Beyond Recombination: Exploring the Impact of Meiotic Frequency on Genome-wide Genetic Diversity

Louis OLLIVIER[1], Brian CHARLESWORTH[2] and Fanny POUYET[1]

[1] Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, 91190 Gif-sur-Yvette, France

[2] Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, Edinburgh EH8 9PU, UK

Corresponding author
E-mail: fanny.pouyet@universite-paris-saclay.fr

Keywords Population genetics, recombination, genetic hitchhiking, simulations, yeast, genetic diversity

Abstract

An important aim of population genetics is to elucidate the processes affecting genetic diversity across regions of the genome and across species. Canonical population genetic models of sexually reproducing species define the rate of meiotic recombination in terms of the frequency of events per site per sexual generation. This paper studies the interplay of several factors with respect to their effects on neutral genetic diversity in a facultatively sexual, diploid, unicellular species such as yeast. The relevant factors are the prevalence of meiosis versus mitosis, the recombination rate, and the selection and dominance coefficients at loci under positive selection. We assume that many generations of mitotic cell divisions are interspersed with episodes of sexual reproduction, in which all individuals in the population undergo meiosis, followed by random matings among the resulting gametes. Our findings reveal that a single hard selective sweep can reduce neutral nucleotide site diversity across the entire genome, provided that the frequency of meiotic events is sufficiently low (half or less than half of the frequency with full sexuality), and that the effects of a selective sweep on levels of neutral diversity at sites linked to the target of selection can be markedly different from those predicted by standard models of sweeps. Species that reproduce by facultative sex are thus likely to exhibit unusual patterns of genetic diversity.

Author summary

In this study, we explored how different sexual strategies influence genetic diversity. Specifically, we looked at how the balance between sexual reproduction (which involves meiosis) and asexual reproduction (which involves mitosis) affect genetic variation in a species like yeast. Our research focused on the effects of recombination rates and the role of selective sweeps — when a beneficial genetic variant spreads rapidly through a population — on genetic diversity. We found that, in species with facultative sex, a selective sweep can dramatically reduce genetic diversity across the genome, but only if sexual reproduction occurs infrequently. We compared our findings to existing models and also developed a new

mathematical framework for understanding the effects of sweeps when sexual reproduction is rare. These results suggest that facultative sex can reduce genetic diversity over the entire genome, which could complicate inferences about population size and evolutionary dynamics in species with mixed reproductive strategies.

Introduction

A major goal of population genetics is to identify and quantify the mechanisms that shape genetic variation, by comparing data on DNA sequence variation in natural populations with theoretical predictions, including computer simulations [1]. In addition to coalescent models [2,3], researchers also use forward-in-time simulations such as SLiM [4] to study complex evolutionary scenarios, including polygenic selection (see, for example, [5,6]) and intricate demographies [7]. All such population genetics models incorporate essential parameters such as mutation rate, population size, selection coefficients, and recombination rate [8,9].

In standard population genetics models, the rate of recombination is expressed in units of events per site per sexual generation. But the passage from one generation to the next is not achieved through meiosis alone, and non-meiotic cell divisions exist in most eukaryotic species. In particular, many organisms display facultative sex. For instance, in the budding yeast *Saccharomyces cerevisiae*, cells can regulate their division strategy, and a sexual generation with meiotic recombination occurs only once every 500 to 10,000 generations [10]. Studies on natural populations of its relative *S. paradoxus* suggest that the frequency of sexual reproduction in that species ranges between 0.001 and 0.0001 per cell division [11,12].

Research on the population genetics of facultative sex remains limited compared to models of obligate sexual or asexual reproduction. For instance, neutral coalescent models of the effects of facultative sex show that gene conversion between homologous chromosomes in the asexual phase can substantially reduce neutral genetic diversity in such systems [13]. This counteracts the effect of long-term asexuality, which causes long

coalescence times for a pair of alleles within the same individual, thereby increasing the net level of sequence divergence among pairs of alleles [14]. Despite these advances, current models of facultative sex [5,13,15] have largely overlooked the role of natural selection in shaping genetic diversity in this context.

In this paper, we explore the interplay between the frequency of sex and other evolutionary forces within the framework of genetic hitchhiking, a term introduced by Maynard Smith and Haigh in 1974 in the context of the spread of a beneficial mutation [16]. More generally, hitchhiking refers to the reduction of genetic variation at linked neutral sites that occurs in the vicinity of a variant under selection. It can either be caused by a linked beneficial mutation that increases in frequency in a population (a selective sweep) or by linked deleterious mutations that decrease in frequency (background selection) [17]. The recombination rate is a core parameter affecting the strength of hitchhiking, due to the effect of recombination in breaking down associations between neutral and selected loci [9,16].

It is well established that recombination rates vary across the genome of a single individual, between individuals of the same species, and between related species – for a review, see [18]. Interestingly, it is still unclear why a limited range of genetic diversity is observed between species, even among those with different census population sizes (Lewontin's paradox) [19–21]. In species like yeast, the frequency of meiotic generations (where recombination events are frequent) versus mitotic generations (where they are rare or absent) must affect their levels of genetic diversity, which will probably be much lower than in species where sex is obligatory, due to the greater effects of hitchhiking.

To start to address this question, we have developed a simulation-based framework of selective sweeps, which includes both the recombination rate per site per meiosis and the frequency of meiosis relative to mitosis. We assume that a population of diploid cells reproduces mitotically, with all members of the population occasionally undergoing meiosis followed by random mating of the haploid gametes, *i.e.* outcrossing; this event is followed by another cycle of mitotic cell divisions. This is intended to capture the properties of the budding yeast *Saccharomyces cerevisiae*, a well-characterized facultatively sexual species

that enhances sexual reproduction in stressful environments, such as under nutrient deprivation [22]. While other reproductive strategies, such as haploselfing, occur in *S. cerevisiae* [10], for simplicity this study concentrates on outcrossing and the clonal reproduction of diploid individuals, in order to provide a proof of concept. This choice allows us to focus on the key processes affecting genetic diversity without the added complexity of alternative reproductive modes.

We chose *S. cerevisiae* because its evolutionary parameters, including mutation rates, have been estimated from laboratory experiments [23], and its high diversity in wild populations [24,25] suggests a large effective population size, which typically supports the efficiency of natural selection unless hitchhiking is pervasive. We show that a low frequency of meioses can create a footprint of natural selection at a genome-wide level, and that its effects on neutral diversity are not necessarily accurately predicted by the standard models of selective sweeps.

Results

A low meiotic frequency can cause a sweep to affect the entire genome

To explore how the frequency of sex and recombination influences genetic diversity after the completion of a selective sweep, we first investigated the interaction between selection and the frequency of sexual reproduction (*i.e*, meiosis), denoted by α. We compared a case with obligatory sex (α = 1) to one with facultative sex (α = 0.01, where the interval between two sexual generations was fixed at $1/\alpha$ = 100 generations). We tested whether the meiotic frequency affects neutral genetic diversity (π) by first comparing a control case (the absence of selection) to a test case with a positively selected site in the middle of the simulated chromosome. The product of the frequency of recombination per basepair per meiosis (ρ_α) and α is denoted by the parameter ρ. For purposes of comparisons among different values of α, ρ is kept constant here; it can be regarded as measuring the net rate of recombination per

basepair over all generations, mitotic and meiotic. A single beneficial mutation with selection coefficient *s* and dominance coefficient *h* is introduced into a population of *N* diploid individuals and allowed to proceed to fixation, such that all individuals in the population are homozygous for the mutation (see Material and Methods for full details).

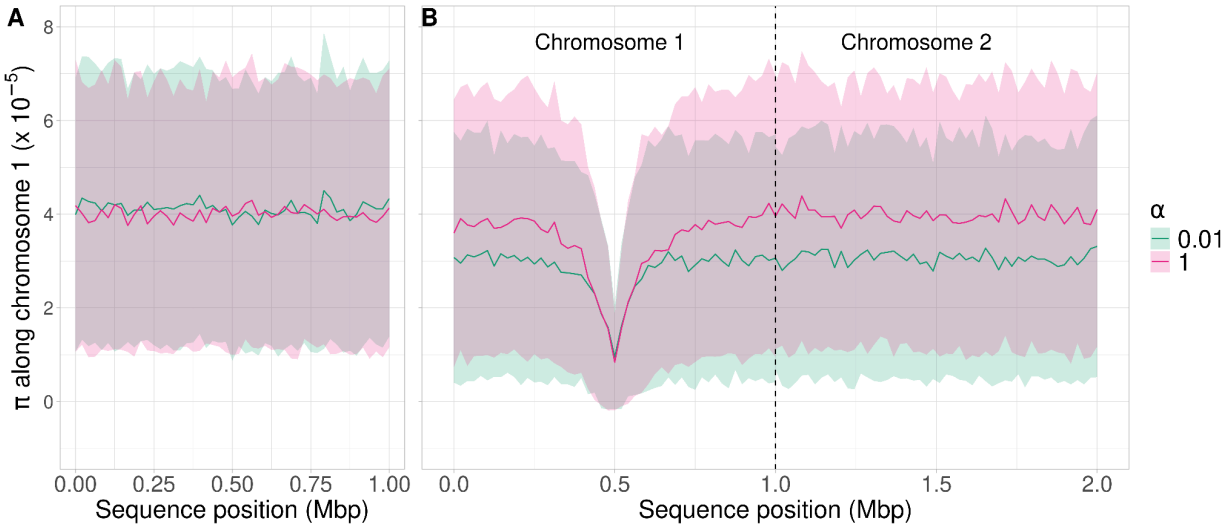


Fig 1. Variation in π along a 1 Mb chromosome with $\rho = 5 \times 10^{-8}$, and along an independent 1 Mb chromosome.

(A) A population of 1,000 individuals that has evolved neutrally for 10,000 generations. (B) A beneficial mutation ($h = 0.5$, $s = 0.05$) at position 0.5 Mb of the first chromosome was introduced into one individual and the population was followed until its fixation. If the mutation is lost before fixation, the simulation is rerun. The second chromosome carried only neutral mutations. In both panels, the pink line is the mean π over 500 simulations with $\alpha = 1$, and the green line is the mean π with $\alpha = 0.01$, such that the interval between two sexual generations is $1/\alpha = 100$ generations. The shaded areas represent standard deviations (see Material and Methods for further details).

In the absence of selection, the nucleotide site diversity across the genome was similar for both frequencies of meiosis, with mean $\pi = 4.01 \times 10^{-5}$ for $\alpha = 1$ and 4.11×10^{-5} for $\alpha = 0.01$ (Fig 1A). This difference is small, less than 3%, but significant ($p = 0.00011$, unilateral Student's t-test). This similarity between the two cases was expected, because recombination has no effect on the magnitude of neutral variability. The slight difference is due to the fact that, when sexual reproduction is rare, a pair of alleles sampled from the same individual coalesce more slowly than a pair from different individuals, resulting in an increase in π

[14,15,26].

When a beneficial mutation was introduced, there was a large loss of genetic diversity around its location, as expected from canonical sweep models [27] (Fig 1B). For both values of α, the diversity valley was centered around the sweep, spanning approximately 0.25 Mb on either side of the selected site. Beyond this region, diversity reached a plateau. In addition, π at the location of the selected site ≈ 9×10^{-6}, about 22% of the value in the absence of selection. The mean numbers of generations taken for the beneficial mutation to reach fixation were similar for the two cases, but the time to fixation, t, was significantly higher (> 3%) with $\alpha = 0.01$ ($t = 430$ for $\alpha = 0.01$ and $t = 418$ for $\alpha = 1$ with $p = 0.026$, Mann–Whitney U test, S1 Fig). The results with a fixed α (Fig 1B) did not differ significantly from those with random variation in the interval between meioses (S2 Fig).

The t value for $\alpha = 1$ is close to the theoretical value t_s for a semidominant mutation (h=0.5) with homozygous selection coefficient s in a sexual population and $2N_e s >> 1$, derived by Hermisson and Penning in 2005 [28]:

$$t_s \approx 4s^{-1} ln[\left(2N_e s\right) + 0.5772] \qquad (1)$$

Here N_e is the effective size of the sexual population (N_e = 1,000 in the present case). With s = 0.05, $t_s \approx 414.6$. In contrast, with $\alpha = 0.01$, t is increased to 430 generations, which reflects the fact that fixation requires at least one meiotic event to produce homozygotes for the beneficial allele (see S1 Appendix).

To assess whether the difference in diversity between the two conditions extends further across the genome, we simulated a second chromosome that segregated independently of the first, and which carried only neutral mutations. Diversity on the second chromosome was reduced from 4.01×10^{-5} with $\alpha = 1$ to 3.06×10^{-5} with $\alpha = 0.01$ ($p < 2.2 \times 10^{-16}$, one-sided Mann-Whitney U test). The diversity plateau seen on the first chromosome was conserved along the second chromosome, so that the fixation of the beneficial allele affected diversity across the entire genome. The effect of selection thus extends across the whole genome when α is small (Fig 1B). At first sight, this behavior might seem unexpected,

since the net frequency of recombination events per basepair per generation, as measured by ρ, is the same for both α values. However, the loci on the 2nd chromosome recombine freely with the target of selection at each meiosis, and they are not constrained by ρ. Facultative sex thus reduces their effective frequency of recombination with the target of selection to $\frac{1}{2}\alpha$, which is the same as the upper limit to the rate of recombination between loci on the first chromosome (see Equation 3 in the Material and Methods).

How other evolutionary parameters modulate the effect of the frequency of meiosis on genome-wide diversity

The extent to which the meiotic frequency α and the selection coefficient s affect π was examined using intermediate values of α (0.02 and 0.1) and s (0.02 and 0.1). To better reflect natural processes, we incorporated variability in α by replacing the fixed interval assumption by sampling the time to the next meiotic event from a normal distribution with mean $1/\alpha$ and a standard deviation of $1/(10\alpha)$ (see Material and Methods). We measured the effect on diversity by dividing π by π_0, where π_0 is the baseline neutral diversity, estimated from the mean π for chromosome 2 with $\alpha = 1$. The diversity at the sweep position was the same at the end of the simulations for all α values and depended only on s (Fig 2A-C). This is because the mean times to fixation of the beneficial mutation for a given s were mostly similar across the different α values; though they sometimes differed significantly, but by less than 3% (S3 Fig).

As expected from the established theoretical results on hard sweeps [8,9,16,29], the valley of diversity around the target of selection on chromosome 1 becomes narrower as s becomes smaller (pink lines in Fig 2A-C), but is relatively unaffected by the value of α. In contrast, for $\alpha < 0.1$, π on the neutral chromosome decreases with decreasing α ($p < 0.05$ for pairwise comparisons, Dunn's post hoc test) – see Fig 2D. The similarity between the results for $\alpha = 1$ and 0.1 extends to the value of t, which suggests that the interval between meiotic events with sufficiently large α does not greatly affect the process of fixation of a beneficial allele (S3 Fig). For all α values, t decreases as s increases, as expected from standard

theory (p < 0.05, Dunn's post hoc test, S3 Fig). The mean number of recombination events during a sweep for a given α increases with larger t (see S1 Appendix) allowing π/π_0 to approach one with small s and free recombination, even for the lowest α value (Fig 2D). In contrast, for large s and α < 0.1, recombination events are too rare for diversity to fully recover.

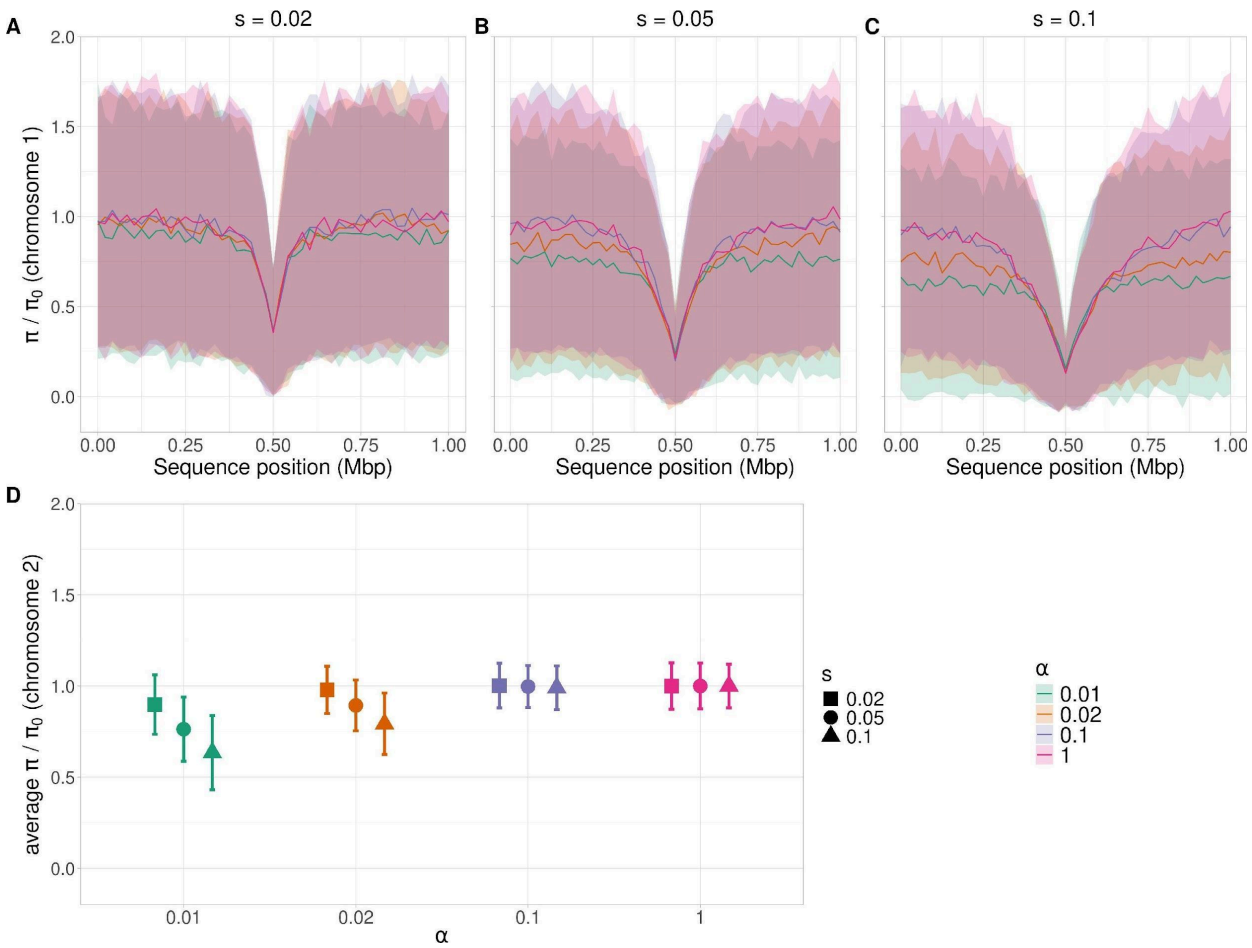


Fig 2. The behavior of diversity relative to the fully neutral value, π/π_0, with different strengths of selection and frequencies of meiosis.

π_0 is the baseline neutral genetic diversity with α = 1 in the absence of selection. Panels A-C show π/π_0 along a 1 Mb chromosome with $\rho = 5 \times 10^{-8}$ and a hard sweep at 0.5 Mb; h = 0.5 and s = 0.02 (A) s = 0.05 (B), and s = 0.1 (C). Panel D shows mean π/π_0 and its standard error for a second, independent, chromosome carrying only neutral mutations. Colors represent π/π_0 with α = 0.01 (green), α = 0.02 (orange), α = 0.1 (blue), and α = 1 (pink). The simulation method was the same as in Fig 1, except that the interval between two sexual generations with α < 1 was drawn from a normal distribution centered on $1/\alpha$ generations

(see Material and Methods for details). Lines in panels A-C are the means from 500 simulations, and the shaded areas are the standard deviations per site. See Material and Methods for details.

In order to interpret these findings, we compared the simulation results presented in Fig 2 to the theoretical predictions of the effects of a sweep in a randomly mating population, given by Equations 14 of [29]. For this purpose, we used the product of α and the recombination rate r per meiosis between a neutral locus and the target of selection (where r is given by Equation 3 in the Material and Methods) in the relevant formulae. This provides a measure of the net rate of recombination per generation between the two loci, and should not be confused with ρ, the net frequency of recombination per basepair per generation. As expected from previous results [29], the model performs well for all three selection coefficients (s) when $\alpha = 1$ (Fig 3 A-C), as this represents the "canonical" case for selective sweeps. Remarkably, the use of $r\alpha$ as a straightforward measure of an effective recombination rate leads to an excellent fit between the theory and the simulation data, except for small values of α with $s = 0.1$, when it starts to deteriorate for reasons discussed below. In the parameter space considered here ($h = 0.5$ and $\alpha \geq 0.01$), the model from [29] is sufficient to predict the reduction in diversity quite accurately, without the need for a more complex model. This finding underscores the utility of incorporating α into the recombination frequency, demonstrating that, even with a quite low meiotic frequency, the standard sweep theory remains robust and reliable for predicting diversity patterns in this context. The plots for different α values, but the same s, in Fig 2 converge when the distance becomes small. The convergence reflects the fact that the recombination rate obtained from equating the product of ρ and distance to d in Equation 3 is the same for all α when d is small, because the recombination rate is then close to the map distance.

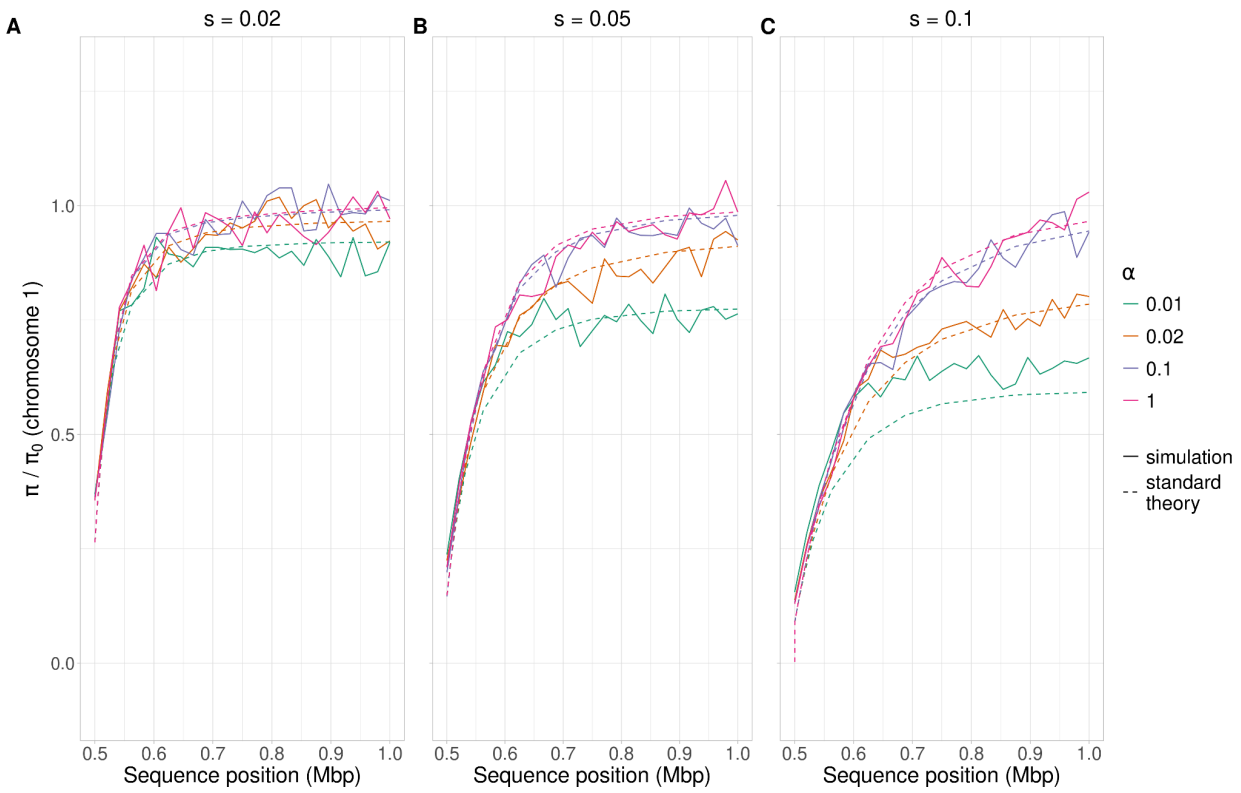


Fig 3. Comparison of the properties of π/π_0 between theory and simulations.

π/π_0 along a 0.5 Mb chromosome (see Methods) with a hard selective sweep at 0.5 Mb; h = 0.5 and $\rho = 5 \times 10^{-8}$ (A) s = 0.02. (B) s = 0.05 (C) s = 0.1. "Standard theory" corresponds to Equations 14 from [29], with the recombination frequency from Equation 3 rescaled as $r\alpha$ (dashed lines). Simulated data are the averages from 500 simulations (solid lines). Colors represent diversity relative to the fully neutral case, with α = 0.01 (green), α = 0.02 (orange), α = 0.1 (blue), and α = 1 (pink).

Additionally, we investigated how the recombination rate (ρ_α) and the dominance coefficient (h) influence genetic diversity during a selective sweep. ρ rather than ρ_α primarily dictates the width of the genetic diversity valley around the sweep (S4 and S5 Figs). The dominance coefficient h has a minimal effect on genetic diversity with α = 1, but its influence becomes more pronounced at lower α, with higher h (a stronger heterozygous effect of a beneficial mutation) causing increased diversity, as is also seen for a sexual population (Fig 1 of ref [29]). We also compared the simulation results with the theoretical model (as shown in Fig 3) for different values of h (S7 Fig). While the fit remains good for α = 1, it deteriorates for

lower α values and becomes less accurate as h deviates from 0.5. The reasons for these discrepancies are not obvious, and will not be discussed here. A detailed analysis of the effects of both evolutionary parameters h *and* ρ_α is given in Supplementary Text 1.

Recovery of diversity during a sweep with a single meiotic event

The standard sweep theory does not accurately predict the genetic diversity (π) for very low α values, as shown in Fig. 4. In these cases, the standard sweep theory using $r\alpha$ as an effective recombination rate strongly underestimates π. The analysis of the case of $\alpha << 1$ in the shows that there is "phase shift" between the behavior of the system with large and small α values, which can be understood as follows. For a neutral site with a recombination frequency r per meiosis with the selected locus, a single meiotic event involving the heterozygous carriers of the beneficial allele has a probability of $2r(1-r)$ of producing homozygotes for the beneficial allele that are non-identical by descent at the neutral site (section 1 of S1 Appendix). Since these homozygotes ultimately replace the other genotypes if there are no further meiotic events, a scenario with at most one meiosis during the initial sweep of heterozygotes for the beneficial allele would give $\pi / \pi_0 = 0.5$ when r takes its maximal value of ½, all other things being equal. However, diversity will increase above this value, due to the accumulation of new mutations at the neutral site (section 2 of S1 Appendix), an effect that is also included in Equations 14 of [29]. This increase is amplified by the fact that, once the meiotic event has occurred, there is a second time interval during which homozygotes for the beneficial allele increase in frequency and become fixed (see Equations S1.5 and S1.8 of S1 Appendix). When multiple meiotic events occur during the sweep, the final π / π_0 ratio will be much higher than predicted by the model based on a single event (A10), and the numbers of such events can be approximated by the product of r, α and the sweep duration, as in the standard sweep model. An approximate analysis of the case of two meiotic events per sweep is given in S2; there is, however, only a relatively narrow range of parameter space to which this model applies.

Fig 4 compares simulation results with small α and large s values with the predictions of Equation S1.10 of the S1 Appendix and those of standard sweep theory. With s = 0.1, h = 0.5, and $\rho = 5 \times 10^{-8}$, the expected time for the initial sweep of heterozygotes for the beneficial allele to go to completion is t_{s1} = 207.2 generations (Equation 1). With α = 0.004, the expected number of meioses per sweep is given by $\alpha\, t_{s1}$= 0.829. Since the requirement of at most one meiotic event during the initial sweep is satisfied, the analytical predictions align well with the simulation results, in contrast to the predictions of standard sweep theory.

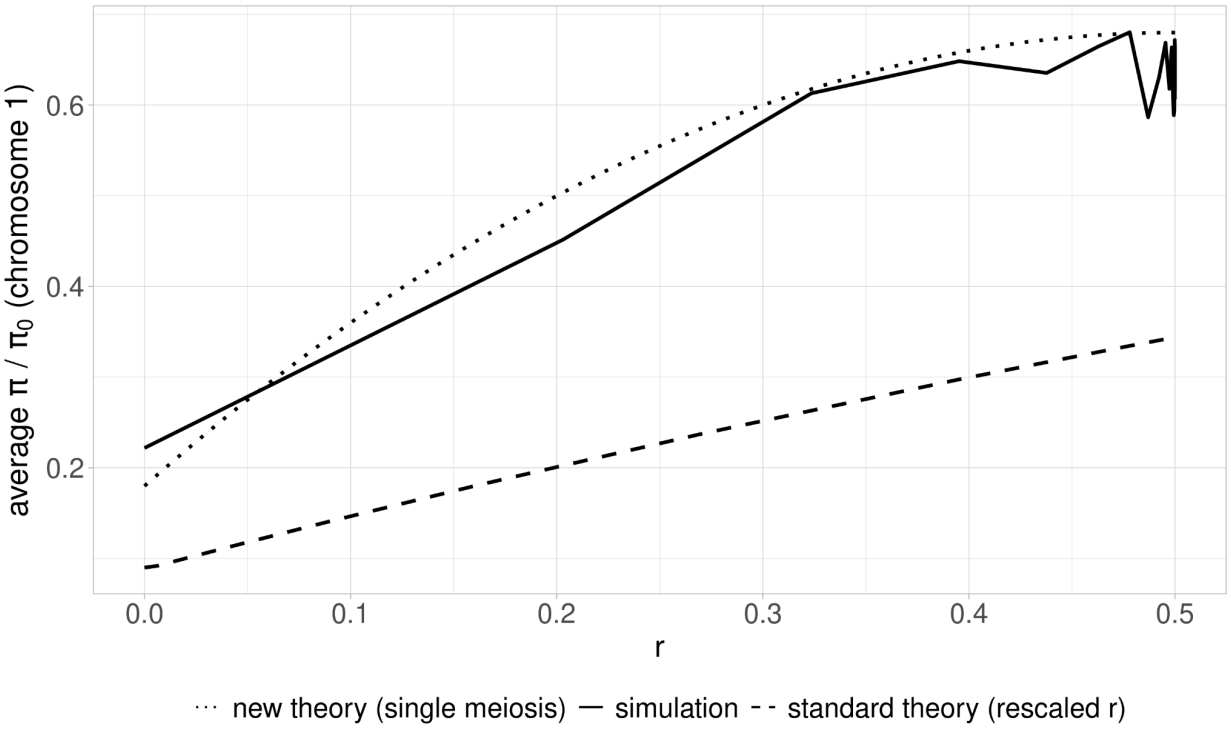


Fig 4. The relation between π / π_0 and r, the recombination frequency per meiosis between the neutral and selected loci, where r is given by Equation 2 in Material and Methods). The lines represent the estimated or simulated values of π / π_0 for small α (0.004). The theoretical values are from Equation S1.10 (dots) or from Equations 14 of [29] with $r\,\alpha$ as the effective recombination rate (dashes). The solid line shows the mean values of π / π_0 over 500 simulations. A beneficial mutation (h = 0.5, s = 0.1) was introduced into one individual at one end of a 0.5Mb chromosome with a recombination rate per basepair of $\rho = 5 \times 10^{-8}$, and the population allowed to evolve until fixation of the beneficial mutation, as in previous figures. The time interval between two sexual generations was drawn from a normal distribution centered on $1/\alpha$ generations. See Material and Methods for details.

Discussion

This study is the first to analyze the effects of a hard selective sweep in a diploid species that reproduces by facultative sex, with long intervals between episodes of sexual reproduction that occur simultaneously in all individuals in the population. Contrary to what is expected from standard models, the effect of a sweep on diversity for a given strength of selection is not simply determined by the net frequency of recombination between a neutral site and the target of selection (*i.e*, by the product of the recombination rate per meiosis and the frequency of sexual generations, α), except when the frequency of meiosis is sufficiently high that several sexual generations occur during a sweep. Otherwise, the expected diversity at a given neutral site is controlled jointly by the probability that a single meiotic event generates homozygotes for the beneficial allele that are not identical by descent at the neutral site and by the time available for diversity to accumulate during the sweep (see Fig 4, the S1 Appendix, and the Supplementary Text 2S2 Appendix). In both cases, a selective sweep can influence not only nearby loci, but also the entire genome, if the number of meioses per sweep is sufficiently small.

The rescaled standard model, where the recombination rate between the selected and neutral locus (r) is multiplied by α, can be most reliably applied when h = 0.5, provided that the $\alpha t_s >> 1$, where t_{s1} is the duration of the initial sweep of heterozygotes for the beneficial mutation given by Equation S1.3b, but it is less accurate when h deviates from 0.5. The use of $r\alpha$ in the sweep equations is reminiscent of the use of r (1 – F) as an approximate effective recombination rate with an inbreeding coefficient of F in standard sweep theory [29]. In contrast, Equation (S1.10) seems to give accurate analytical predictions for situations in which at most one meiosis occurs during a sweep. Interestingly, the major role of $2r(1 - r)$ in Equation S1.10 violates the widely-used principle that properties such as π / π_0 should be determined by the products of N_e and deterministic parameters such as r, rather than their absolute values [8]. This reflects the fact the process controlling diversity when there is at most one meiosis per sweep is largely deterministic rather than stochastic.

These results raise the broader question of how the values of parameters such as α, ρ

and r used in the simulations with a population size of N_s = 1,000 can be related to their values in a natural population with a much greater size, N_n. When αt_{s1} is >> 1 and standard sweep theory is applicable, at least as an approximation, the scaled parameters $2N_e s$ and $2N_e \alpha r$ determine the effect of a sweep on π / π_0. When this is the case, the simulation value of α would also apply to the natural population, whereas s and r should be rescaled by multiplication by N_s/N_n to capture their natural population values (fixation times are correspondingly multiplied by N_n/N_s). In contrast, when αt_{s1} is ≤ 1, standard sweep theory no longer applies, due to the rarity of recombination events. In this case, Equation S1.10 suggests that r should not be rescaled, whereas α should be multiplied by N_s/N_n. This leads to the somewhat unsatisfactory conclusion that the relations between the simulation results and the real world are dependent on the value of αt_{s1} or, equivalently, the product of the scaled parameters $2N_e\alpha$ and $T_{s1} = t_{s1} /2N_e$. Since t_{s1} is largely determined by the reciprocal of hs, the effects on diversity of strongly selected beneficial mutations are likely to be largely controlled jointly by r and $1/(2N_e\alpha)$ (Equation A10), whereas the effects of weakly selected mutations are likely to be controlled by $2N_e s$ and $2N_e \alpha r$.

Further work is required to extend our theoretical understanding to situations where αt_{s1} is not too far from 1, where the transition to diversity that is largely controlled by the recombination rate r and $1/(2N_e\alpha)$ takes place. In addition, the effects of recurrent sweeps need to be studied, together with population genetics statistics such as levels of linkage disequilibrium, the extent of distortions of the shapes of gene genealogies by the effects of selection at linked sites, and the extent of Hill-Robertson interference among the selected sites themselves.

The motivation for modeling *Saccharomyces cerevisiae*, as is done here, stems from its composite life cycle and its status as a model organism. As a facultatively sexual species, *S. cerevisiae* alternates between clonal reproduction and sexual reproduction, making it ideal for exploring the impact of meiotic frequency on patterns of genetic diversity. Although this study focuses exclusively on the effects of outcrossing events, we acknowledge that other modes of sexual reproduction such as haploselfing (the intra-tetrad mating of the products of

meioisis) also play a role in shaping genetic diversity. Many evolutionary parameters for this species, including mutation rates, are well established through laboratory experiments [22,30,31]. Additionally, the large effective population sizes of *S. cerevisiae* and its relatives ensure that natural selection should operate efficiently, providing a robust framework for studying the genome-wide effects of selective sweeps. In future work, it would be useful to extend the model by incorporating the full range of reproductive strategies and life cycle stages of *S. cerevisiae* and its relatives. With the increasing availability of genomic data, addressing questions about the persistence of rare recombination events in yeasts [32] could become another avenue for exploration. Understanding whether these events contribute to maintaining genetic diversity or if they are simply evolutionary byproducts could help explain the evolutionary advantage of maintaining recombination.

Our results challenge the assumption that genome-wide neutral diversity is unaffected by selection. Studies on wild *S. paradoxus* have provided estimates of rates of clonal reproduction as high as 0.9999 (α = 0.0001) [12]. These studies, however, assumed that neutral diversity at synonymous sites is not influenced by selection. Here, we provide evidence that selection can alter global genetic diversity, highlighting the importance of considering selective effects in similar analyses. A limitation is that our study focuses on hard sweeps and does not take into account the fact that most mutations are probably nearly neutral or deleterious in *S. cerevisiae* [33]. It would be desirable for future work to explore the effects of negative selection on genome-wide genetic diversity (π), including the effects of background selection (BGS) [34] and associative overdominance (AOD) [35,36]. These mechanisms, which can either decrease or increase genetic diversity respectively, are influenced by recombination rates and meiotic frequency, with AOD becoming more prominent under low recombination conditions. Understanding the transition from BGS to AOD, and its dependence on α and the recombination rate per meiosis remains an area for further investigation.

This study also provides a new perspective on Lewontin's paradox—the observation

that genetic diversity across species of eukaryotes does not increase as fast with census population size as expected from neutral theory [19]. While previous research has attributed this discrepancy to factors like historical changes in effective population size [20,21], our findings suggest that differences in breeding systems, particularly meiotic frequency, may also play a significant role. Species with a low meiotic frequency, even those with large census sizes, might exhibit reduced genetic diversity compared with species that reproduce fully sexually. This highlights the need to estimate meiotic frequency and analyze its relationship with census size and N_e across species to better understand the dynamics of genetic diversity. By focusing research efforts on meiotic frequency, we can further elucidate the factors driving patterns of genetic variation across diverse. As discussed above, the mean number of recombination events during a sweep increases with longer sweep durations, allowing π/π_0 to approach one when hs is small, even for the lowest α value (Fig 2D). In contrast, for large hs and $\alpha < 0.1$, recombination events are too rare to restore diversity to fully recover. By focussing research efforts on meiotic frequency, we can further elucidate the factors driving patterns of genetic variation across diverse taxa

Material and Methods

Modeling reproduction and meiotic recombination

Two new parameters are introduced compared to canonical population genetic models: the recombination rate per site per sexual generation, ρ_α, and the frequency of sexual reproduction (*i.e*, meiosis), α. We assume an absence of mitotic recombination, as it is four orders of magnitude less frequent than meiotic recombination in yeast [30,37]. In most simulations, we assumed a constant average recombination rate per nucleotide site per generation, weighted by the frequency of meiotic events:

$$\rho = \alpha \, \rho_\alpha \qquad (2)$$

The motivation for this procedure is to examine the extent to which the size of the

reduction in diversity caused by a sweep is determined by the ratio of the recombination rate per meiosis, modulated by the frequency of meiosis to the strength of selection on the beneficial allele, as expected from classical sweep theory (α = 1) [27,29]. When α < 1, reproduction occurs both sexually and asexually; at each generation, all diploid individuals reproduce in the same way. For example, when α = 0.01, meiosis and recombination occur on average once every 100 mitotic divisions. We used values of α = [1, 0.1, 0.02, 0.01] in most simulations.

We modeled the occurrence of sexual generations when α < 1 using two approaches. Initially, sexual generations were set to occur at fixed intervals of $1/\alpha$ generations. Later, this assumption was relaxed to allow variability in α. In that case, we determined the next sexual generation by sampling from a normal distribution centered on x (where $x = 1/\alpha$) with a standard deviation $\sigma = x/10$. This approach enabled the exploration of both regular and stochastic patterns of sexual reproduction.

Simulation methods

We performed forward-in-time, individual-based simulations using SLiM [4]. In a sexual generation, two diploid parents were sampled from the population with a probability proportional to their fitness and allowed to mate to produce a single diploid offspring individual; this procedure was repeated until N individuals were formed. In an asexual generation (clonal reproduction), N offspring were produced by sampling a single parent, drawn from the population in proportion to its fitness. This is a Wright-Fisher sampling scheme, where the effective size of the population is N for a sexual generation and $0.5N$ for an asexual generation [8].

We chose the values of the mutation rate [23] and recombination rate per basepair [30] to mirror *S. cerevisiae* evolutionary dynamics, using the principle that the products of effective population size (N_e) and deterministic parameters control the dynamics of evolution if time is measured in units of $2N_e$ generations [38]. N_e was estimated by plugging the π value from [25]

and mutation rate from [23] into the formula: $N_e = \pi/4\mu$. We thus modeled a plausible natural population effective size of 10^6 as far as diversity is concerned but caution is needed when applying these results to a natural population of this size, as the recombination rates per meiosis for closely linked sites in the simulations are probably much lower than appropriate for modeling a natural population, except when α is very small, given that the mean rate of recombination per Mb in yeast is approximately 0.005 per Mb (0.004 in [39] and 0.0061 in [30]). We simulated two 1Mb independent chromosomes for 1,000 diploid individuals using the reproduction scheme described above. The second chromosome carries only neutral mutations.

In the absence of selection, simulations were run for 10,000 generations. When natural selection was included, a site under strong positive selection was introduced into the middle of the first chromosome of one individual at generation 2,000. Selection acts at each cell division (both mitoses and meioses). The selection and dominance coefficients, s and h, determine the fitnesses of the mutant genotypes relative to wild-type as $1 + hs$ for heterozygotes and $1 + s$ for homozygotes. The beneficial mutation was semidominant, with $h = 0.5$ and $s = 0.05$ ($2Nhs = 50$), unless specified otherwise. Simulations were stopped when the beneficial mutation reached fixation, when all individuals in the population are mutant homozygotes. In the rare cases where the mutation failed to fix before 10,000 generations, the simulations were run again from the beginning until fixation occurred. 500 replicates for each parameter set were run.

To estimate diversity at the neutral sites, we retrieved tree sequences and the generation when fixation occurred from the simulation output [40]. Following the most recent recommendations of SLiM [4], we used tskit (https://github.com/tskit-dev/tskit), pyslim (https://github.com/tskit-dev/pyslim), and msprime [41] Python (version 3.8.13) packages to get fully coalesced population trees (the recapitulation step) and to remove extinct lineages (the simplification step). We then added neutral mutations to the resulting tree (the overlay step) with a neutral mutation rate per site per generation of $\mu = 1 \times 10^{-8}$.

Summary statistics

For each simulation, we sampled 100 individuals from the population at the time of fixation of the beneficial allele, and computed the average nucleotide site diversity at neutral sites for the sampled individuals [42] using tskit (https://github.com/tskit-dev/tskit) Python package. We used the site mode to estimate the mean pairwise genetic diversity for 50 windows along the first chromosome and for the whole second chromosome. We computed the mean and standard deviation of π for each window between the 500 replicates. We quantified the reduction in genetic diversity for a given parameter set by dividing π by π_0, where π_0 is the baseline neutral diversity in the absence of sweeps, estimated from the mean π for chromosome 2 when $\alpha = 1$.

Mathematical modeling

An analytical model with an arbitrary value of α is hard to construct, due to the complex dynamics when there is more than one meiosis during a sweep. However, if the product of α and the duration of a sweep is sufficiently low that at most a single meiotic event occurs during a sweep, a relatively simple formula can be obtained for the expected neutral diversity at a site which recombines with the selected site with frequency r per meiosis – see Equation (S1.10) of S1 Appendix. The simulation method assumed a complete lack of crossover interference, so that r in the simulations is related to the map distance d between the neutral and selected sites by Haldane's mapping function [43]:

$$r = \frac{1}{2}[1 - exp(-2d)] \qquad (3)$$

where d is equal to the distance from the sweep in base pairs multiplied by the recombination rate per site per sexual generation, given by $\rho_\alpha = \rho/\alpha$.

Statistical tests

Differences in mean π values on the second (neutral) chromosome between two different α

values were tested by first checking that they follow a normal distribution (Shapiro test), either using a Student's *t*-test (if Gaussian) or Mann-Whitney *U* test (if not). We compared the average time to fixation of hard selective sweeps between two conditions by the same method. When comparing more than two cases at once, we performed Kruskal-Wallis tests followed by a Dunn's post hoc test [44] (these cases were never all Gaussian). All these statistics were analyzed using R (version 4.3.2) [45].

Figs

Plots were made using the following packages from R (version 4.3.2) [45]: cowplot [46], ggplot2 [47], readr [48], tibble [49], dplyr [50] and RColorBrewer [51]. The colors should be color-blind-friendly.

Cluster computations

Simulations were run on a SLURM-based (https://github.com/SchedMD/slurm) HPC cluster at IFB (https://www.france-bioinformatique.fr/cluster-ifb-core/) within the software environment management system conda (https://github.com/conda) for reproducibility. Scripts are available here: https://github.com/Louis-XIV-bis/meiotic-freq-simulation.

Availability and Implementation

Scripts and environment are available here:

https://github.com/Louis-XIV-bis/meiotic-freq-simulation

Acknowledgments

The simulations were performed on the Core Cluster of the Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013). We would like to thank George Marchment for correcting the English as well as Florian Massip, Laurent Excoffier, Guillaume Achaz, Gilles

Fischer, Alexander Palazzo and Nina Vittorelli for helpful discussions about the project.

Author Contributions

LO : Formal Analysis (simulations, statistical tests), Software, Visualization, Writing. BC: Methodology (mathematical model), Writing. FP: Conceptualization, Funding Acquisition, Methodology, Writing and Supervision.

References

[1] Johri P, Aquadro CF, Beaumont M, Charlesworth B, Excoffier L, Eyre-Walker A, et al. Recommendations for improving statistical inference in population genomics. PLOS Biol 2022;20:e3001669. https://doi.org/10.1371/journal.pbio.3001669.

[2] Kingman JFC. The coalescent. Stoch Process Their Appl 1982;13:235–48. https://doi.org/10.1016/0304-4149(82)90011-4.

[3] Wakeley JH. Coalescent theory : an introduction. Roberts & Co. Publishers; 2009.

[4] Haller BC, Messer PW. SLiM 4: Multispecies Eco-Evolutionary Modeling. Am Nat 2023;201:E127–39. https://doi.org/10.1086/723601.

[5] Hartfield M, Glémin S. Polygenic selection to a changing optimum under self–fertilisation. PLOS Genet 2024;20:e1011312. https://doi.org/10.1371/journal.pgen.1011312.

[6] Reid BN, Star B, Pinsky ML. Detecting parallel polygenic adaptation to novel evolutionary pressure in wild populations: a case study in Atlantic cod (Gadus morhua). Philos Trans R Soc B Biol Sci 2023;378:20220190. https://doi.org/10.1098/rstb.2022.0190.

[7] Marchi N, Winkelbach L, Schulz I, Brami M, Hofmanová Z, Blöcher J, et al. The genomic origins of the world's first farmers. Cell 2022;185:1842-1859.e18. https://doi.org/10.1016/j.cell.2022.04.008.

[8] Ewens WJ. Mathematical Population Genetics. vol. 27. New York, NY: Springer; 2004. https://doi.org/10.1007/978-0-387-21822-9.

[9] Charlesworth B, Charlesworth D. Elements of Evolutionary Genetics. Roberts and Company; 2010.

[10] Fischer G, Liti G. The budding yeast life cycle: More complex than anticipated? Yeast 2020. https://doi.org/10.1002/yea.3533.

[11] Tsai IJ, Bensasson D, Burt A, Koufopanou V. Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. Proc Natl Acad Sci 2008;105:4957–62. https://doi.org/10.1073/pnas.0707314105.

[12] Koufopanou V, Lomas S, Pronina O, Almeida P, Sampaio JP, Mousseau T, et al. Population Size, Sex and Purifying Selection: Comparative Genomics of Two Sister Taxa of the Wild Yeast Saccharomyces paradoxus. Genome Biol Evol 2020;12:1636–45.

https://doi.org/10.1093/gbe/evaa141.
[13] Hartfield M, Wright SI, Agrawal AF. Coalescent Times and Patterns of Genetic Diversity in Species with Facultative Sex: Effects of Gene Conversion, Population Structure, and Heterogeneity. Genetics 2016;202:297–312. https://doi.org/10.1534/genetics.115.178004.
[14] Mark Welch DB, Meselson M. Evidence for the Evolution of Bdelloid Rotifers Without Sexual Reproduction or Genetic Exchange. Science 2000;288:1211–5. https://doi.org/10.1126/science.288.5469.1211.
[15] Hartfield M. Approximating the Coalescent Under Facultative Sex. J Hered 2021;112:145–54. https://doi.org/10.1093/jhered/esaa036.
[16] Maynard J, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res 2007;89:391–403. https://doi.org/10.1017/S0016672308009579.
[17] Charlesworth B, Jensen JD. Effects of Selection at Linked Sites on Patterns of Genetic Variability. Annu Rev Ecol Evol Syst 2021;52:177–97. https://doi.org/10.1146/annurev-ecolsys-010621-044528.
[18] Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos Trans R Soc B Biol Sci 2017;372:20160455. https://doi.org/10.1098/rstb.2016.0455.
[19] Lewontin RC. The Genetic Basis of Evolutionary Change. 4th Printing edition. New York: Columbia Univ Pr; 1974.
[20] Buffalo V. Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's paradox. eLife 2021. https://doi.org/10.7554/eLife.67509.
[21] Charlesworth B, Jensen JD. How Can We Resolve Lewontin's Paradox? Genome Biol Evol 2022;14:evac096. https://doi.org/10.1093/gbe/evac096.
[22] Weidberg H, Moretto F, Spedale G, Amon A, Werven FJ van. Nutrient Control of Yeast Gametogenesis Is Mediated by TORC1, PKA and Energy Availability. PLOS Genet 2016;12:e1006075. https://doi.org/10.1371/journal.pgen.1006075.
[23] Zhu YO, Siegal ML, Hall DW, Petrov DA. Precise estimates of mutation rate and spectrum in yeast. Proc Natl Acad Sci U S A 2014;111:E2310–8. https://doi.org/10.1073/pnas.1323011111.
[24] Duan S-F, Han P-J, Wang Q-M, Liu W-Q, Shi J-Y, Li K, et al. The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. Nat Commun 2018;9:2690. https://doi.org/10.1038/s41467-018-05106-7.
[25] Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, et al. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 2018;556:339–44. https://doi.org/10.1038/s41586-018-0030-5.
[26] Butlin R. The costs and benefits of sex: new insights from old asexual lineages. Nat Rev Genet 2002;3:311–7. https://doi.org/10.1038/nrg749.

[27] Charlesworth B, Harvey PH, Barton NH. Genetic hitchhiking. Philos Trans R Soc Lond B Biol Sci 2000;355:1553–62. https://doi.org/10.1098/rstb.2000.0716.

[28] Hermisson J, Pennings PS. Soft Sweeps: Molecular Population Genetics of Adaptation From Standing Genetic Variation. Genetics 2005;169:2335–52. https://doi.org/10.1534/genetics.104.036947.

[29] Charlesworth B. How Good Are Predictions of the Effects of Selective Sweeps on Levels of Neutral Diversity? Genetics 2020;216:1217–38. https://doi.org/10.1534/genetics.120.303734.

[30] Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature 2008;454:479–85. https://doi.org/10.1038/nature07135.

[31] Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. Science 1996;274:546, 563–7. https://doi.org/10.1126/science.274.5287.546.

[32] Zeyl CW, Otto SP. A short history of recombination in yeast. Trends Ecol Evol 2007;22:223–5. https://doi.org/10.1016/j.tree.2007.02.005.

[33] Wloch DM, Szafraniec K, Borts RH, Korona R. Direct Estimate of the Mutation Rate and the Distribution of Fitness Effects in the Yeast Saccharomyces cerevisiae. Genetics 2001;159:441–52. https://doi.org/10.1093/genetics/159.2.441.

[34] Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics 1993;134:1289–303. https://doi.org/10.1093/genetics/134.4.1289.

[35] Gilbert KJ, Pouyet F, Excoffier L, Peischl S. Transition from Background Selection to Associative Overdominance Promotes Diversity in Regions of Low Recombination. Curr Biol CB 2020;30:101-107.e3. https://doi.org/10.1016/j.cub.2019.11.063.

[36] Becher H, Jackson BC, Charlesworth B. Patterns of Genetic Variability in Genomic Regions with Low Rates of Recombination. Curr Biol 2020;30:94-100.e3. https://doi.org/10.1016/j.cub.2019.10.047.

[37] Sui Y, Qi L, Wu J-K, Wen X-P, Tang X-X, Ma Z-J, et al. Genome-wide mapping of spontaneous genetic alterations in diploid yeast cells. Proc Natl Acad Sci U S A 2020;117:28191–200. https://doi.org/10.1073/pnas.2018633117.

[38] Cury J, Haller BC, Achaz G, Jay F. Simulation of bacterial populations with SLiM. Peer Community J 2022;2. https://doi.org/10.24072/pcjournal.72.

[39] Cubillos FA, Vásquez C, Faugeron S, Ganga A, Martínez C. Self-fertilization is the main sexual reproduction mechanism in native wine yeast populations. FEMS Microbiol Ecol 2009;67:162–70. https://doi.org/10.1111/j.1574-6941.2008.00600.x.

[40] Haller BC, Galloway J, Kelleher J, Messer PW, Ralph PL. Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. Mol Ecol Resour 2019;19:552–66. https://doi.org/10.1111/1755-0998.12968.

[41] Baumdicker F, Bisschop G, Goldstein D, Gower G, Ragsdale AP, Tsambos G, et al. Efficient ancestry and mutation simulation with msprime 1.0. Genetics 2022;220:iyab229. https://doi.org/10.1093/genetics/iyab229.

[42] Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 1979;76:5269–73. https://doi.org/10.1073/pnas.76.10.5269.

[43] Haldane JBS. The Combination of Linkage Values, and The Calculation of Distances Between The Loci of Linked Factors. J Genet 1919;8:299–309.

[44] Dinno A. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums 2017.

[45] Team RC. R: A Language and Environment for Statistical Computing 2021.

[46] Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2" 2024.

[47] Hadley W. ggplot2: Elegant Graphics for Data Analysis 2016. https://ggplot2.tidyverse.org/ (accessed March 6, 2024).

[48] Wickham H, Hester J, Francois R, Bryan J, Bearrows S, Software P, et al. readr: Read Rectangular Text Data 2024.

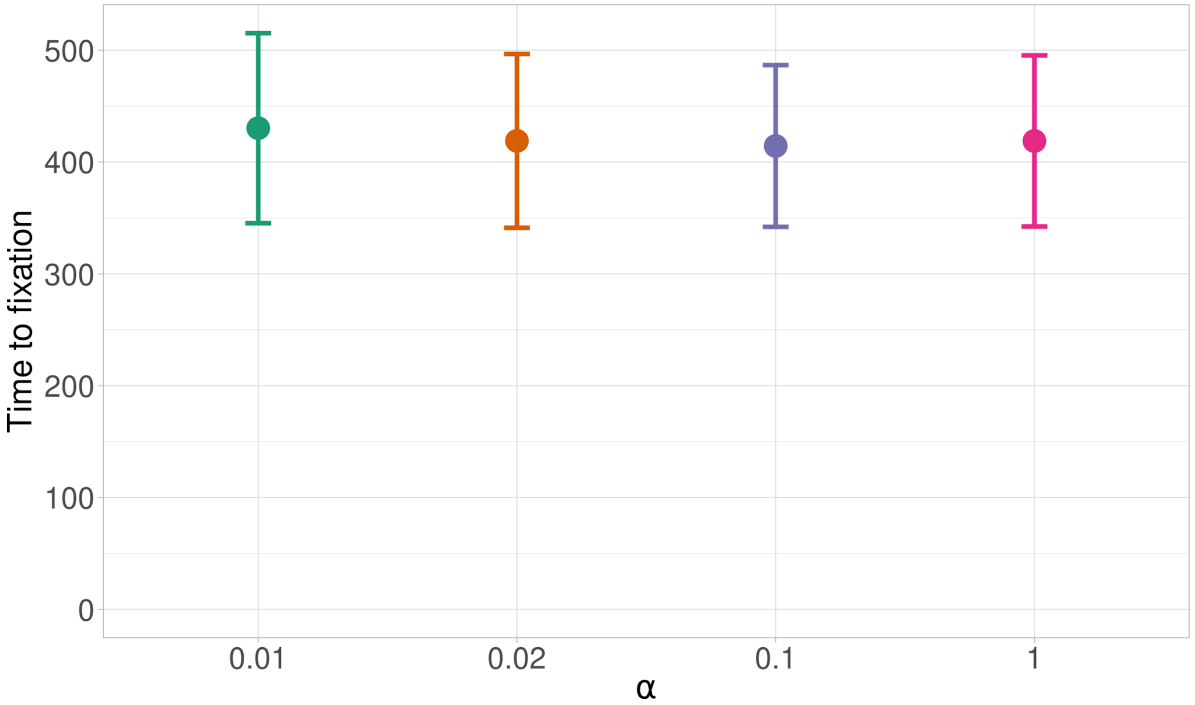
[49] Müller K, Wickham H, Francois R, Bryan J, RStudio. tibble: Simple Data Frames 2023.

[50] Wickham H, François R, Henry L, Müller K, Vaughan D, Software P, et al. dplyr: A Grammar of Data Manipulation 2023.

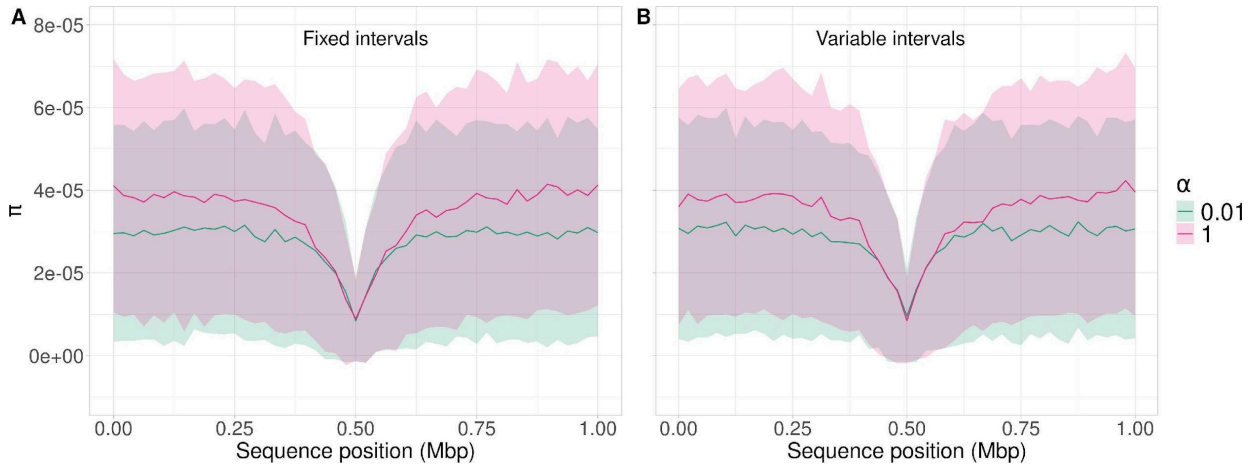
[51] Neuwirth E. RColorBrewer: ColorBrewer Palettes 2022.

[52] Charlesworth B. How Long Does It Take to Fix a Favorable Mutation, and Why Should We Care? Am Nat 2020;195:753–71. https://doi.org/10.1086/708187.

Supporting information

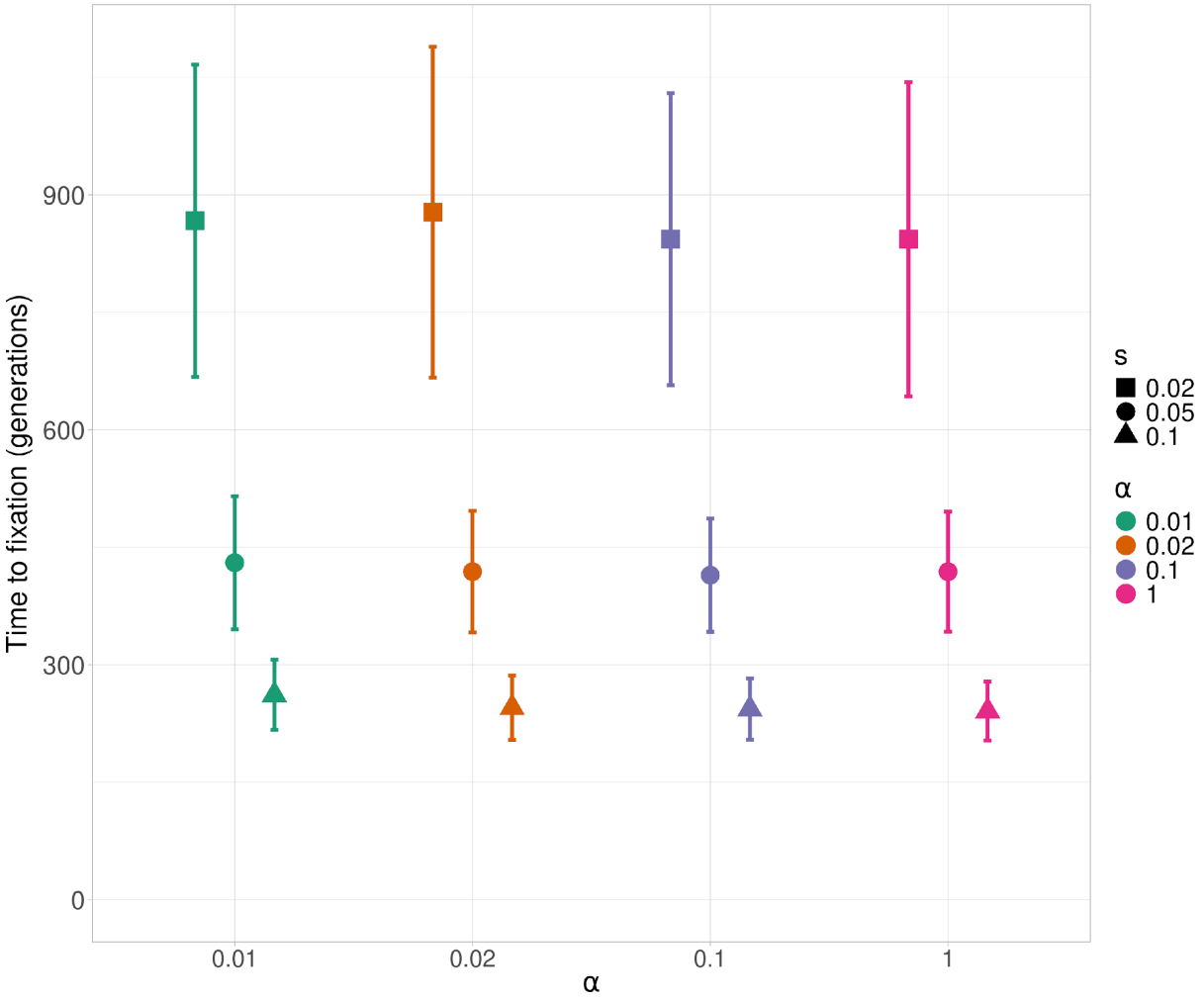


S1 Fig. Times to fixation of the beneficial mutation for different *α* values.
Each case used the following settings: a beneficial mutation is introduced into one individual and the population was allowed to evolve until its fixation, with $\rho = 5\times10^{-8}$, $s = 0.05$ and $h = 0.5$. If the beneficial mutation was lost, the simulation was rerun. Dots represent the average time to fixation from 500 simulations, with $\alpha = 0.01$ (green), $\alpha = 0.02$ (orange), $\alpha = 0.1$ (blue), and $\alpha = 1$ (pink). The error bars are standard errors (see Material and Methods for details).



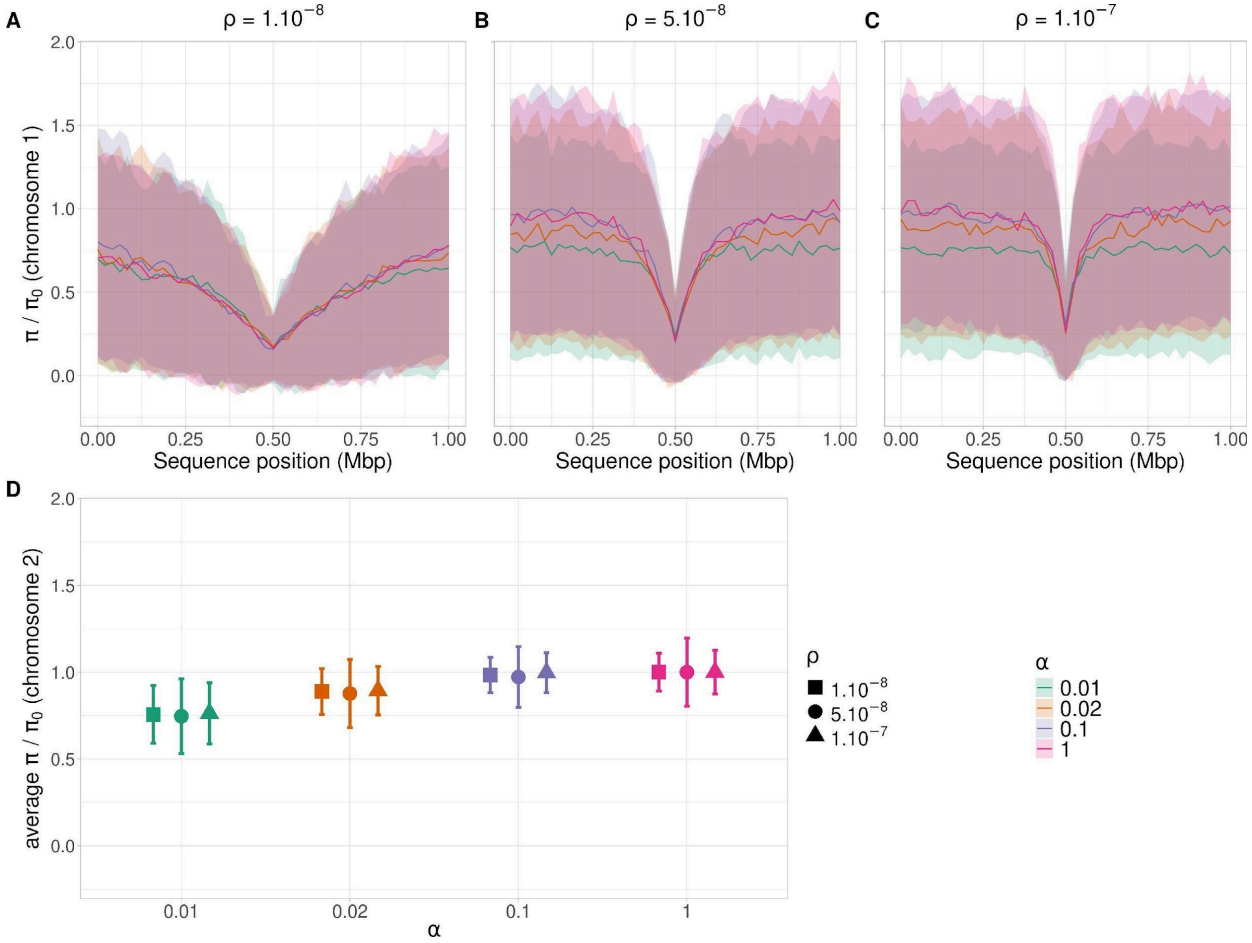
S2 Fig. The effects of the two models of *α* with hard selective sweeps on genetic diversity (*π*) along a 1 Mb chromosome.

The same procedure for introducing a beneficial mutation as in S1 Fig was used, with h = 0.5 and s = 0.05. Colors and lines are the same as in Fig 1. With α = 0.01, the interval between two sexual generations was fixed at $1/\alpha$ = 100 generations (A) or sampled from a Gaussian distribution centered around $1/\alpha$ = 100 and a standard deviation of 10 (B)



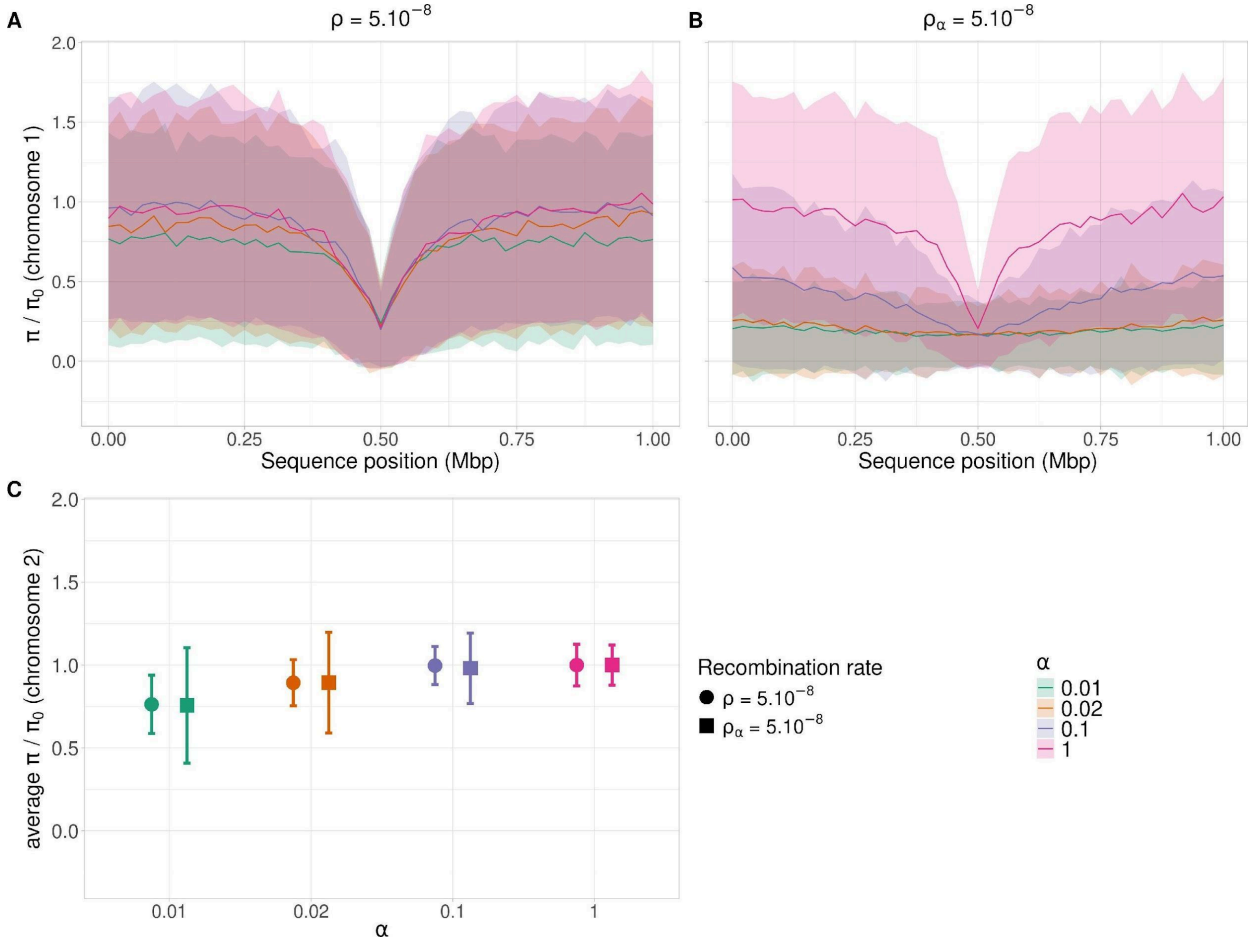
S3 Fig. Times to fixation of the beneficial mutations under different strengths of selection.

The same procedure for introducing a beneficial mutation as in the other figures was used, with $\rho = 5 \times 10^{-8}$, $h = 0.5$ and a given s. Dots represent the average time to fixation from 500 simulations with $\alpha = 0.01$ (green), $\alpha = 0.02$ (orange), $\alpha = 0.1$ (blue), and $\alpha = 1$ (pink) and error bars are standard errors (see Material and Methods for details).



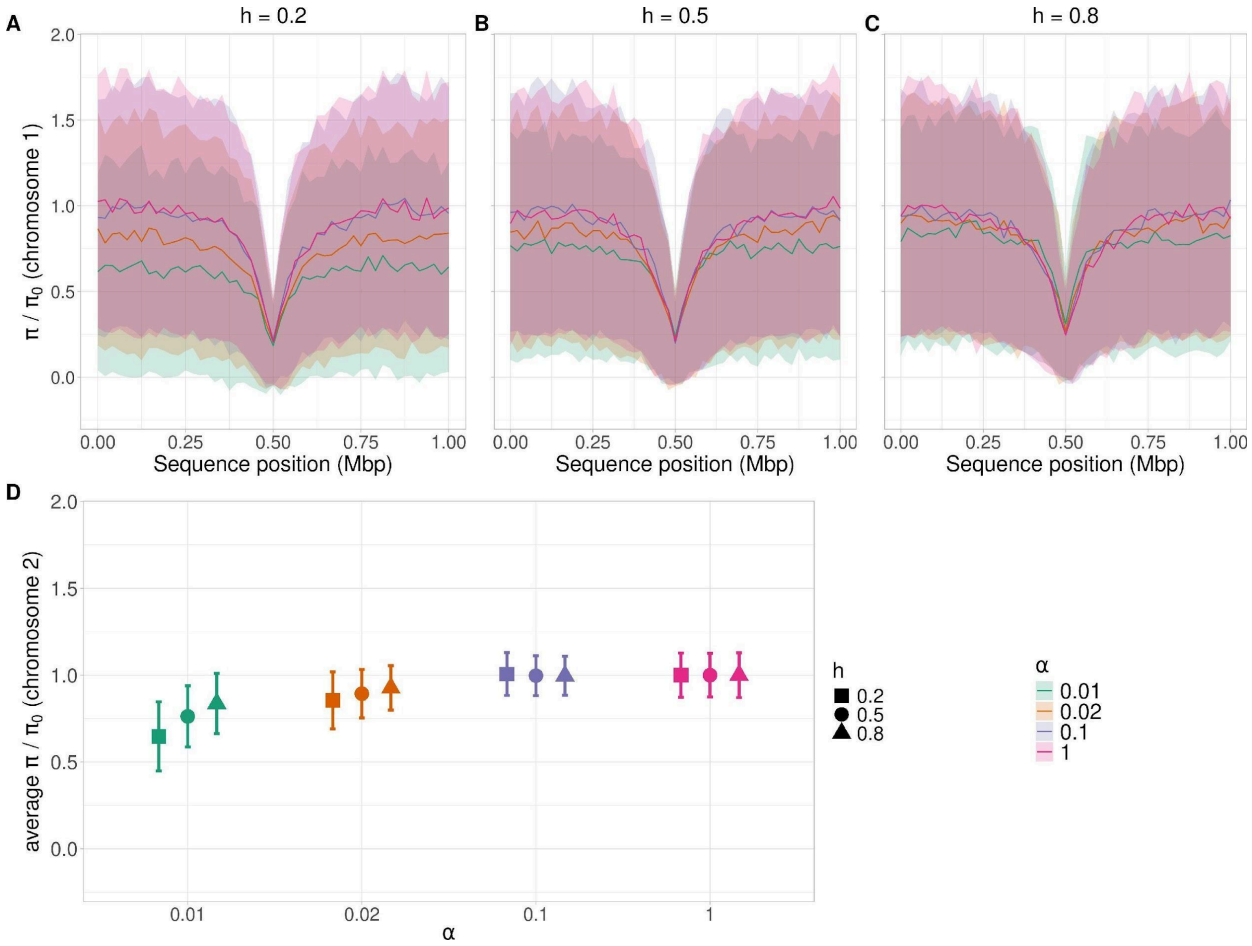
S4 Fig. The effect of the recombination rate per basepair on π/π_0, with s = 0.05 and h = 0.5.

(A-C) π/π_0 along a 1Mb chromosome with a hard sweep at 0.5 Mbp. (D) Average π/π_0 and its standard error for a second chromosome that is unlinked to the first one and which carries only neutral mutations. For $\alpha < 1$, the interval between two sexual generations was drawn from a normal distribution centered on $1/\alpha$ generations (see Material and Methods for details). The same procedure for introducing a beneficial mutation as in the other figures was used. The three cases have different recombination rates per generation: $\rho = 1 \times 10^{-8}$ (A), $\rho = 5 \times 10^{-8}$ (B), and $\rho = 1 \times 10^{-7}$ (C). Colored lines are the same as in Fig 2.



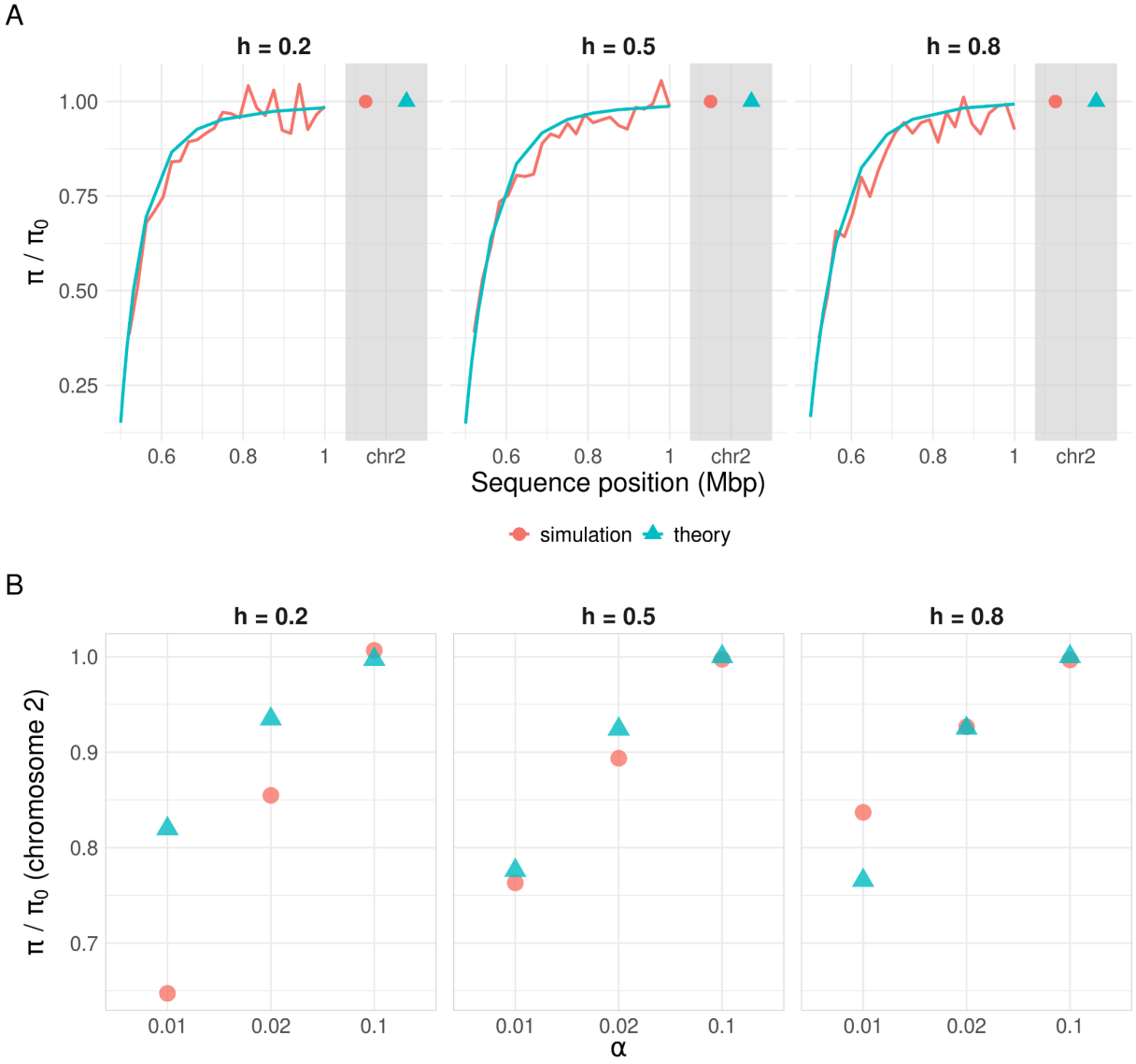
S5 Fig. Effect of ρ versus ρ_α on π/π_0.

(A-B): π/π_0 along a 1Mb chromosome with a selective sweep at 0.5 Mbp, with h = 0.5, s = 0.05. (A) Recombination rate per site per generation $\rho = \alpha\, \rho_\alpha$ is constant at 5×10^{-8} whereas $\rho_\alpha = 5 \times 10^{-6}$ for $\alpha = 0.01$, $\rho_\alpha = 2.5 \times 10^{-6}$ for $\alpha = 0.02$, $\rho_\alpha = 5 \times 10^{-7}$ for $\alpha = 0.1$ and $\rho_\alpha = 5 \times 10^{-8}$ for $\alpha = 1$. (B) Recombination rate per site per meiosis ρ_α is constant at 5×10^{-8} whereas $\rho = 5 \times 10^{-10}$ for $\alpha = 0.01$, $\rho = 1 \times 10^{-9}$ for $\alpha = 0.02$, $\rho = 5 \times 10^{-9}$ for $\alpha = 0.1$ and $\rho = 5 \times 10^{-8}$ for $\alpha = 1$. (C) Average π/π_0 and its standard error for a second, independent chromosome carrying only neutral mutations when ρ is constant (circles) or ρ_α is constant (squares). For $\alpha < 1$, the interval between two sexual generations was drawn from a normal distribution centered on $1/\alpha$ generations (see Material and Methods for details). Colored lines are the same as in Fig 2. The same procedure for introducing a beneficial mutation as in other figures was used,



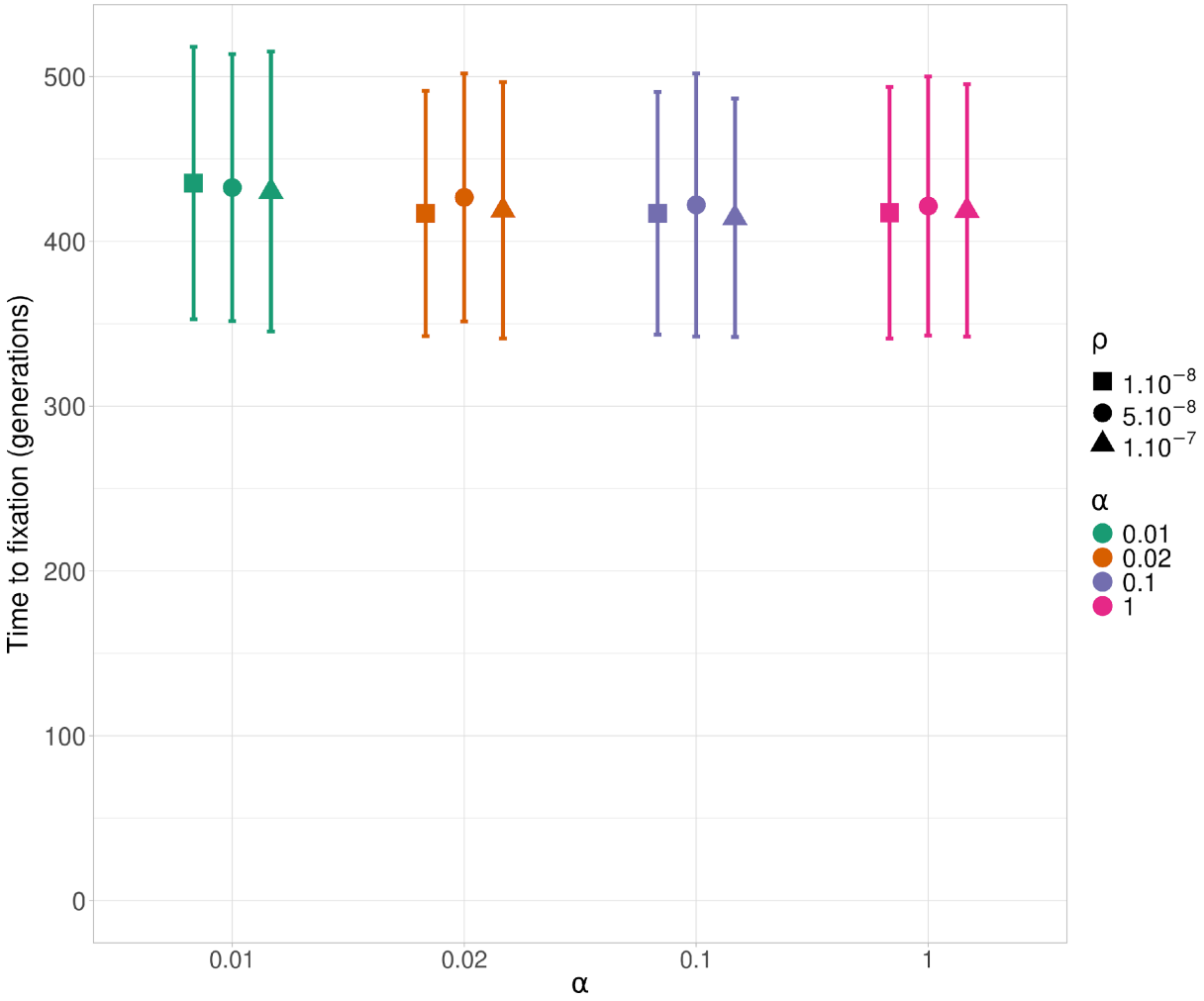
S6 Fig. Effects of the dominance coefficient *h*.

(A)-(C) Values of π/π_0 along a 1Mb chromosome that carries a beneficial mutation (s = 0.05) at 0.5 Mbp with h = 0.2 (panel A), 0.5 (B), and 0.8 (C) and $\rho = 5 \times 10^{-8}$. (D) The mean value of π/π_0 (with standard error) on a second, independent chromosome. The same procedure for introducing a beneficial mutation as in other figures was used. For $\alpha < 1$, the interval between two sexual generations was drawn from a normal distribution centered on $1/\alpha$ generations (see Material and Methods for details). Colors and lines are the same as in Fig 1.



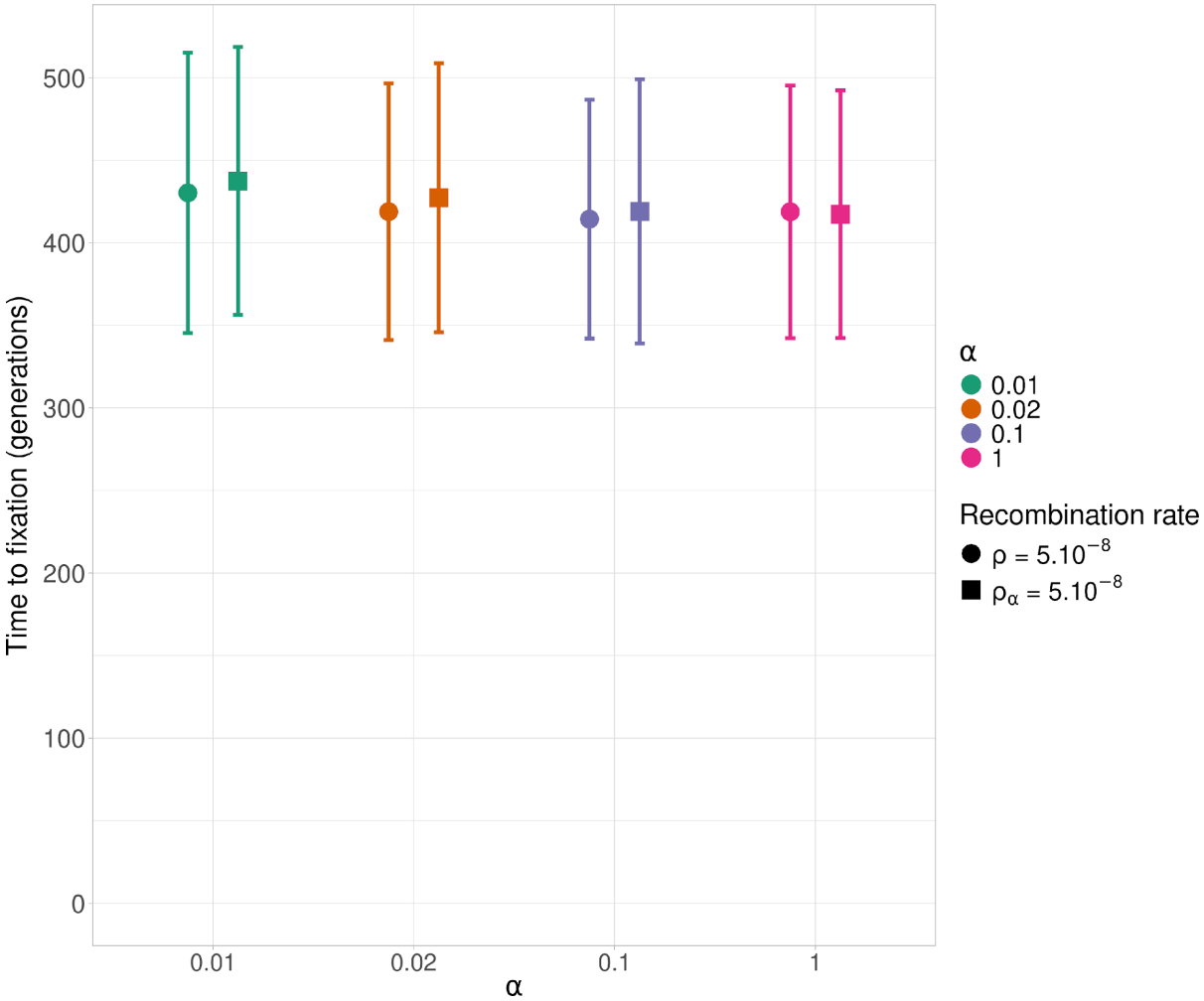
S7 Fig. The behavior of π/π_0 with different dominance coefficients (*h*) and frequencies of meiosis between theoretical model (Equations 14 from [29], in blue and simulation results [from Fig 2] in red).

π/π_0 was computed in the same way as in Fig 2 (see Methods). (A) Comparisons for chromosome 1 (linked) and 2 (unlinked) for α = 1. (B) Comparisons for chromosome 2, with a combination of α and *h*. Values for simulated data (red) are the averages from 500 simulations. The same procedure for introducing a beneficial mutation as in other figures was used.



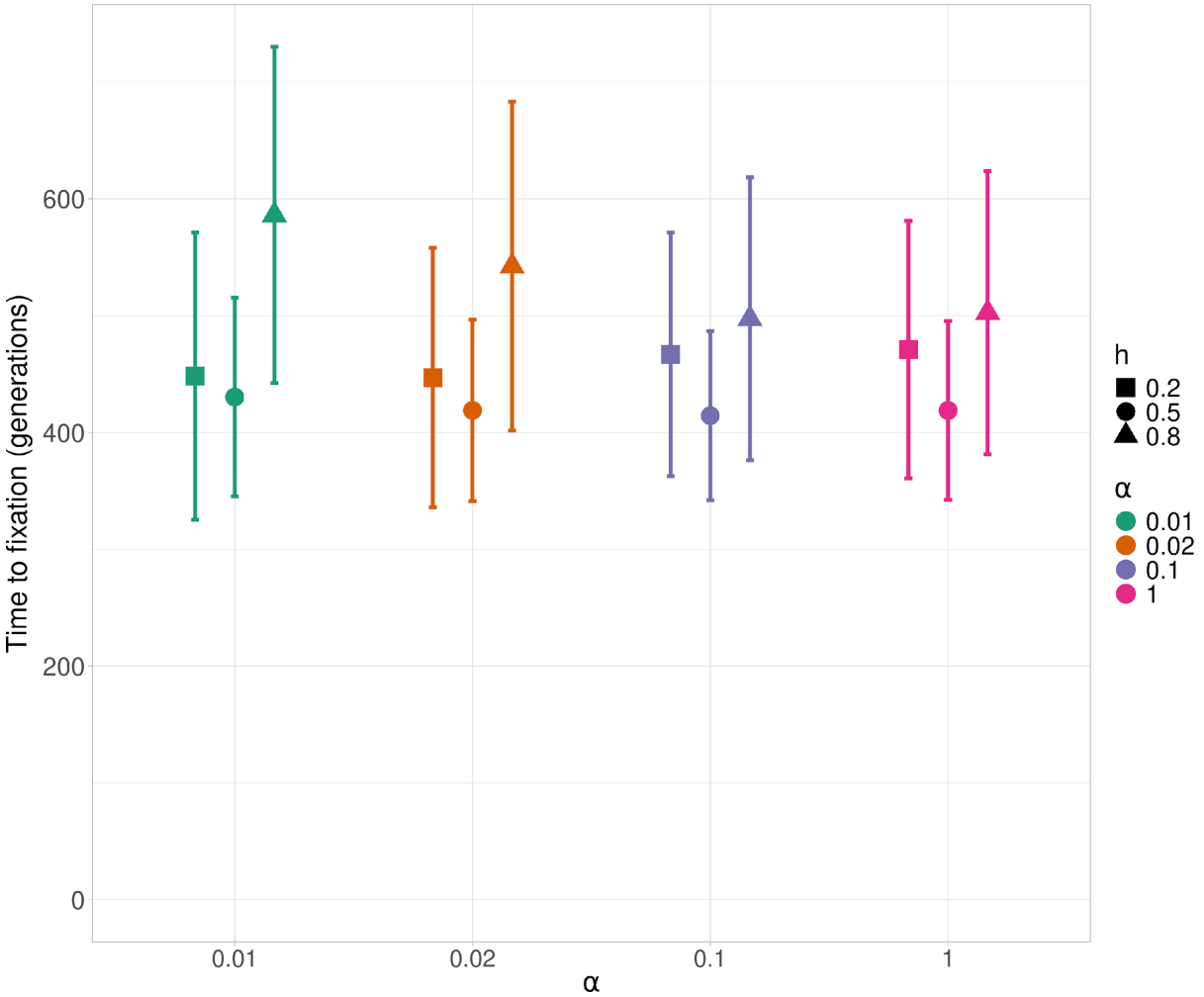
S8 Fig. Times to fixation of beneficial mutations under different recombination rates.

The same procedure for introducing a beneficial mutation as in other figures was used, with h = 0.5, s = 0.05 and a given value of ρ. Dots represent the average time to fixation from 500 simulations time with α = 0.01 (green), α = 0.02 (orange), α = 0.1 (blue), and α = 1 (pink), and error bars are standard deviation (see. Material and Methods for details).



S9 Fig. Effect of ρ versus ρ_α on the time to fixation of a beneficial mutation.

The same procedure for introducing a beneficial mutation as in other figures was used, with h = 0.5, s = 0.05. Either ρ was fixed (circles) or ρ_α was fixed (squares). If recombination rate per site per generation $\rho = \alpha\ \rho_\alpha$ is fixed at 5×10^{-8} , we have $\rho_\alpha = 5 \times 10^{-6}$ for $\alpha = 0.01$, $\rho_\alpha = 2.5 \times 10^{-6}$ for $\alpha = 0.02$, $\rho_\alpha = 5 \times 10^{-7}$ for $\alpha = 0.1$ and $\rho_\alpha = 5 \times 10^{-8}$ for $\alpha = 1$. Dots represent the average time to fixation from 500 simulations time with $\alpha = 0.01$ (green), $\alpha = 0.02$ (orange), $\alpha = 0.1$ (blue), and $\alpha = 1$ (pink), and error bars are standard errors (see. Material and Methods for details).



S10 Fig. Times to fixation of a beneficial mutation with different dominance coefficients. The same procedure for introducing a beneficial mutation as in other figures was used, with $\rho = 5 \times 10^{-8}$, $s = 0.05$ and a given h. Dots represent the average time to fixation from 500 simulations time with $\alpha = 0.01$ (green), $\alpha = 0.02$ (orange), $\alpha = 0.1$ (blue), and $\alpha = 1$ (pink); error bars are standard errors (see. Material and Methods for details).

Supplementary Text 1

In canonical models with α = 1, the recombination rate per nucleotide site per generation (ρ) relative to the strength of selection controls the width of the valley of diversity around the location of the beneficial mutation. We wanted to disentangle the roles played by ρ_α, α and ρ on genetic diversity, for a given selection regime. We expected to find that ρ controls the width of the valley regardless of α, and we wanted to explore the impact of ρ_α and α on π/π_0 measured on the neutral chromosome. Panels A-C in S4 A-C show that, the smaller ρ, the wider the valley around the location of the sweeping mutation on the chromosome, for all α values. In Panel A, the whole chromosome is affected by selection, as recombination is too infrequent to counteract its effect. Neutral loci are in linkage disequilibrium with the beneficial allele until fixation, and genetic diversity does not reach a plateau because recombination events are too rare. In contrast, with a high rate of recombination, diversity reaches a plateau close to its purely neutral value, and the valley is narrower. As expected, diversity on the independent, neutral chromosome (Panel D) is not affected by ρ (no significant difference between each ρ for each α, Kruskal-Wallis tests). As is also expected, the time to fixation of the sweep is not affected by ρ, taking an average of 425 generations to fix ($p > 0.05$ for all α values, Kruskall-Wallis tests, S8 Fig).

Conversely, for a given ρ, π/π_0 on the neutral chromosome decreases with α ($p < 0.05$ for all pairwise α comparisons, except between α = 0.1 and α = 1 for all ρ, Dunn's post hoc test). Indeed, the frequency of recombination events between the swept and neutral chromosomes is solely determined by the frequency of sexual reproduction (*i.e*, by α and not by ρ) (see S1 Appendix). These results show that α and s determine the genetic diversity on neutral chromosomes, while ρ affects the width but not the depth of the valley. This is in accord with the predictions of the standard sweep model, given by Equations 14 of [29]

Recombination rate per meiosis determines the width of the diversity valley

We assumed that the effects of sweeps would be controlled by the product ρ of the

recombination rate per sexual generation ρ_α and α, which measures the recombination rate per generation. We assessed whether this rescaling assumption affects our results and compared a constant ρ (panel A, same as before) to a constant ρ_α (panel B) in S5 Fig. As expected, π behaves completely differently when ρ_α is not scaled with $\alpha < 1$ (Panel B). Indeed, with $\rho_\alpha = 5 \times 10^{-8}$ and $\alpha < 1$, recombination happens only very rarely. For instance, with $\alpha = 0.01$, ρ equals 5×10^{-10} on average per site per generation. With a genome of 2Mbp and approximately 420 generations for the sweep duration (S9 Fig), we have a mean of 0.4 recombination events per individual per simulation ($400 \times 2 \times 10^6 \times 5 \times 10^{-10}$). The rare recombination events that happen do not effectively break genetic linkage between the beneficial mutation and neutral mutations. This has a strong effect so that, when $\alpha = 0.01$, π/π_0 is ~ 0.2 over the whole of chromosome 1 (Panel B of S5 Fig). As expected given the results from S4 Fig, for a given α, π/π_0 on the neutral chromosome is the same on average between the two conditions but the variance of the average π/π_0 is larger when ρ_α is fixed (Panel C of S5 Fig). Once again, the patterns for the mean of π/π_0 are consistent with standard sweep theory.

The dominance coefficient slightly affects the loss of genome-wide genetic diversity

Next, we explore the effect of the dominance coefficient h and α on genetic diversity. So far, we have considered semi-dominant beneficial mutations with $h = 0.5$. We compared π/π_0 for simulations with either partially recessive $h = 0.2$ or partially dominant $h = 0.8$ mutations, see S6 Fig. With $\alpha = 1$ and 0.1, h has no effect on genetic diversity at the plateau nor on the width of the valley (pink line, all panels), in agreement with standard sweep theory [29]. Overall, the genetic diversity (π/π_0) on the neutral chromosome ranges from 0.977 to 1.0 for all h and both $\alpha \geq 0.1$. With $\alpha < 0.1$, the genetic diversity on the second chromosome increases with h (Panel D of S6 Fig). These differences are significant for all three h values with all p-values lower than 8×10^{-4} (Dunn's post hoc test). We also compared the simulation results with the theoretical predictions of the standard sweep model (as in Fig 3) for different

values of h (S7 Fig). The fit remains good for $\alpha = 1$ and $\alpha = 0.1$ for all h values. For $h = 0.2$ and $\alpha < 0.1$, the standard model tends to overestimate the diversity on the unlinked chromosome for the lowest α values. The opposite is observed for $h = 0.8$ and $\alpha < 0.1$, with the standard model underestimating the diversity on the neutral chromosome with small α.

The effects of α and h on the fixation times of beneficial mutations are rather different (S10 Fig). For each h, the fixation time is the same between $\alpha = 1$ and $\alpha = 0.1$ ($p > 0.20$, Dunn's post hoc test). If $h = 0.8$, the difference to fix is significant when $\alpha < 0.1$. If $h = 0.5$, the difference is significant only when $\alpha = 0.01$ compared to other α values. If $h = 0.2$, the difference is significant between all the pairwise comparisons ($p < 6 \times 10^{-4}$, Dunn's post hoc test, except between $\alpha = 0.01$ and 0.02 as well as $\alpha = 0.1$ and 1).

For each α, the fixation time is larger for $h = 0.8$ than for the other two values ($p < 1\times10^{-4}$, Dunn's post hoc test), consistent with the slow approach to fixation of dominant mutations in randomly mating populations, due to the sheltering from selection of homozygotes for the recessive alternative allele [52]. The fixation time is larger for $h = 0.2$ than for $h = 0.5$ ($p < 1 \times 10^{-4}$, Dunn's post hoc test, except for $\alpha = 0.01$ where $p = 0.0509$) , even though this difference is smaller than to the difference between $h = 0.8$ and 0.5 (< 4% between $h = 0.02$ and $h = 0.05$). These effects of h are predicted by the theory for randomly mating populations (Table 1 of); as in that case, the effects are all small.