$  respectively equal to

$$p_i = \frac{i}{2N}, \quad q_i = 1 - p_i \quad (9)$$

$$b_i = \frac{p_i^2(1+s) + p_i q_i(1+hs)}{p_i^2(1+s) + 2p_i q_i(1+hs) + q_i^2}. \quad (10)$$

The transition probabilities to go from  $i$  to  $i + 1$ , to go from  $i$  to  $i - 1$ , or to stay at  $i$  are equal to

$$T_i^+ = b_i(1 - p_i), \quad T_i^- = (1 - b_i)p_i, \quad T_i^0 = 1 - T_i^+ - T_i^-. \quad (11)$$

We also consider the simplified jump process introduced in the main text, which ignores the chance to stay at  $i$ . In this case, unless the focal allele is in an absorbing state, its frequency is never the same in two consecutive time steps. Thus if at time  $t$  there are  $iA$  alleles, with  $i \neq 0$  and  $i \neq 2N$ , at time  $t + 1$  there can be only  $i - 1$  or  $i + 1$   $A$  alleles. Hence we normalize the new transition probabilities as  $\tilde{T}_i^+ = T_i^+ / (T_i^+ + T_i^-)$  and  $\tilde{T}_i^- = 1 - \tilde{T}_i^+$ .

Keeping  $T_i^-$  and  $T_i^+$  to describe in general unconditional transition probabilities, for both the general and the simplified model, the arrival probability at state  $j$  can be obtained by solving the recursion equation

$$\phi_i^j = (1 - T_i^+ - T_i^-) \phi_i^j + T_i^+ \phi_{i+1}^j + T_i^- \phi_{i-1}^j \quad (12)$$

(Ewens 2004); thus

$$\phi_i^j = \frac{\sum_{k=0}^{i-1} \gamma_{1,k}}{\sum_{k=0}^{j-1} \gamma_{1,k}}, \quad \text{if } i < j \quad (13)$$

$$\phi_i^j = \frac{\sum_{k=i}^{2N-1} \gamma_{j+1,k}}{\sum_{k=0}^{j-1} \gamma_{j+1,k}}, \quad \text{if } i > j, \quad (14)$$

where  $\gamma_{i,k} = \Pi_{i=m}^k (T_i^- / T_i^+)$ . The fixation and extinction probabilities are then denoted as  $\phi_i^{2N}$  and  $\phi_i^0 = 1 - \phi_i^{2N}$ .

The conditional fixation and extinction times,  $t_i^{2N}$  and  $t_i^0$  can be calculated similarly to the sojourn times, as in Altrock *et al.* (2012). The average sojourn time in  $j$ , starting in  $i$ , is

$$t_{ij} = \frac{\phi_i^j}{T_j^+ (1 - \phi_{j+1}^j) + T_j^- (1 - \phi_{j-1}^j)}. \quad (15)$$

The conditional fixation  $t_{ij}^{2N}$  and extinction  $t_{ij}^0$  sojourn times can be obtained by conditioning as in Ewens (2004):

$$t_{ij}^{2N} = \frac{\phi_i^{2N}}{\phi_j^{2N}} t_{ij}, \quad t_{ij}^0 = \frac{\phi_j^0}{\phi_i^0} t_{ij}. \quad (16)$$

The average times until absorption, fixation and extinction are simply the sum of the sojourn times, either unconditioned on the absorbing state, or conditioned on fixation or extinction, respectively:

$$t_i = \sum_{j=1}^{2N-1} t_{ij}, \quad t_i^{2N} = \sum_{j=1}^{2N-1} t_{ij}^{2N}, \quad t_i^0 = \sum_{j=1}^{2N-1} t_{ij}^0. \quad (17)$$

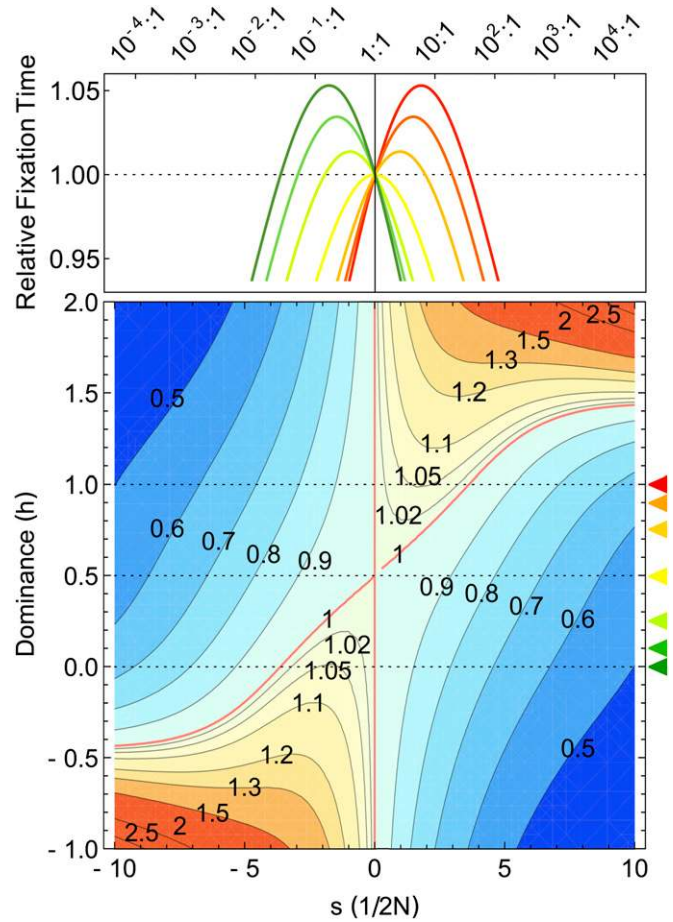
## Simulations

We used SLiM (Messer 2013) and custom code for unconditional simulations and MSMS (Ewing and Hermisson 2010) for simulations conditioned on fixation. Wright–Fisher simulations were performed for 10,000, 500, or 50 individuals. We simulated  $1.1 \times 10^7$  unconditional simulations (sojourn times, age of segregating alleles),  $10^7$  runs conditioned on extinction, and  $10^5$  runs conditioned on fixation, to verify the accuracy of diffusion approximations.

## Results

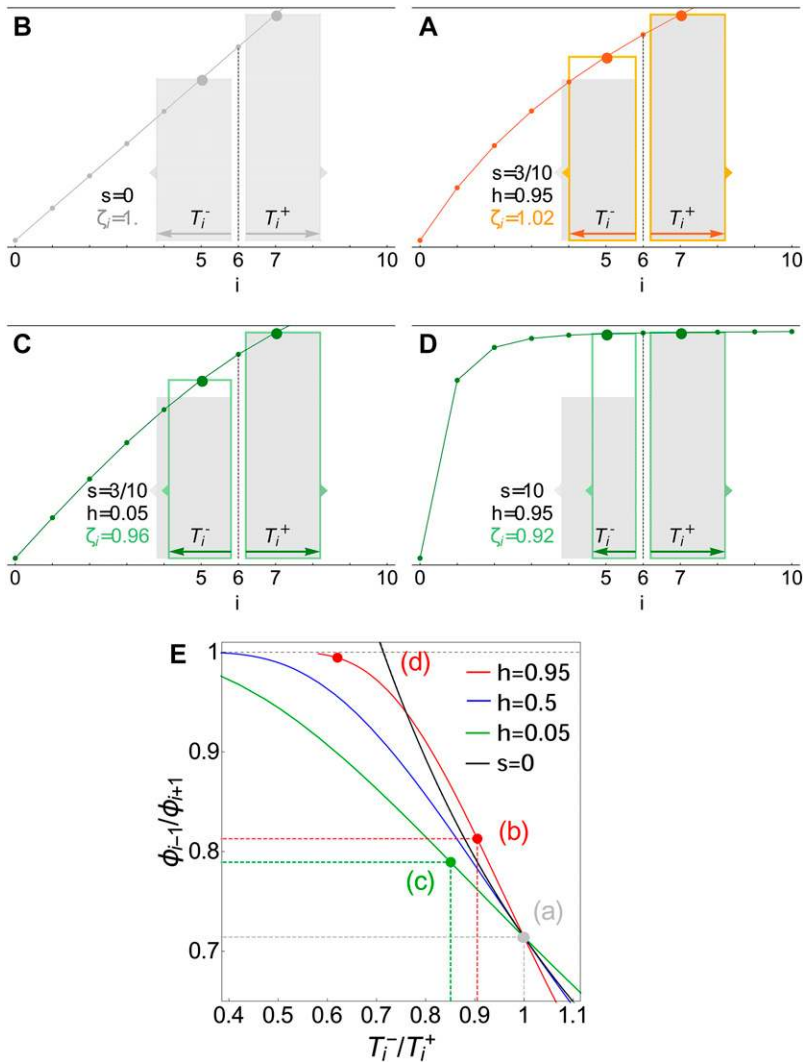
### Dominant weakly advantageous alleles reach fixation more slowly than neutral ones

We investigate a classic single-locus two-allele Wright–Fisher model. The wild-type homozygote with genotype  $aa$  has fitness 1, and we study the fate of a newly introduced allele  $A$ . The fitness of the homozygote  $AA$  is  $1 + s$ , where  $s$  denotes the selection coefficient, while the heterozygote  $Aa$  has fitness  $1 + h s$ , where  $h$  denotes dominance. In the absence of mutations, extinction and fixation, when the frequency of allele



**Figure 1** Time of fixation relative to neutrality for different selection (x-axis) and dominance (y-axis) coefficients, for a population of size  $N_e$ . (A) Different curves indicate different levels of dominance, from red to green. (B)  $h = 1, 0.9, 0.75, 0.5, 0.25, 0.1$ , and  $0$ , indicated by the colored triangles. In A, y-axis selection is also measured in terms of the odds for a site to be fixed in A rather than in B, at the stationary state (for  $h = 0.5$  and  $2N_e = 10^4$ ). (B) Time of fixation relative to neutrality for different values of positive selection (x-axis) and dominance (y-axis). A red contour (fixation times equal to neutrality) separates shorter (shades of blue) from longer (shades of orange) average fixation times than under neutrality. Dotted lines separate (from top to bottom) overdominance for positive selection, dominance, recessivity, and underdominance. Both graphs show numerical results of the diffusion approximation.

$A = 0$  or  $1$ , respectively, are two absorbing states. We explored the conditional expected time until either one or the other event occurs, with diffusion approximation and simulations, referring to them for brevity as extinction and fixation time, respectively. In Figure 1 we show diffusion approximations for the fixation time of weakly selected positive ( $s > 0$ ) and negative ( $s < 0$ ) mutations, relative to neutrality. For a completely dominant positive ( $h = 1$ ) allele  $A$  a peak in the time of fixation is observed around  $N_e s \simeq 1$ , where  $N_e$  is the variance effective population size. An analogous effect is observed for completely recessive ( $h = 0$ ) deleterious mutations. This result depends only on  $N_e s$  and  $h$ , and hence it holds also for small  $N_e$  (see Supporting Information, Figure S1 for simulations). Remarkably these effects occur even for



**Figure 2** Jump process with  $2N_e = 10$  for  $i = 6$  (black vertical line) advantageous  $A$  alleles. A–D compare the probabilities of decreasing in frequency ( $i \rightarrow i - 1$ , left rectangles) or increasing ( $i \rightarrow i + 1$ , right rectangles) given fixation, when  $A$  is neutral (A), weakly selected and dominant (B), or recessive (C) and strongly selected (D). For all plot, colored silhouettes refer to the different values of  $h$  and  $s$ , while full gray rectangles indicate neutrality. Horizontal arrows (width of rectangles) represent the chance of an increase/decrease in the number of  $A$  alleles ( $x$ -axis). The fixation probability at each frequency is shown on the  $y$ -axis. For all conditions, to facilitate comparisons, transition and fixation probabilities are normalized by the respective quantities for an increase in frequency (right rectangle): transition probabilities are rescaled by  $T_i^+$  and fixation probabilities by  $\phi_{i+1}$ , so that the right rectangle has area 1 while the left one has  $P_i^-/P_i^+$ . Hence, a slow-down occurs at  $i$  when the left colored silhouette is larger than the shaded gray rectangle in its background (*i.e.*,  $\zeta_i > 1$ ). Note that for weak selection  $T_i^-(s, h) \approx T_i^-(s = 0)$ , while dominance (B) exposes relatively more the effects of the advantageous mutation at lower frequencies, increasing the height of the left rectangle, which hence is larger than the gray one. (E) Why can the left rectangle be larger than for neutrality? Dominance exposes the effects of selection at lower frequencies; hence as  $s$  increases from 0 (gray circle) the relative probability of fixation at  $i - 1$  vs.  $i + 1$  ( $y$ -axis) increases more rapidly than the relative probability of a decrease in  $i$  ( $x$ -axis,  $T_i^-/T_i^+$ ). Along each of the colored curves selection varies (increasing from bottom to top) for different  $h$ :  $h = 0.95$  (red),  $h = 0.5$  (blue), and  $h = 0.05$  (green). Points (a–d) indicate the respective plots A–D above. The areas delimited by the dotted lines coincide with the left rectangles in A–D, *i.e.*,  $P_i^-/P_i^+$ . The black curve delimits fixation speed  $P_i^-/P_i^+$  equivalent to neutrality. Hence as  $s$  increases from zero, the red curve goes above the black curve (slowdown), the green one goes below it (speedup), and the blue curve is tangent.

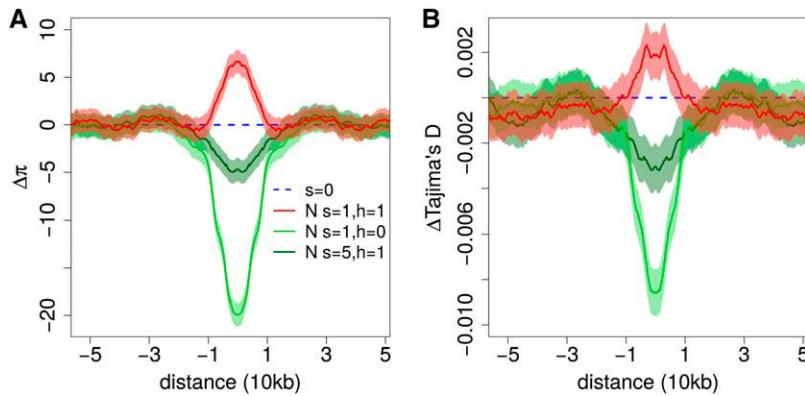
intermediate levels of dominance, although to a smaller degree as  $h$  approaches  $1/2$ . The symmetry of fixation times has been already noted by Maruyama (1974), who showed that inverting the coefficient and the sign of the selection coefficient keeps the fixation time unchanged. However, this symmetry is centered at  $s = 0$  only when  $h = 0.5$ . Hence a balancing selection-like pattern, with increased fixation times compared to neutrality, is observed in a nontrivial region in the  $s - h$  parameter space (Figure 1): increased fixation times occur in an hourglass-shaped parameter region, rather than in a simpler rectangular fashion (overdominance,  $h > 1, s > 0$ ), due to the stochastic slowdown shown here for weak selection and the decrease in fixation time for overdominant strong selection described by Robertson (1962).

### An intuitive explanation of the stochastic slowdown in diploids

Why does this counterintuitive phenomenon occur? Altrock *et al.* (2012) pointed out that in haploids the sojourn times, the amount of time spent at the different population frequencies,

increase at higher frequencies. Here we provide a simple explanation for the stochastic slowdown in diploids, by looking at the conditional trajectories. A conditional trajectory is slower whenever the frequency of the selected allele decreases, only to increase again later and eventually reach fixation. Therefore the fixation probability of a mutation going to lower frequencies has a key role in the fixation time: in the case of positive selection dominance increases the probability that once a drop in frequency occurs, the selected allele  $A$  will reach fixation anyway. This occurs since even if there is a higher chance for  $A$  to be paired with the less advantageous allele  $a$ , the fixation probability is still relatively high because the average fitness of the heterozygote is biased toward the more advantageous allele. This effect is apparent only for weak selection, since when  $s$  is strong, the probability to decrease in frequency, compared to the probability of increasing, is so low that this effect can be neglected. A similar explanation is applied to recessive deleterious mutations. In this case recessivity masks the deleteriousness of the selected allele  $A$ , so that when a drop in frequency occurs, its absolute fitness will increase since it will be biased toward the advantageous allele.





**Figure 3** Patterns of nucleotide diversity ( $\pi$ ) and Tajima's  $D$  around the site of a selected allele that just reached fixation. We considered a neutral (blue), a dominant (red), and a recessive (dark green) weakly selected allele and a dominant allele with stronger selection (light green) as indicated in the key. The population size is  $N_e = 10^4$ , and mutation rate is  $1.2 \times 10^{-8}$ . Shaded regions indicate the 95% confidence interval of the mean of  $5 \times 10^5$  simulations. On the y-axis, the statistics are reported symmetrically for each distance as differences compared to neutrality. Distribution of fixation times and raw measures are shown in Figure S8, Figure S9, and Figure S10.

Hence, its probability of fixation would be relatively high (Ohta 1973; Ohta and Kimura 1975).

To illustrate this intuitive explanation, we use a simplified Moran process, with fitness mirroring that of a diploid biallelic population with  $2N_e$  alleles (Hartl and Clark 2007). In this jump process the number of  $A$  alleles, indicated as  $i$ , can change only one allele at a time, so that the only nonzero transition probabilities are those of moving from  $i$  to  $i + 1$  ( $T_i^+$ ) or from  $i$  to  $i - 1$  ( $T_i^-$ ). We denote the fixation and extinction probabilities in  $i = 2N_e$  and  $i = 0$  as  $\phi_i$  and  $1 - \phi_i$ , respectively. In general, the conditional transition probabilities, given that fixation will eventually occur, are obtained by multiplying the unconditional transition probabilities by the probability of fixation once in the new state and by normalizing for their sum, equivalent to  $\phi_i$ . In this jump process, the probabilities of a decrease or of an increase in  $i$  conditional on fixation, i.e.,  $P_i^-$  and  $P_i^+$ , respectively, are simply equal to the normalized areas of the rectangles  $T_i^- \phi_{i-1}$  and  $T_i^+ \phi_{i+1}$ . These are shown in Figure 2, A–D: the larger the left rectangle is compared to the right one, the slower a trajectory is because of the more frequent decreases in allele frequency followed by fixation. When this area (left colored silhouettes) is larger than that under neutrality (Figure 2, A–D, gray lines under gray rectangles), a slowdown occurs at that frequency. This condition can be expressed as  $\zeta_i = P_i^-(s, h)/P_i^-(s = 0) > 1$ , where  $P_i^-(s, h)$  is calculated for selection and dominance coefficients  $s$  and  $h$ .

It is easy to observe how dominance buffers the decrease in frequency by decreasing the probability of extinction for lower  $A$  frequencies, thus increasing the height of the rectangle  $T_i^- \phi_{i-1}$  (Figure 2A) and the overall area above that under neutrality. When selection is stronger, the transition probability  $T_i^-$  is much smaller, and the area of the rectangle  $T_i^- \phi_{i-1}$  cannot be compensated enough by increasing its relative height (Figure 2D). Recessivity exerts the opposite effect, accelerating fixation (Figure 2C). Vice versa, recessivity determines slightly longer fixation times in the case of a deleterious  $A$  allele (Figure S4). Figure 2E summarizes the previous cases, showing how the conditional probability of a decrease in frequency relative to that of an increase (rectangles delimited by dotted lines) increases above that of neutrality (black curve) due to a steeper increase of the ratio

$\phi_{i-1}/\phi_{i+1}$  compared to  $T_i^-/T_i^+$ , in the presence of dominance and weak selection (red curve).

We note that the stochastic slowdown is stronger at higher frequencies (Figure S5 and Figure S6), as it can be also seen from the distribution of conditional sojourn times reported for the Wright–Fisher model (Figure S7). However, this effect is not an artifact of considering trajectories until complete fixation: this effect is present even at intermediate frequencies (for example, 60% as in Figure 2) or when a trajectory is conditioned only to reach partial fixation (Figure S2).

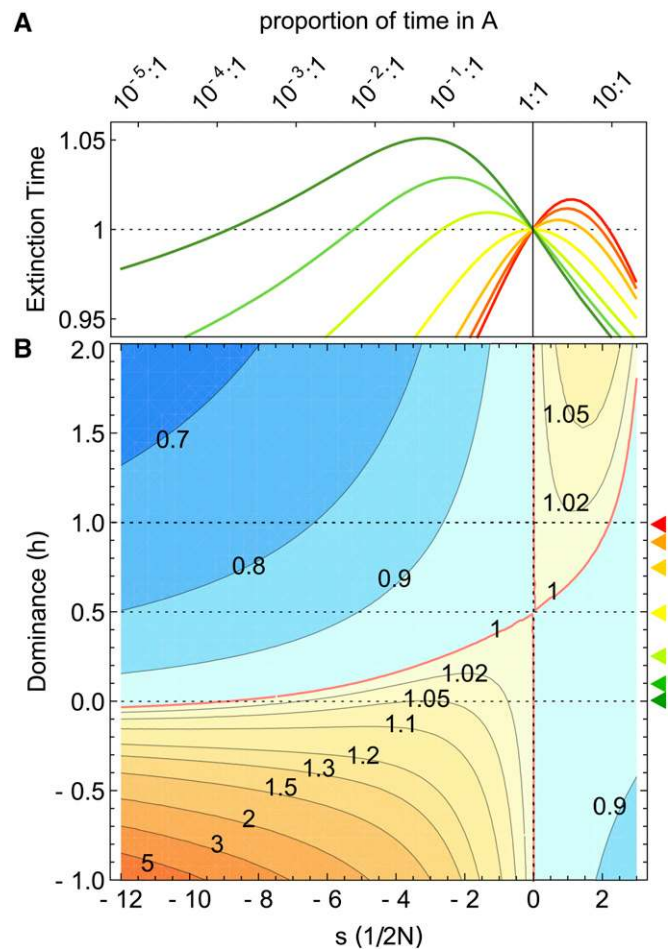
#### Slow sweeps/selective strolls signatures

Consistent with the slightly longer time until fixation, genetic diversity is slightly increased compared to neutrality around a selected site, in opposite fashion to classical selective sweeps (Figure 3). Hence, even classical statistics to detect positive selection such as Tajima's  $D$  show opposite patterns for slow and classical selective sweeps. We used a hypothetical distribution of dominance and selective effects to estimate the fraction of fixation events due to either slow or classical selective sweeps (Figure S11). We assumed the distribution of selective effects (DFE) estimate in Racimo and Schraiber (2014) and explored a truncated normal distribution for dominance effects, by excluding overdominance and underdominance cases. When the variance of the latter is high, this converges on a uniform distribution, and the fraction of fixation events due to slow sweeps/selective strolls is  $\sim 1/2$  (Figure S11E). This fraction decreases rapidly as dominant mutations are rarer. Since the DFE is estimated for mostly deleterious mutations, we also conservatively take into consideration the possibility that the distribution of selection coefficients is much more skewed toward stronger effects for positive selection. Hence we considered exponential DFE with mean 5 and 10 times higher than in Racimo *et al.* (2014). In these cases the fraction of fixation events due to slow sweeps/selective strolls is lower, when  $h$  is uniformly distributed about 1/5 and 1/10, respectively (Figure S11F).

#### Recessive weakly deleterious alleles persist in a population slightly longer than neutral ones

It is estimated that most newly arising mutations in functional regions are slightly deleterious and partially recessive,

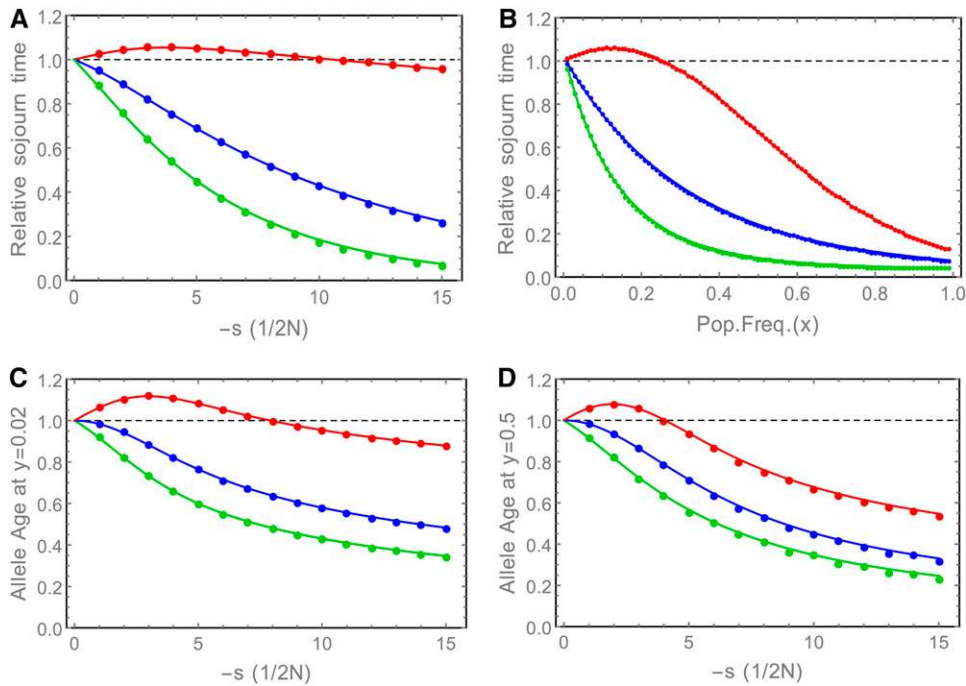
consistent with the nearly neutral theory of molecular evolution (Ohta 1973; Davies *et al.* 1999; Eyre-Walker and Keightley 2007; Agrawal and Whitlock 2011; Racimo and Schraiber 2014). A remarkable implication of the selective slowdown is that these mutations disappear more slowly than neutral ones from a population: the fixation of a slightly advantageous selected mutation is equivalent to the extinction of a slightly deleterious recessive one. Furthermore we have seen that the slowdown process is stronger for a high frequency of the advantageous mutation and hence for low frequencies of the deleterious one. Hence we can ask, What is the time until extinction of a newly appeared mutation? This is also determined by the arrival time, the time that this mutation takes to initially invade. We have already observed how this process is slightly slower, when considered until fixation. Therefore we hypothesize that the extinction time of a weakly deleterious recessive mutation is slightly longer than neutrality. We show this seemingly paradoxical result in Figure 4 for  $N_e = 10^4$ . The extinction times are longer than neutrality for recessive deleterious mutations. When recessivity is complete, even mutations with a selective coefficient almost as deleterious as  $-10/2N_e$  disappear more slowly than neutral ones. While conditional fixation times relative to neutrality are dependent only on  $h$  and  $N_e s$ , the stochastic slowdown for extinction is stronger for smaller absolute population sizes. For a population with  $2N_e = 100$  the extinction time can be even 10% longer than neutrality (Figure S3), while for populations with  $N_e > 10^4$  the effect of the stochastic slowdown is at most 5% (Figure 4). Since most deleterious mutations reach extinction, recessive deleterious mutations have also slightly longer unconditional sojourn times than neutral ones at low frequencies (Figure 5, A and B). The amount of polymorphisms at a given frequency is directly proportional to the unconditional sojourn times, *i.e.*, sojourn times independent of whether fixation or extinction is reached. Hence, this phenomenon determine a slight excess of weakly deleterious polymorphisms compared to neutral ones (Figure 5, A and B). A further consequence is that, contrary to what is generally thought, these alleles are on average older than neutral ones (Figure 5, C and D). This phenomenon is apparent at low frequencies, where the range of selection coefficients for which deleterious alleles are older expands (Figure 5C). The slightly larger amount of weakly deleterious recessive polymorphisms (Figure S12) has a direct, although small, impact on methods to detect selection that rely on calculations based on polymorphisms. An example is the McDonald–Kreitman test that is predicted to be biased by the presence of slightly deleterious polymorphisms (Charlesworth and Eyre-Walker 2008; Messer and Petrov 2013). Here we estimate the effect of recessivity and the contribution of selective strolls in underestimating the fraction of adaptive events,  $\alpha$  (Figure S13). The contribution to this bias is small and apparent only for strongly recessive alleles. Since the sojourn times are increased mostly at low frequencies, this bias is well accounted for by removing rare polymorphisms from the McDonald–Kreitman analyses (Charlesworth and Eyre-Walker 2008).



**Figure 4** (A) Time of extinction relative to neutrality for different selection coefficients (x-axis) and dominance for a population of size  $10^4$ . Different curves indicate different levels of dominance, from red to green. (B)  $h = 1, 0.9, 0.75, 0.5, 0.25, 0.1$ , and  $0$ , indicated by the colored triangles. In A, y-axis selection is also measured in terms of the odds for a site to be fixed in A rather than in B, at the stationary state (for  $h = 0.5$  and  $2N_e = 10^4$ ). (B) Time of fixation relative to neutrality for different values of positive selection (x-axis) and dominance (y-axis). A red contour (fixation times equal to neutrality) separates shorter (shades of blue) from longer (shades of orange) fixation times than under neutrality. Dotted lines separate (from top to bottom) overdominance, dominance, recessivity, and underdominance.

## Discussion

Our results show that the mild frequency dependence of the fitness of a selected allele, even when due only to small levels of recessivity and dominance, is enough to alter qualitatively the expected time until fixation and extinction of nearly neutral mutations. The effects of dominance on fixation (and extinction) times have been already investigated by previous studies that showed how the fixation of recessive and codominant mutations is faster than that of dominant ones (Teshima and Przeworski 2006; Ewing *et al.* 2010). These studies are consistent with the general assumption in population genetics that positive selection leads to a shortening of fixation times (Zhao *et al.* 2013). Most methods aimed at detecting molecular signatures of positive selection rely on



**Figure 5** (A–D) Sojourn times (A and B) and ages (C and D) of a mutated allele, relative to neutrality, for completely recessive (red), codominant (blue), and dominant (green) deleterious mutation. Circles indicate the average of  $10^6$  simulations, while lines show diffusion approximations. In A, C, and D different selective coefficients are shown on the x-axis, while for B the relative sojourn times are shown for different population frequencies, with  $s = -2/N$ . The population frequency of the selected allele is 0.1 (A), 0.02 (C), and 0.5 (D), respectively.

this assumption (Nielsen 2005; Vitti *et al.* 2013). Here we show that weakly selected dominant alleles, with  $N_e s \approx 1$ , violate the assumption of a rapid spread in the population: these alleles behave as slow sweeps, on average reaching fixation slightly more slowly than neutral ones. Hence, the expected signatures of these slow selective events are not only weak, but also even in the opposite direction of what is expected: the diversity around a fixed selected allele is higher than around a fixed neutral one.

The magnitude of this phenomenon is small ( $\sim 5\%$  at its peak), possibly explaining why it went unnoticed despite the large number of related theoretical studies. Its limited effects also suggest that identifying fixation events subjected to a stochastic slowdown is extremely unlikely in real data, as only strongly selected sweeps can be detected unambiguously. Rather, our results show that certain adaptive events are virtually indistinguishable from neutral ones, not only because of the lack of resolution or power, but in addition because of their inherently opposite genetic signatures. Previous studies documented difficulties in detecting hard sweeps, usually explained in terms of the prevalence of polygenic adaptation (Pritchard *et al.* 2010) or soft sweeps (Hermisson and Pennings 2005; Fu and Akey 2013) or possibly because many sweeps have a selection pressure too weak to detect. These mechanisms alone probably represent the most important reasons behind the observed scarcity in hard sweeps and difficulties in detecting adaptive events. Our results complement these findings, showing that besides these mechanisms, even in a simple biallelic system, certain classes of adaptive events are expected to be undetectable. Furthermore, not only methods apt at detecting positive selection based on local genetic diversity but also those relying on normalization over polymorphic variants (*e.g.*, McDonald–Kreitman)

are predicted to slightly underestimate the fraction of adaptive events, because of the longer survival of the larger amount of deleterious polymorphisms. This effect is present only for very low recessivity coefficients and can be almost entirely accounted for by removing low-frequency polymorphisms from the analyses.

In which cases are slow sweeps/selective strolls more relevant? Since this phenomenon appears only under weak selection ( $N_e s \approx 1$ ), it is unlikely to play an important role for large population sizes. In these cases only mutations with very small fitness effects would be affected. Furthermore in these cases fixation occurs on very long timescales; hence the slowdown effect is by then likely overwhelmed by the effects of other sites and environmental changes. For small population sizes we are more likely to observe selective strolls: even though selection is weak, the slowdown occurs for sites at which we are 10 times more likely to observe the selected allele than the deleterious one (see Figure 1). As current measurements observe that the majority of mutations have weak effects (Eyre-Walker and Keightley 2007; Racimo and Schraiber 2014), selective strolls could take up a large fraction of selected sites in genomes of species with a small  $N_e$ . An exact estimate of the fraction of selective events due to slow sweeps/selective strolls is limited by the lack of a precise knowledge of the distribution of dominance effects for newly arising mutations. However, exploring hypothetical distributions, we estimate that the fraction of slow sweeps/selective strolls is 10–20%, for conservative assumptions.

We have also shown that weakly deleterious recessive and partially recessive alleles have slightly longer average extinction times than neutral ones, both in the case of present deleterious variants and in the case newly arising mutations. Previous studies documented a negative dependence between

dominance and fitness effects, with a larger fraction of weakly selected mutations having additive or even dominant effects (Phadnis and Fry 2005; Agrawal and Whitlock 2011). Nevertheless, most deleterious and weakly deleterious ones are thought to be at least partially recessive (Agrawal and Whitlock 2011). For this reason, an increase in the length of extinction times of such mutated alleles is extremely relevant, increasing rather than decreasing the amount of certain classes of deleterious polymorphisms at certain frequencies. In particular, the sojourn times of these mutations at low frequencies are longer compared to neutrality, implying an accumulation of weakly deleterious variants at the population level (Fu *et al.* 2014). Note that we are referring here to the unconditional sojourn times, directly proportional to the site frequency spectrum.

So far, empirical studies failed to detect our theoretical predictions of slightly older weakly recessive deleterious mutations compared to neutral ones (Fu *et al.* 2013; Kiezun *et al.* 2013; Mathieson and McVean 2014; Rasmussen *et al.* 2014). The detection of this phenomenon would require high resolution in terms of estimates of allele ages; their fitness effects; and whether they are recessive, codominant, or dominant. It is currently difficult to estimate the degree of dominance of mutations with weak effects. Furthermore for extinction times, recent population sizes become more relevant, and most studies have been carried out in humans that underwent a strong population expansion. Further complications arise from the flatness of the distribution of allele ages (Qiu and Fedorov 2015). Future studies, and better estimates of the distribution of selective and dominance effects, might be able to confirm our theoretical predictions.

Hence our study describes a mechanism challenging some basic predictions about the speed of the stochastic dynamics of selected alleles, showing how dominance and recessivity can lead to genetic signatures opposite to the ones occurring in their absence. These results suggest new testable predictions about the permanence of deleterious mutations in current populations and suggest inherent limitations in our ability to detect adaptive events even in simple biallelic systems, particularly for small population sizes.

## Acknowledgments

We thank Chaitanya Gokhale, Arne Traulsen, two anonymous reviewers, and the editors for helpful comments. We also thank Kay Prüfer for support.

## Literature Cited

- Agrawal, A. F., and M. C. Whitlock, 2011 Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* 187: 553–566.
- Altrock, P. M., C. S. Gokhale, and A. Traulsen, 2010 Stochastic slowdown in evolutionary processes. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 82: 011925.
- Altrock, P. M., A. Traulsen, and T. Galla, 2012 The mechanics of stochastic slowdown in evolutionary games. *J. Theor. Biol.* 311: 94–106.
- Charlesworth, J., and A. Eyre-Walker, 2008 The McDonald–Kreitman test and slightly deleterious mutations. *Mol. Biol. Evol.* 25: 1007–1015.
- Davies, E. K., A. D. Peters, and P. D. Keightley, 1999 High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science* 285: 1748–1751.
- Ewens, W. J., 2004 *Mathematical Population Genetics. Interdisciplinary Applied Mathematics*, Ed. 2. Springer-Verlag, Berlin/Heidelberg, Germany/New York.
- Ewing, G., and J. Hermisson, 2010 MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* 26: 2064–2065.
- Ewing, G., J. Hermisson, P. Pfaffelhuber, and J. Rudolf, 2010 Selective sweeps for recessive alleles and for other modes of dominance. *J. Math. Biol.* 63: 399–431.
- Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8: 610–618.
- Fu, W., and J. M. Akey, 2013 Selection and adaptation in the human genome. *Annu. Rev. Genomics Hum. Genet.* 14: 467–489.
- Fu, W., T. D. O’Connor, G. Jun, H. M. Kang, G. Abecasis *et al.*, 2013 Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493: 216–220.
- Fu, W., R. M. Gitterman, M. J. Bamshad, and J. M. Akey, 2014 Characteristics of neutral and deleterious protein-coding variation among individuals and populations. *Am. J. Hum. Genet.* 95: 421–436.
- Hartl, D. L., and A. G. Clark, 2007 *Principles of Population Genetics*. Sinauer Associates, Sunderland, MA.
- Hermisson, J., and P. S. Pennings, 2005 Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169: 2335–2352.
- Hider, J. L., R. M. Gitterman, T. Shah, M. Edwards, A. Rosenbloom *et al.*, 2013 Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. *BMC Evol. Biol.* 13: 150.
- Key, F. M., J. C. Teixeira, C. de Filippo, and A. M. Andrés, 2014 Advantageous diversity maintained by balancing selection in humans. *Curr. Opin. Genet. Dev.* 29: 45–51.
- Kiezun, A., S. L. Pulit, L. C. Francioli, F. van Dijk, M. Swertz *et al.*, 2013 Deleterious alleles in the human genome are on average younger than neutral alleles of the same frequency. *PLoS Genet.* 9: e1003301.
- Kimura, M., and T. Ohta, 1969a The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* 63: 701–709.
- Kimura, M., and T. Ohta, 1969b The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61: 763–771.
- Maruyama, T., 1974 The age of a rare mutant gene in a large population. *Am. J. Hum. Genet.* 26: 669–673.
- Mathieson, I., and G. McVean, 2014 Demography and the age of rare variants. *PLoS Genet.* 10: e1004528.
- Messer, P. W., 2013 SLiM: simulating evolution with selection and linkage. *Genetics* 194: 1037–1039.
- Messer, P. W., and D. A. Petrov, 2013 Frequent adaptation and the McDonald–Kreitman test. *Proc. Natl. Acad. Sci. USA* 110: 8615–8620.
- Nielsen, R., 2005 Molecular signatures of natural selection. *Annu. Rev. Genet.* 39: 197–218.
- Ohta, T., 1973 Slightly deleterious mutant substitutions in evolution. *Nature* 246: 96–98.
- Ohta, T., and M. Kimura, 1975 Theoretical analysis of electrophoretically detectable polymorphisms: models of very slightly deleterious mutations. *Am. Nat.* 109: 137–145.
- Phadnis, N., and J. D. Fry, 2005 Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. *Genetics* 171: 385–392.



- Pritchard, J. K., J. K. Pickrell, and G. Coop, 2010 The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* 20: R208–R215.
- Qiu, S., and A. Fedorov, 2015 Maruyama's allelic age revised by whole-genome GEMMA simulations. *Genomics* 105: 282–287.
- Racimo, F., and J. G. Schraiber, 2014 Approximation to the distribution of fitness effects across functional categories in human segregating polymorphisms. *PLoS Genet.* 10: e1004697.
- Racimo, F., M. Kuhlwilm, and M. Slatkin, 2014 A test for ancient selective sweeps and an application to candidate sites in modern humans. *Mol. Biol. Evol.* 31: 3344–3358.
- Rasmussen, M. D., M. J. Hubisz, I. Gronau, and A. Siepel, 2014 Genome-wide inference of ancestral recombination graphs. *PLoS Genet.* 10: e1004342.
- Robertson, A., 1962 Selection for heterozygotes in small populations. *Genetics* 47: 1291–1300.
- Smith, J. M., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. *Genet. Res.* 23: 23–35.
- Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Teshima, K. M., and M. Przeworski, 2006 Directional positive selection on an allele of arbitrary dominance. *Genetics* 172: 713–718.
- van Herwaarden, O. A., and N. J. van der Wal, 2002 Extinction time and age of an allele in a large finite population. *Theor. Popul. Biol.* 61: 311–318.
- Vitti, J. J., S. R. Grossman, and P. C. Sabeti, 2013 Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47: 97–120.
- Zhao, L., M. Lascoux, A. D. J. Overall, and D. Waxman, 2013 The characteristic trajectory of a fixing allele: a consequence of fictitious selection that arises from conditioning. *Genetics* 195: 993–1006.

*Communicating editor: J. Hermisson*

# GENETICS

Supporting Information

[www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178160/-/DC1](http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178160/-/DC1)

## **Selective Strolls: Fixation and Extinction in Diploids Are Slower for Weakly Selected Mutations Than for Neutral Ones**

Fabrizio Mafessoni and Michael Lachmann

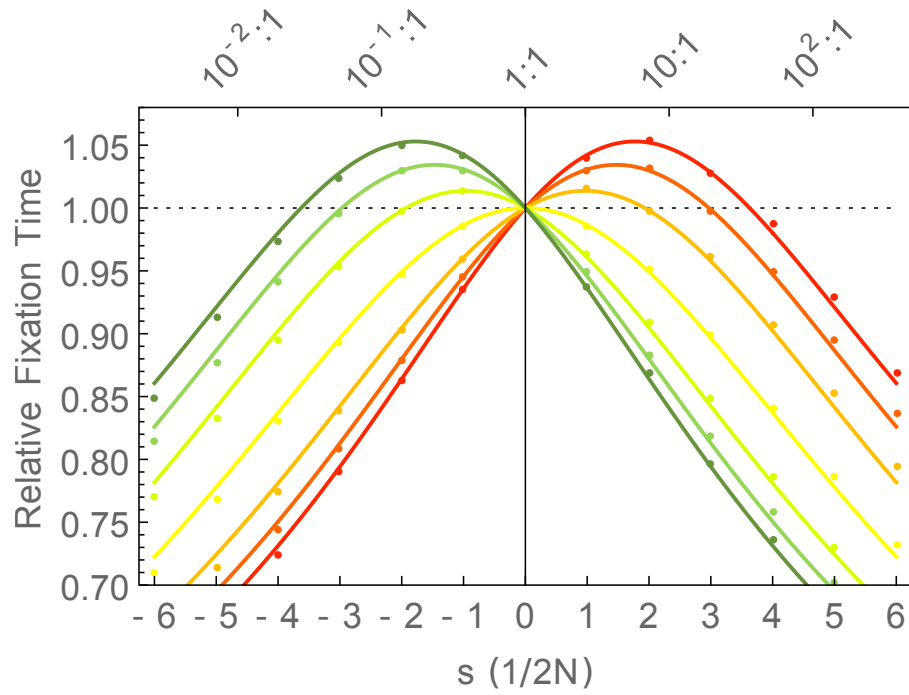


Figure S1: Diffusion approximation (continuous lines) and the average of  $10^5$  simulations (dots) for the fixation times relative to neutrality for a single initial allele in a population with  $N_e = 50$ . Error bars indicate the 95% confidence interval. Colors indicate values of  $h$  corresponding to figure 1, from red to the green equal to 1, 0.9, 0.75, 0.5, 0.25, 0.1 and 0.

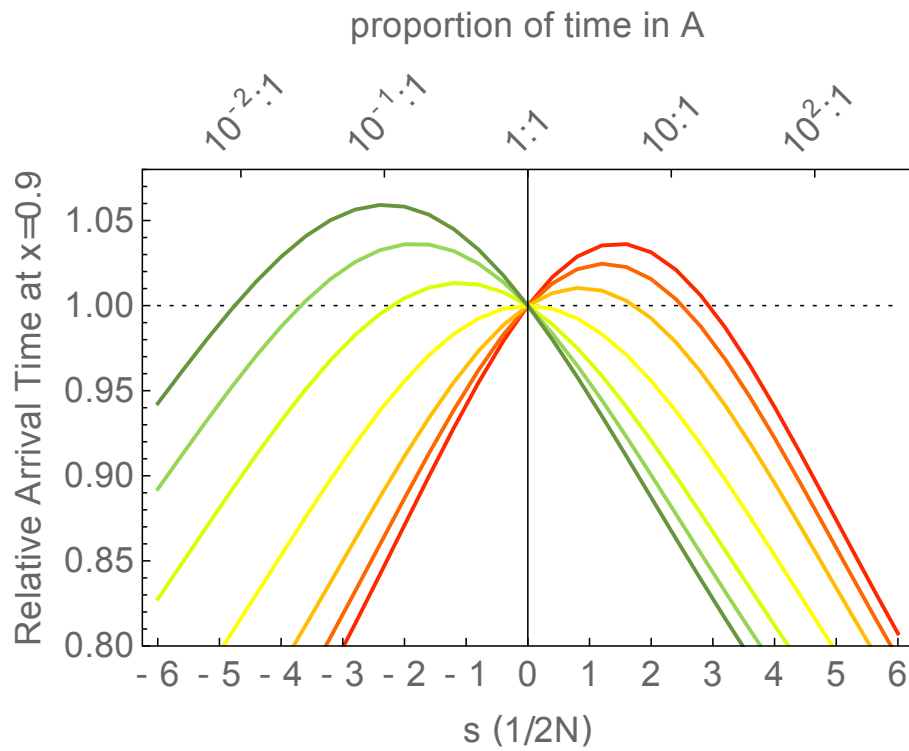


Figure S2: Average time till frequency 0.9 is reached, relative to neutrality for different selection (x-axis) and dominance (y-axis) coefficients, for a population of size  $N_e$ , calculated numerically using diffusion approximation. Colors indicate values of  $h$  corresponding to figure 1, from red to the green equal to 1, 0.9, 0.75, 0.5, 0.25, 0.1 and 0.



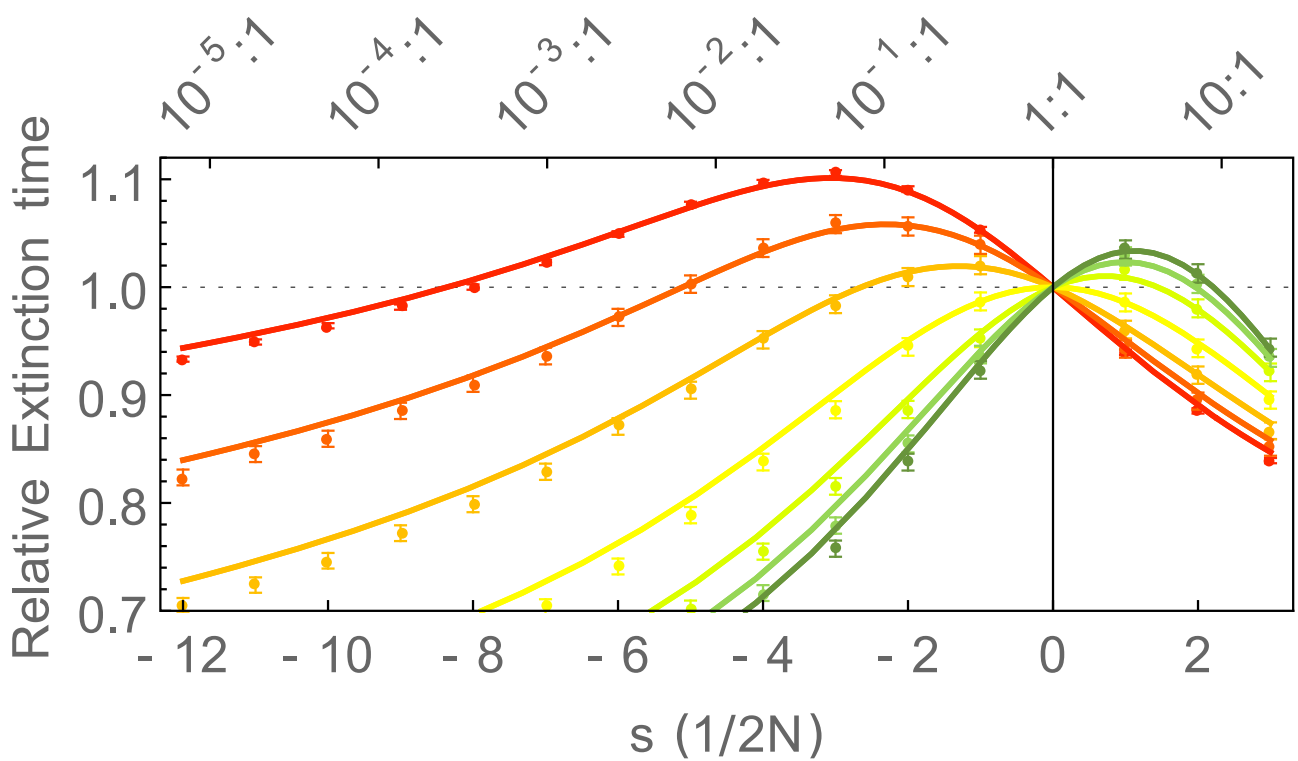


Figure S3: Diffusion approximation (continuous lines) and the average of  $10^7$  simulations (dots) for the extinction times relative to neutrality for a single initial allele in a population with  $Ne = 50$ . Error bars indicate the 95% confidence interval. Colors indicate values of  $h$  corresponding to figure 1, from red to the green equal to 1, 0.9, 0.75, 0.5, 0.25, 0.1 and 0.

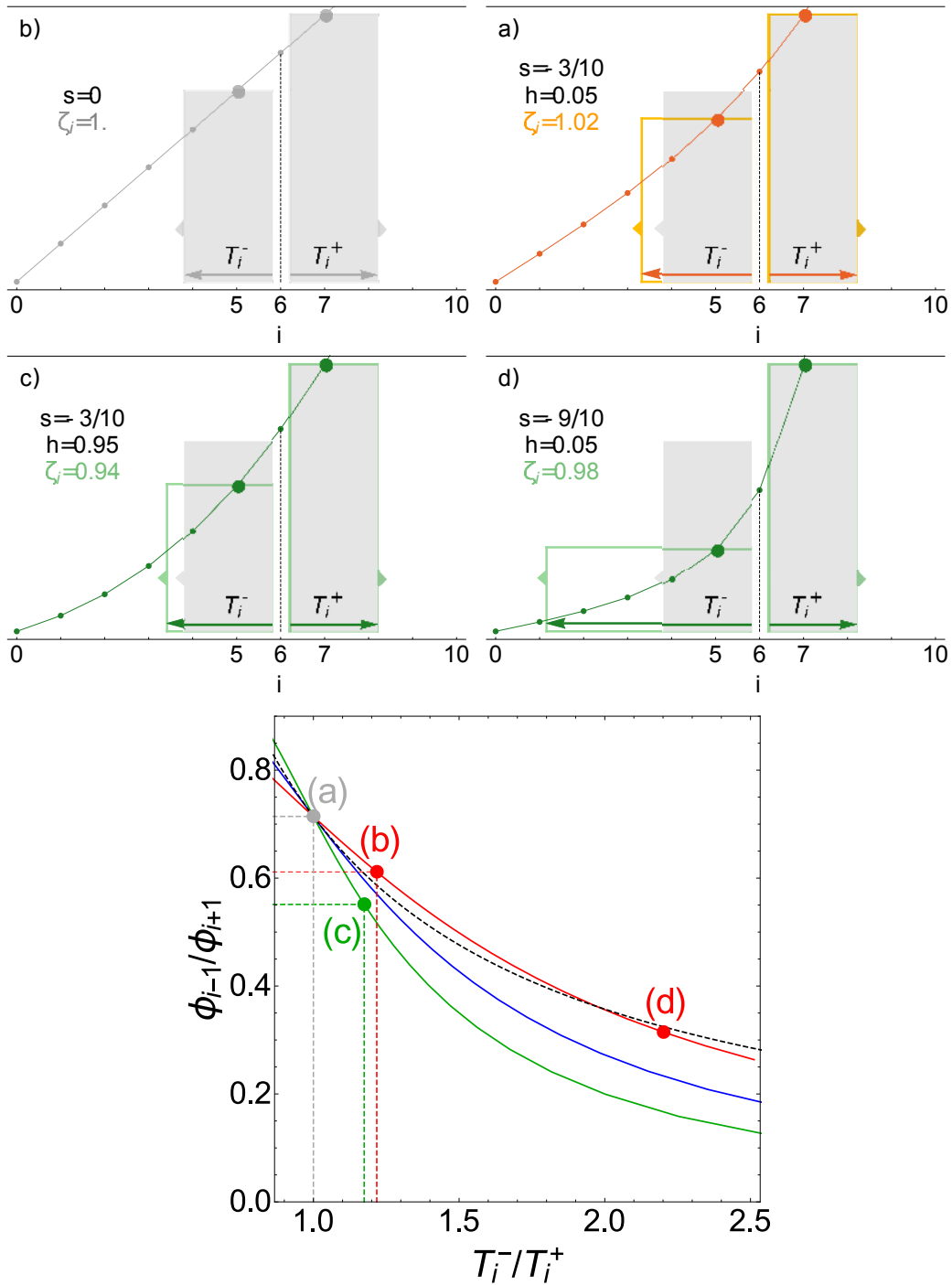


Figure S4: Jump process with  $2N_e=10$  for  $i=6$  (black vertical line) a deleterious A-alleles. Fig.(a-d) compare the probabilities of decreasing in frequency ( $i \rightarrow i-1$ , left-rectangles) or increasing ( $i \rightarrow i+1$ , right-rectangles) given fixation, when  $A$  is neutral (a), weakly selected and recessive (b) or dominant (c), and strongly selected and recessive(d). For all figures colored silhouette refers to the different values of  $h$  and  $s$ , while full gray rectangles indicates neutrality, as in figure 2. (e) Why can the left-rectangle be larger than for neutrality? Recessivity hides the effects of a deleterious alleles comparatively more at lower frequencies, hence as an allele becomes more deleterious and  $|s|$  increases from 0, (moving away from the gray dot along the curves), the relative probability of fixation at  $i-1$  and  $i+1$  (y-axis) increases more rapidly than the the probability of a decrease in  $i$  (x-axis, chance for step  $i=6 \rightarrow 5$  relative to  $6 \rightarrow 7$ ). The colored curves indicate varying varying selective coefficients  $s$  (increasing from bottom to top) for  $h = 0.95$  (red),  $h = 0.5$  (blue) and  $h = 0.05$  (green), while points (a-d) correspond to the respective figures. Hence as  $s$  becomes more deleterious from zero (gray dot), the red curve goes above black curve (neutrality). Dotted lines coincides with the ratio of the areas of the rectangles in (a-d), i.e.  $P_i^-/P_i^+$ , hence the neutrality curves indicate combinations of  $T_i^-/T_i^+$  and  $\phi_{i-1}/\phi_{i+1}$  that provide equivalent  $P_i^-/P_i^+$  to the neutral case.

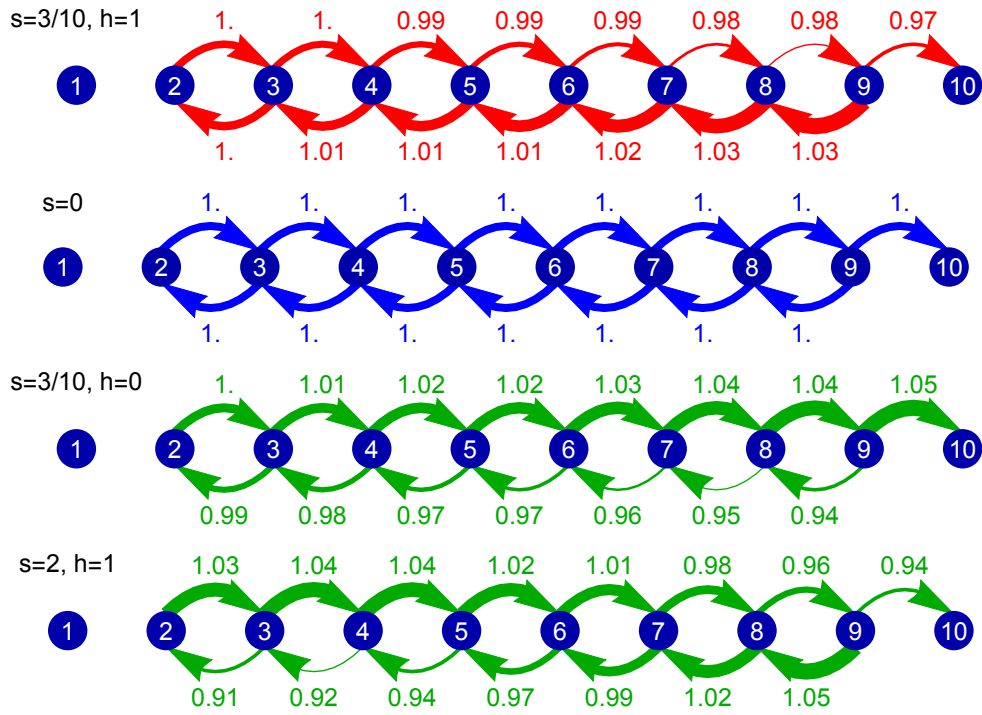


Figure S5: Transition probabilities conditional on fixation of an increase (top arrows) or a decrease (bottom arrows) in  $A$  alleles relative to neutrality, for the simplified Moran process with  $2N = 10$  and values of  $h$  and positive selection corresponding to Fig.2.

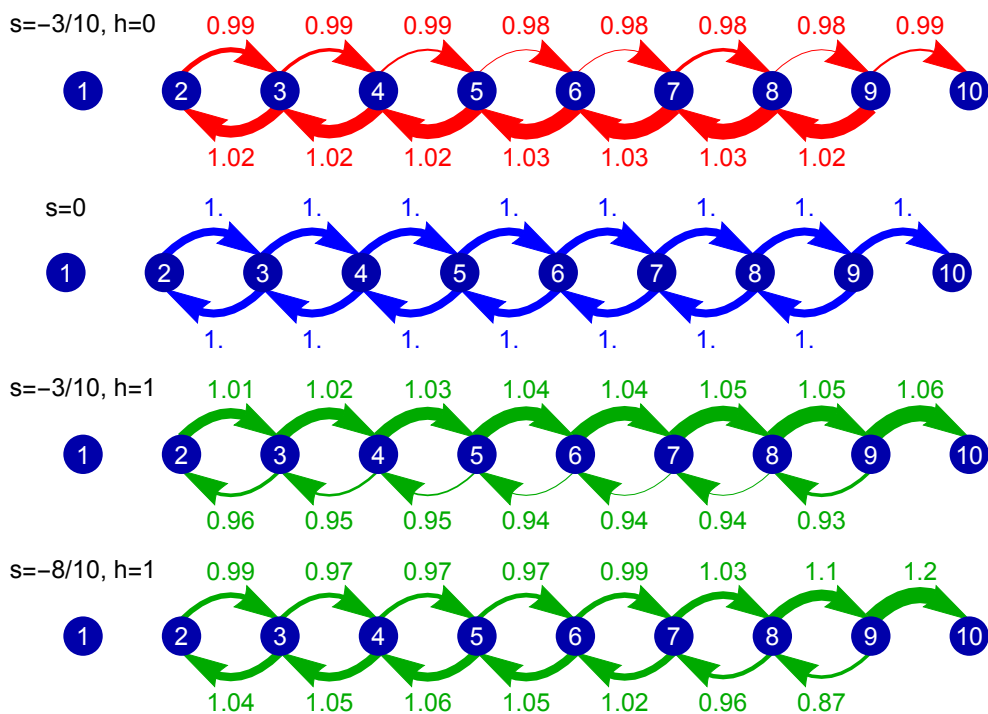


Figure S6: Transition probabilities conditional on fixation of an increase (top arrows) or a decrease (bottom arrows) in  $A$  alleles relative to neutrality, for the simplified Moran process with  $2N = 10$  and values of  $h$  and negative selection corresponding to Fig.S4



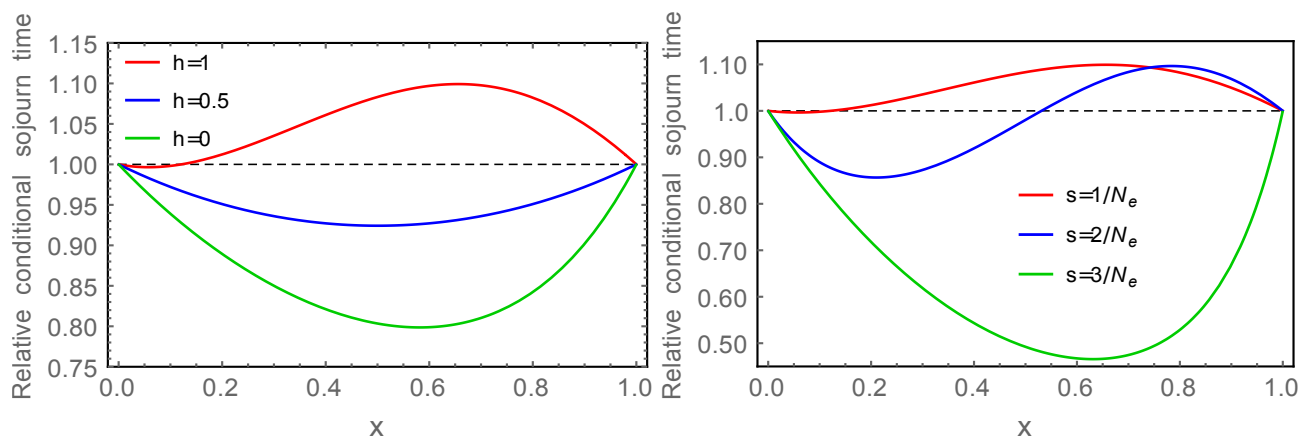


Figure S7: Diffusion approximation of the conditional sojourn time relative to neutrality at different population frequencies ( $x$ ). On the left side the sojourn times are shown for  $s = 1/N_e$ , for a single emerging allele in the case of complete dominance (red), codominance (blue) and recessivity (green). On the right side the sojourn times are shown for a completely dominant allele with different values of  $s$ . In the case of weak selection and  $h > 0.5$  the conditional sojourn times are longer than under neutrality, especially at higher population frequencies.

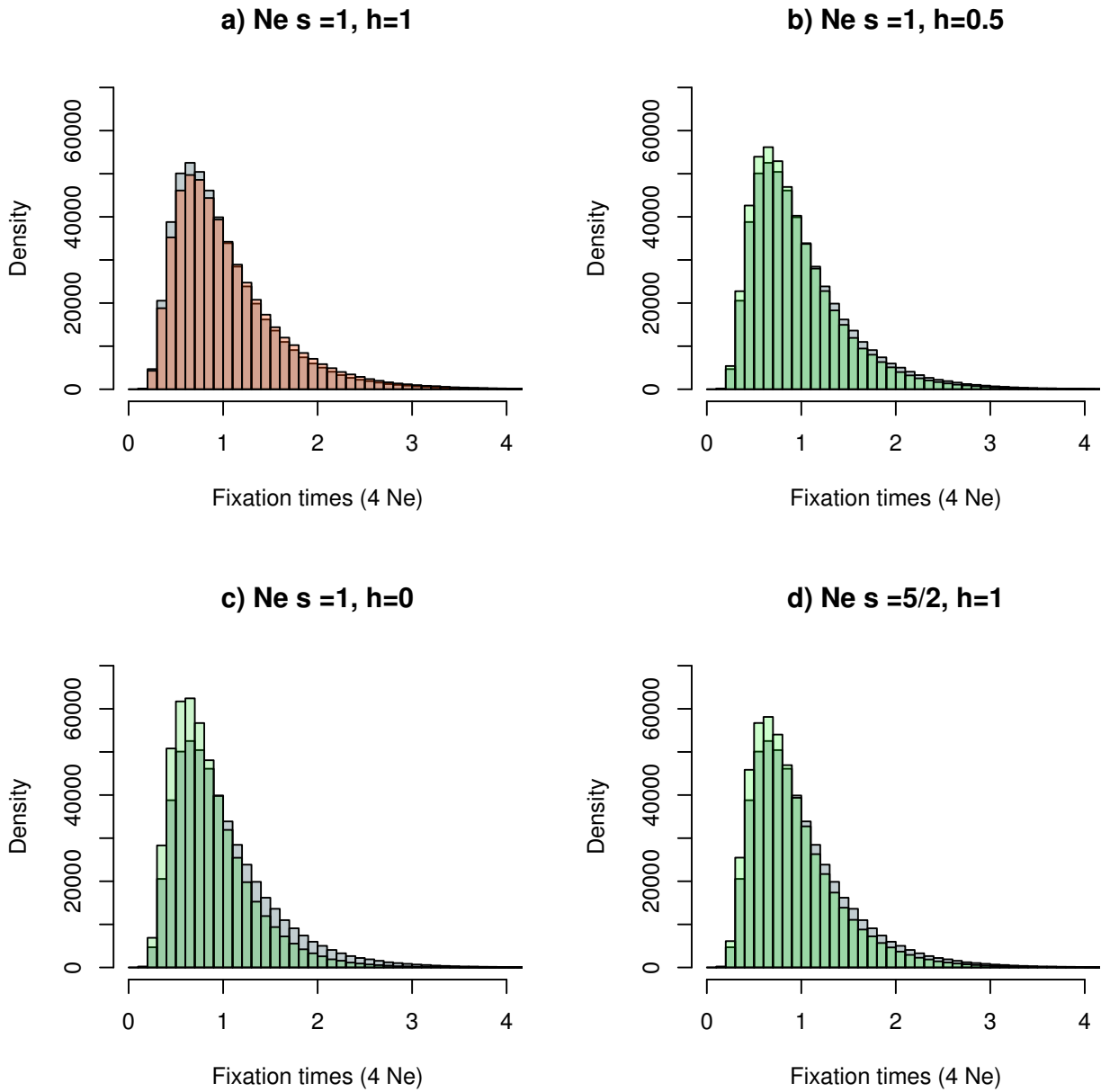


Figure S8: Distribution of fixation times for  $5 * 10^5$  Wright-Fisher simulations with  $N_e = 10^5$  and parameters as indicated in the figures. Neutral distributions are indicated in gray, while red and green indicate combinations of  $s$  and  $h$  for which an increase (a) or a decrease in conditional fixation time is expected, respectively (b-d).

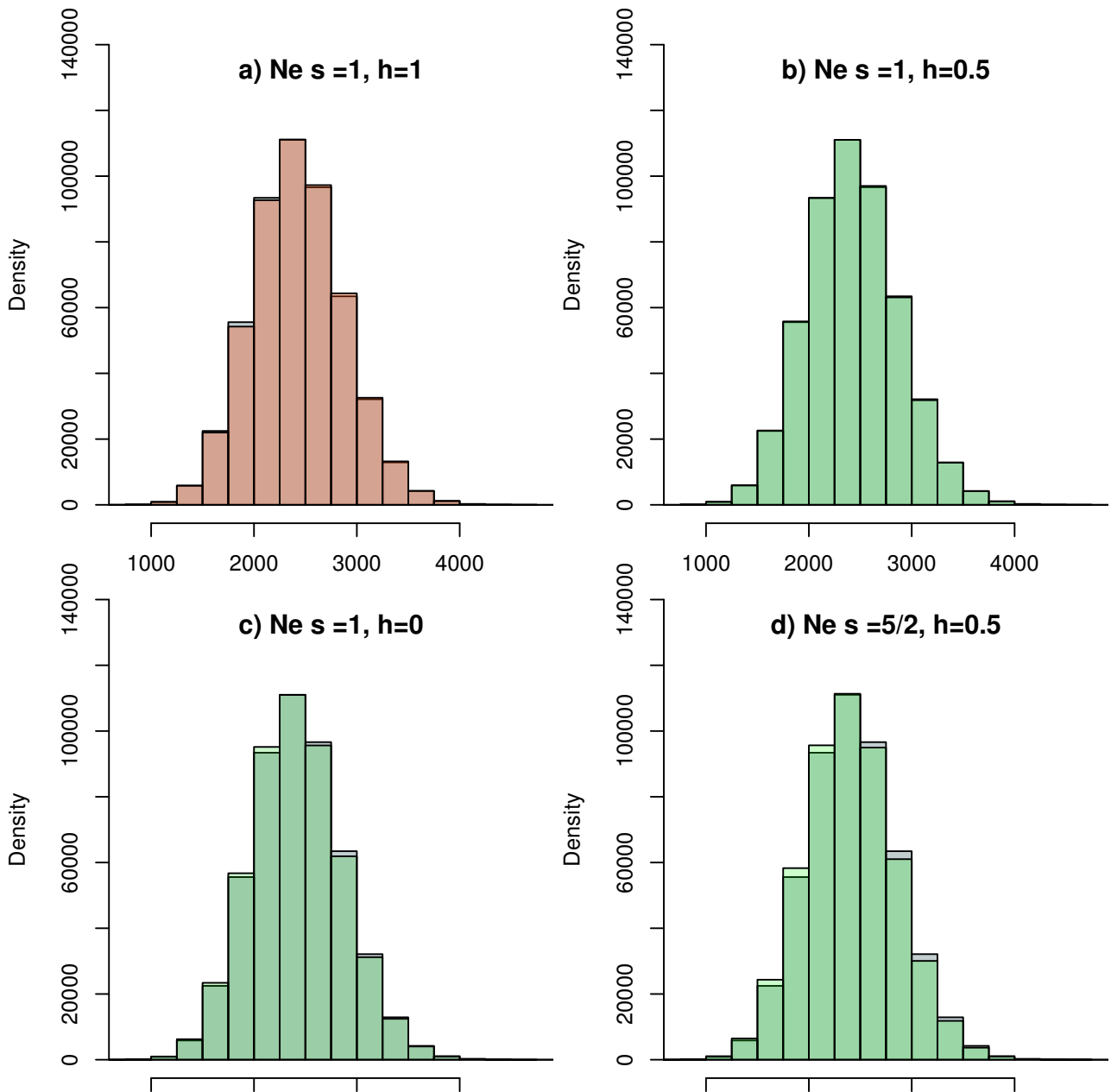


Figure S9: Distribution of  $\pi$  for  $5 * 10^5$  Wright-Fisher simulations with  $N_e = 10^5$  and parameters as indicated in the figures. Neutral distributions are indicated in gray, while red and green indicate combinations of  $s$  and  $h$  for which an increase (a) or a decrease in conditional fixation time is expected, respectively (b-d).

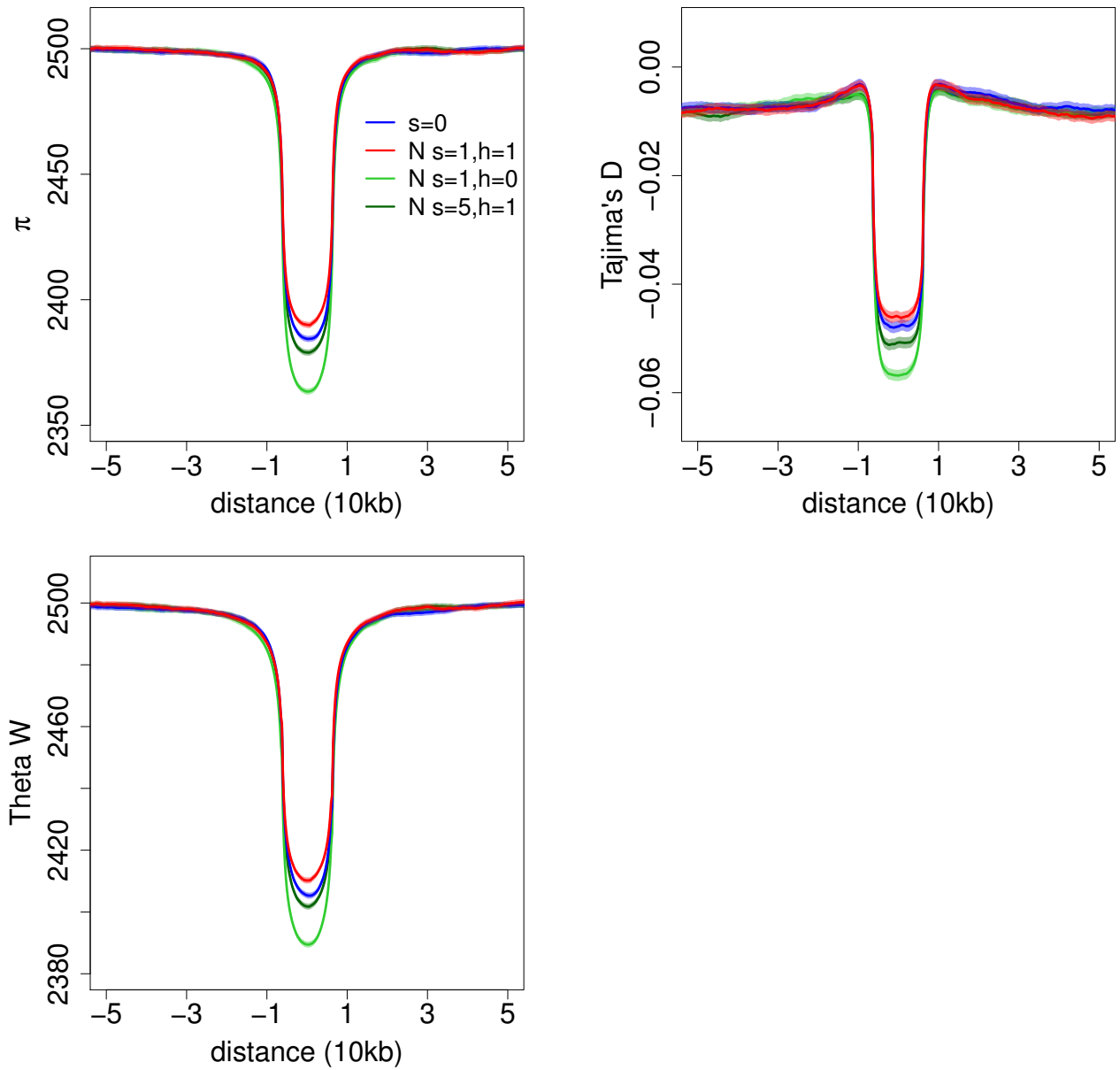


Figure S10:  $\pi$ , Tajima's D and Watterson's  $\theta$  around the site of a selected allele that just reached fixation. We considered a neutral (blue), dominant (red) and recessive (dark green) weakly selected allele, and a dominant allele with stronger selection (light green) as indicated in the legend. Shaded regions indicate the 95% confidence interval of  $5 \times 10^5$  simulations for a population with  $N_e = 10000$ , mutation rate  $1.2 \times 10^{-8}$ . Note that as expected a reduction in diversity occurs around a neutral site as well after fixation [?].



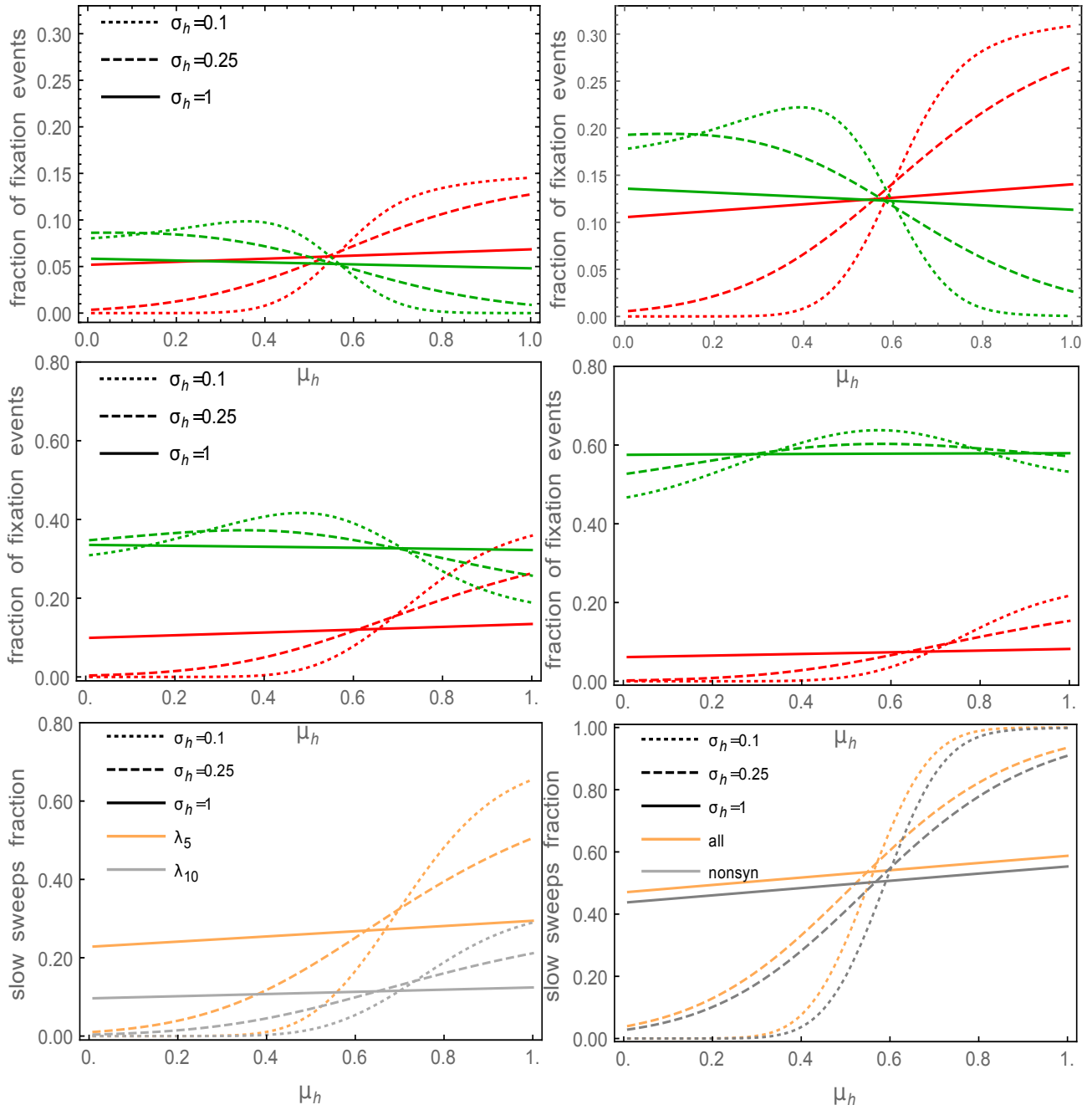


Figure S11: Fraction of fixation events (y-axis) due to drift and selection with longer (red) and shorter (green) fixation times than neutrality for different distribution of selective effects (DFE) and dominance. We exclude overdominance and underdominance cases, assuming dominance to be distributed following a truncated normal distribution with mean  $\mu_h$  and variance  $\sigma_h^2$ , i.e.  $\hat{N}(h) = N(h) / \int_0^1 N(h) dh$  for  $0 \leq h \leq 1$ , with  $N(h) = \frac{1}{\sqrt{2\pi}\sigma_h} e^{-\frac{(h-\mu_h)^2}{2\sigma_h^2}}$ . Selection follows the DFE estimated in Racimo and Schraiber[?] at the genome wide level, for all mutations (a), and non-synonymous mutations (b). We also considered exponentially distributed selection coefficients, with mean 5 (c) and 10 (d) higher than estimated in [?]. For each value of  $s$  and  $h$  we calculated the fraction of neutral fixation events and selective ones as  $1/2N_e$  and  $P_1(1/2N_e) - 1/2N_e$ , respectively.

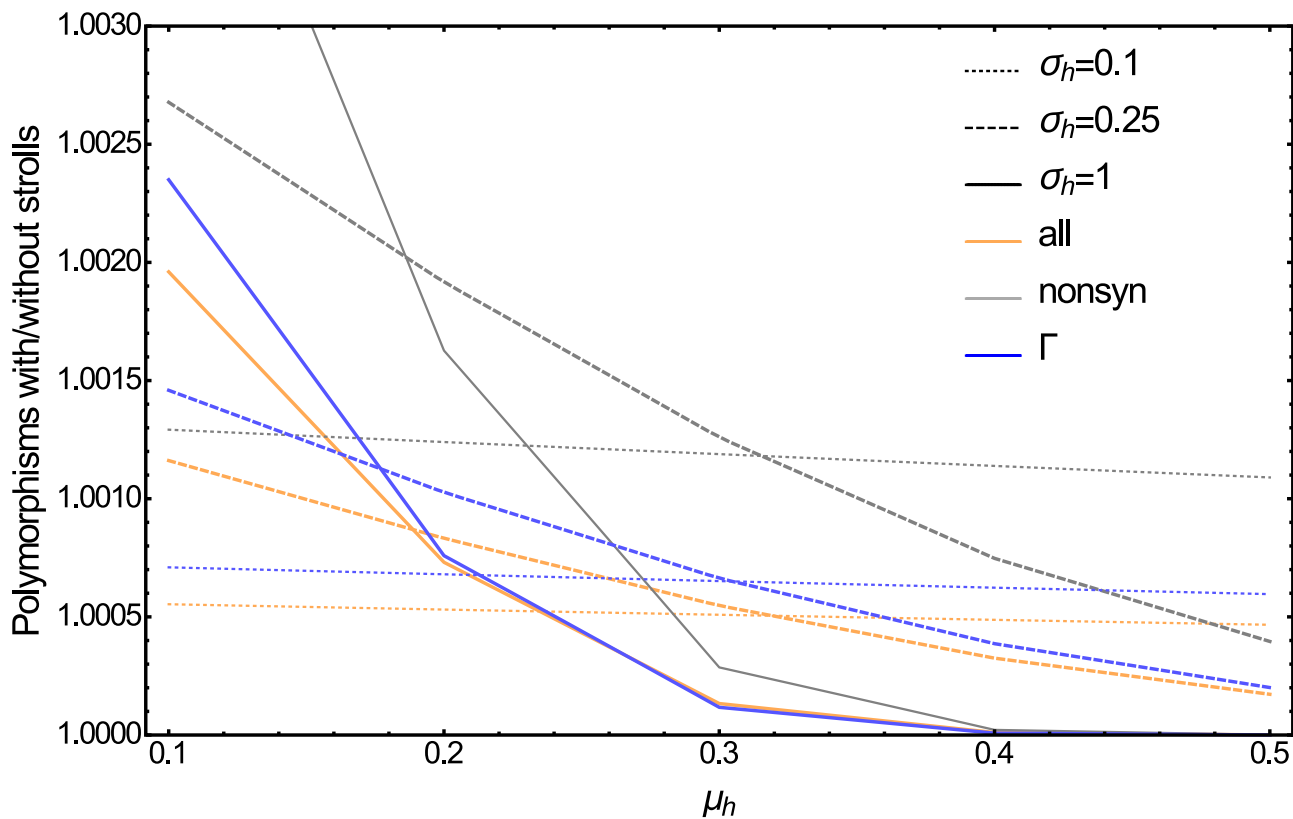


Figure S12: Total excess of polymorphisms due to selective strolls at  $x = 0.1$ , for distribution of dominance as in Fig.S11, and selection coefficients distributed as in Racimo and Schraiber [?] (orange and gray, for all mutations and non-synonymous ones) or with a Gamma (blue) distribution as in Charlesworth and Eyre-Walker [?] (see Fig.S13). The x-axis represents an increasing mean of the dominance distribution, while the y-axis the ratio between the amount of expected amount of polymorphisms and the amount of polymorphisms without strolls. The latter is calculated by assuming that all classes of mutations, for all frequencies, have sojourn times at most equal to those under neutrality, i.e.  $\int_s \int_h \int_x t_{\text{nostrolls}} = \min(t(s, h), t(s = 0))$ . Note that this is a substantial underestimate of the effects of the stochastic slowdown, since only the excess of time spent above neutrality is considered (and not the fact that selective strolls lead deleterious mutations to survive at least *as long as* neutral ones). Different dashings indicates different levels of variance in the distribution of dominance.

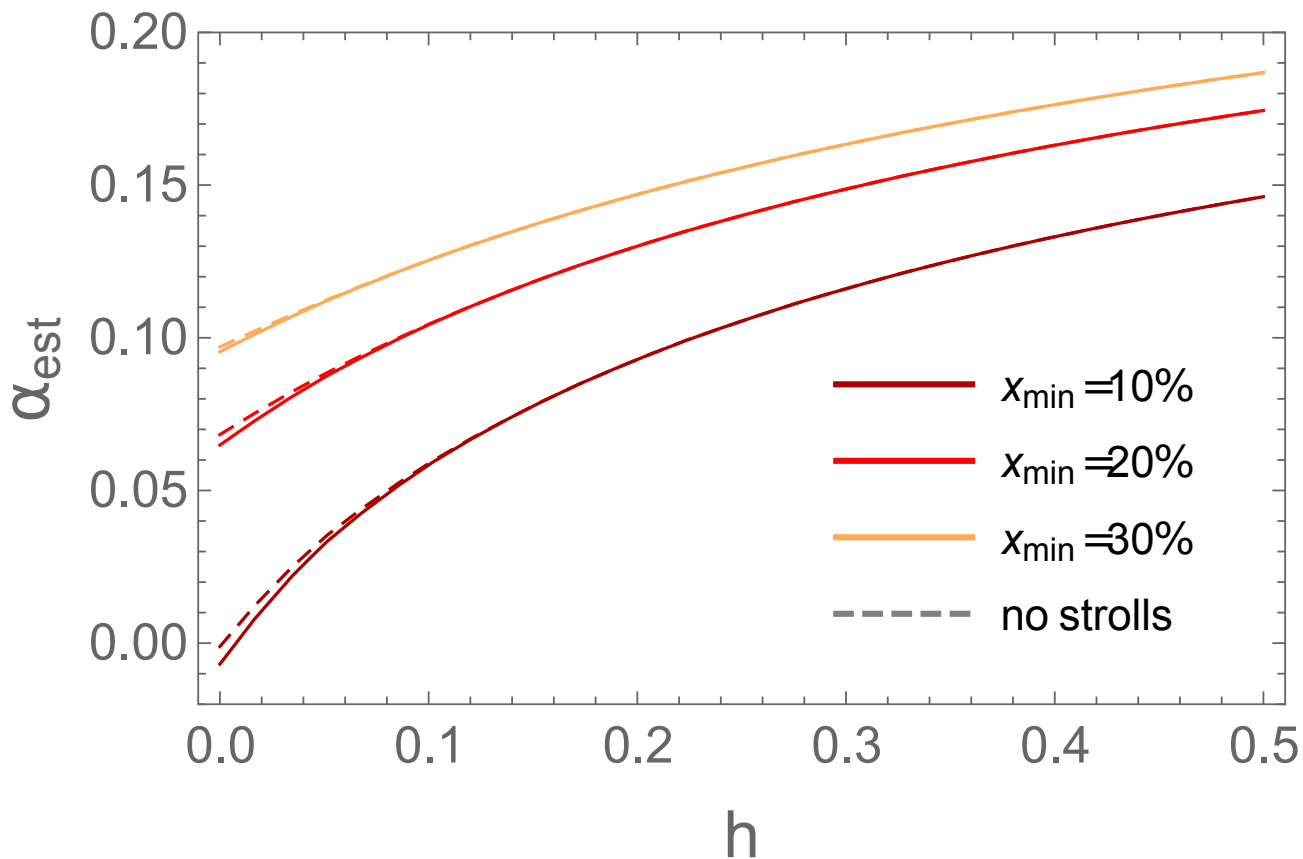


Figure S13: Estimated fraction of adaptive events due to strong positive selection ( $\alpha_{est}$ ) (y-axis) for various values of  $h$  (x-axis), by assuming a true value of  $\alpha = 0.25$ . Different colored lines indicate McDonald-Kreitman tests performed by filtering out alleles with frequencies below 10%, 20%, or 30%. The contribution of *selective strolls* is determined as in Fig.S12, by assuming that all classes of mutations, for all frequencies, have sojourn times at most equal to those under neutrality, i.e. by neglecting the excess of polymorphisms due to the *selective strolls* (dashed lines). Hence the difference on the y-axis between dashed and continuous lines represents the bias due to the strolls. For the distribution of fitness effects we followed [?], by assuming a gamma distribution with shape parameter 0.125 and mean selection strength such that the ratio of non silent and silent divergence equals 0.2.

## File S1. References

- [1] F. Racimo, M. Kuhlwilm, and M. Slatkin, “A test for ancient selective sweeps and an application to candidate sites in modern humans,” *Molecular Biology and Evolution*, p. msu255, Aug. 2014.
- [2] J. Charlesworth and A. Eyre-Walker, “The McDonaldKreitman Test and Slightly Deleterious Mutations,” *Molecular Biology and Evolution*, vol. 25, pp. 1007–1015, June 2008.