

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17

**Coalescence with background and balancing selection in systems with bi- and uniparental reproduction: contrasting partial asexuality and selfing**

Aneil F. Agrawal<sup>a, 1</sup>, and Matthew Hartfield<sup>a,b</sup>

a - Department of Ecology & Evolutionary Biology, University of Toronto

b - Bioinformatics Research Centre, University of Aarhus, 8000C Aarhus, Denmark

<sup>1</sup>email: [a.agrawal@utoronto.ca](mailto:a.agrawal@utoronto.ca)

key words: coalescence, asexual reproduction, self-fertilization, effective population size, background selection, balancing selection

**Abstract:** Uniparental reproduction in diploids, via asexual reproduction or selfing, reduces the independence with which separate loci are transmitted across generations. This is expected to increase the extent to which a neutral marker is affected by selection elsewhere in the genome. Such effects have previously been quantified in coalescent models involving selfing. Here we examine the effects of background selection and balancing selection in diploids capable of both sexual and asexual reproduction (i.e., partial asexuality). We find that the effect of background selection on reducing coalescent time (and effective population size) can be orders of magnitude greater when rates of sex are low than when sex is common. This is because asexuality enhances the effects of background selection through both a recombination effect and a segregation effect. We show that there are several reasons that the strength of background selection differs between systems with partial asexuality and those with comparable levels of uniparental reproduction via selfing. Expectations for reductions in  $N_e$  via background selection have been verified using stochastic simulations. In contrast to background selection, balancing selection increases the coalescent time for a linked neutral site. With partial asexuality, the effect of balancing selection is somewhat dependent upon the mode of selection (e.g., heterozygote advantage vs. negative frequency dependent selection) in a manner that does not apply to selfing. This is because the frequency of heterozygotes, which are required for recombination onto alternative genetic backgrounds, is more dependent on the pattern of selection with partial asexuality than with selfing.

## Introduction

Coalescence theory forms the basis for much of the modern evolutionary inference made from genomic data. Although the coalescence of neutral alleles is simple in ideal populations, the process is altered by a myriad of biological realities such as population structure, changes in population size, and non-Poisson variance in reproductive success. Existing theory has tackled these and many other important deviations from the simplest case (WAKELEY 2008). In addition to ecological factors such as those above, selection at one site in the genome can profoundly influence the coalescence times of linked neutral sites, thereby affecting patterns of genomic diversity. Although selective sweeps can dramatically reduce coalescence times (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989), here we will focus on two other forms of selection: background and balancing selection.

Background selection refers to selection on deleterious alleles occurring in the genetic background of a focal neutral allele, and its resulting consequences for coalescence (reviewed in CHARLESWORTH 2012). Assuming selection is strong relative to drift, deleterious alleles are doomed to extinction and linked neutral sites will share the same fate unless they can escape via recombination (CHARLESWORTH *et al.* 1993). Looking backwards in time, this means that extant copies of a neutral site must trace their ancestry either avoiding deleterious backgrounds or recombining onto them only in the relatively brief period prior to their selective elimination. Thus, there are fewer potential ancestors for a neutral site than the actual number of individuals present in the preceding generations. In other words, background selection reduces the effective population size, shortening coalescent times. As this

logic suggests, the consequences of background selection are much stronger if there is tight linkage between the focal neutral site and the locus experiencing deleterious mutations (HUDSON and KAPLAN 1995; NORDBERG *et al.* 1996a). Because of heterogeneity across the genome in the local deleterious mutation rate (e.g., due to variation in gene density and recombination rate), the strength of background selection is expected to vary across the genome. This is believed to be a major source of the variation in sequence diversity across chromosomes in species such as *Drosophila melanogaster* (CHARLESWORTH and CAMPOS 2014; ELYASHIV *et al.* 2014) and humans (McVICKER *et al.* 2009).

Balancing selection preserves variation at a selected site, allowing two selected alleles to persist much longer than expected under drift alone. This means that coalescent times of the alternative selected alleles will be much longer than expected under neutrality. The same logic applies to two copies of a neutral site linked to alternative alleles at the selected site. Coalescence of more loosely linked neutral sites is less affected by the balancing selection that prevents coalescence at a selected site (HUDSON and KAPLAN 1988; KAPLAN *et al.* 1988; HEY 1991). By increasing coalescent times, balancing selection is expected to result in increased levels of neutral diversity at sites closely linked to targets of selection. Consistent with this, elevated neutral diversity is observed around loci strongly expected to be under balancing selection such as self-incompatibility alleles (RICHMAN *et al.* 1996; KAMAU and CHARLESWORTH 2005) and major histocompatibility loci (HUGHES and NEI 1988). In other cases, elevated levels of diversity have been used to identify potential candidates of balancing selection (ANDRÉS *et al.* 2009).

Most coalescent models implicitly assume that organisms are either haploid or diploid but obligately sexual and randomly mating. However, some of the most reproductively prolific diploids reproduce uniparentally much of the time via selfing or asexual reproduction (e.g., duckweeds, monogonant rotifers, ostracods, *Daphnia*, many grasses and fungi). Asexuality and selfing involve very different types of inheritance and, consequently, their “direct” effects on coalescent times are in opposite directions; asexuality increases coalescence time whereas selfing reduces it (see below). However, both forms of uniparental reproduction reduce the independence with which two alleles at separate loci within the same individual trace their ancestry backwards in time; that is uniparental reproduction increases the effective degree of linkage. As a consequence, uniparental inheritance is expected to magnify the importance of linked selection on coalescence.

Previous work has examined background and balancing selection for organisms that reproduce by both outcrossing and selfing (NORDBORG 1997). This work showed, for example, that the effect of background selection is stronger when selfing rates are high. Because effective population size can be dramatically reduced in selfers via background selection when considering genome-wide mutation, selfers suffer from concomitant reductions in the probability of fixation of beneficial mutations as well as increases in the probability of fixation of deleterious alleles (GLÉMIN and RONFORT 2012; KAMRAN-DISFANI and AGRAWAL 2014). This illustrates that studying the effect of linked selection on coalescent times of neutral sites is useful for understanding how uniparental reproduction affects adaptation via  $N_e$ , in

addition to its more obvious utility with respect to understanding levels of neutral diversity.

Here we develop coalescent models for organisms capable of both sexual and asexual reproduction. Like selfing, asexuality enhances the consequences of selection elsewhere across the genome, sometimes dramatically so. Partial asexuality and selfing are often lumped together, at least colloquially, because both are systems involving uniparental inheritance with reduced rates of genetic mixing. However, these alternative modes of uniparental reproduction heighten linked selection in somewhat different ways and we highlight the differences, though we do not present any original results with respect to selfing.

## Theory and Results

### *Direct effects*

In the absence of any linked selection, the expected coalescent time for two randomly sampled alleles in a diploid sexually outcrossing species is  $E[T] = 2N$  (WAKELEY 2008). For a species that produces a fraction  $\sigma$  of its offspring sexually and the remainder asexually, the expected coalescence time of two alleles sampled from different individuals is

$$E[T_{Sex\&Asex}] = 2N + \frac{1}{\sigma} \quad [1]$$

as shown by CEPLITIS 2003 (see also BENGTTSSON (2003) and HARTFIELD *et al.* (2015) for further discussion of direct effects of asexuality).

As shown by NORDBORG and DONNELLY (1997), for a species that produces a fraction  $o$  of its offspring via outcrossing and the remainder by selfing, the expected coalescence time of two alleles sampled from different individuals is

$$E[T_{Outcross\&Self}] = 2N \left( \frac{1+o}{2} \right) \quad [2]$$

Contrasting [1] and [2], we see the direct effects of selfing and asexuality differs in several ways (Fig. 1). Increasing the rate of uniparental reproduction (i.e., reducing  $o$  or  $\sigma$ ), reduces coalescent time in the case of selfing but has the opposite effect with partial asexuality. With selfing, coalescent time decreases linearly with declining rates of outcrossing. In contrast, the effect of partial asexuality on coalescent time is negligible unless the rate of sex is very small (i.e., the direct effect of partial asexuality can be ignored unless  $\sigma \sim O(1/N)$ ). Whereas very high selfing ( $o \rightarrow 0$ ) reduces the coalescent time by a maximum of 50%, the maximal increase in  $E[T]$  with very high asexuality ( $\sigma \rightarrow 0$ ) is unlimited in principle. However, if gene conversion is incorporated in the partial sex model, increases in coalescent time from asexuality are much more limited and can even be reversed (HARTFIELD *et al.* 2015).

We refer to the effects described above as the “direct” effects of uniparental reproduction to distinguish them from the “indirect” effects of uniparental reproduction in mediating the consequences of linked selection.

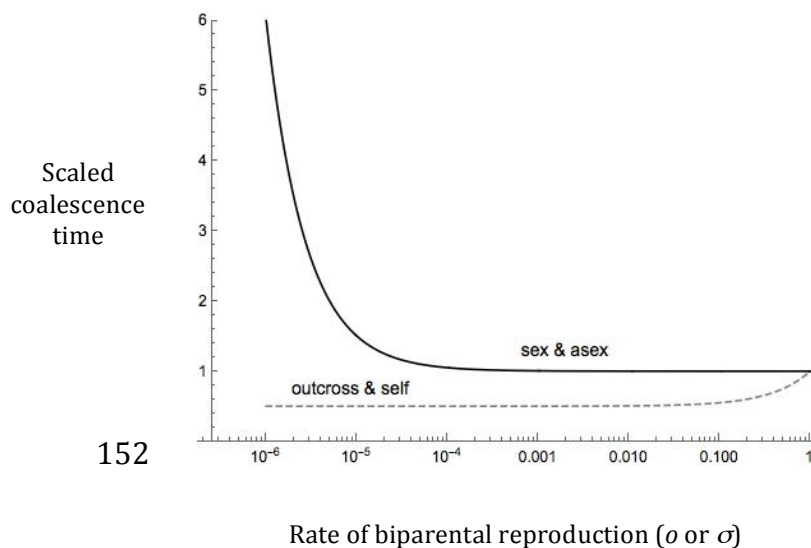


Figure 1. The direct effect of uniparental reproduction via asexuality or selfing on the scaled coalescence time ( $E[T]/2N$ ). Partial asexuality has a negligible direct effect unless the rate of sex ( $\sigma$ ) is on the order of the reciprocal of the population size; here  $N = 10^5$ . High selfing reduces  $N_e$  to 50% of  $N$ .



## Background Selection

The effects of background selection in sexual outcrossing species were first modeled in classic papers by HUDSON and KAPLAN (1994; 1995) and by NORDBORG *et al.* (1996a) using different approaches. To examine background selection in partially asexual species, we follow an approach similar to that used by NORDBORG (1997) who studied background and balancing selection in partial selfers. The main idea is that two sampled alleles can be found in a variety of different states (e.g., in the same or different individuals, in wild-type or mutant backgrounds). Based on transition probabilities, we can calculate the expected time to leave the current state and the probability with which the process will enter each of the other states, given that it has left its current state. If there are  $k$  states (other than coalescence), then we can create a system of  $k$  linear equations of the form

$$E[T_{Sex\&Asex,i}] = \frac{1}{1-v_{i,i}} + \sum_{j \neq i} \frac{v_{i,j}}{1-v_{i,i}} E[T_{Sex\&Asex,j}] \quad [3]$$

where  $v_{i,j}$  is the probability of the process transitioning from state  $i$  to state  $j$  (moving backwards in time). The system can be solved for the expected coalescent times from each state.

We make assumptions similar to those of previous models (HUDSON and KAPLAN 1994; 1995; NORDBORG *et al.* 1996a; NORDBORG 1997). We consider a single selected locus that can mutate from the wild-type  $A_1$  to the deleterious allele  $A_2$  at rate  $\mu$ . The neutral locus is at a recombination distance  $r$  from the selected locus. Two neutral alleles could be in the same individual or two separate individuals. However, for each neutral allele we also need to consider its genetic background, both the selected allele present on the same chromosome as well as the selected

allele on the homolog. We assume that  $A_1/A_2$  individuals have a fitness of  $1 - hs$  and that  $A_2/A_2$  individuals have fitness  $1 - s$  but are sufficiently rare that they can be ignored. The frequency of the deleterious allele is  $q = \mu/hs$ . Because the allele frequency is assumed to be stable, this requires  $N$  is sufficiently large so that  $Nq \gg 1$ . Technically, we assume  $1/N \sim O(\xi^3)$ , and  $hs, q \sim O(\xi)$  where  $0 < \xi \ll 1$ .

For fully sexual outcrossing species, diploid genotypes are transient so only haplotypes need to be considered. This “haplotype perspective” means that there are only three (non-coalesced) states in which two samples can exist: (i) both samples are found on  $A_1$ -haplotypes, (ii) both samples are found on  $A_2$ -haplotypes, and (iii) the two samples are on alternative haplotypes. With selfing there is the potential difficulty of needing to consider samples in the same or different individuals, potentially complicating the model with additional states. However, (NORDBORG 1997) used a separation of timescales argument to avoid this difficulty. With selfing, two samples from within the same individual either coalesce or transition into separate individuals at rates close to instantaneous with respect to coalescent times. Thus, it is possible to accommodate selfing into a system involving just three states (“the haplotype perspective”) by the appropriate adjustments of the transition probabilities.

In contrast, the separation of timescales assumptions is not applicable with partial asexuality as two samples may remain non-coalesced in the same individual for an extended period if rates of sex are low. Consequently, the system becomes considerably more complicated as we must keep an explicitly diploid perspective. Assuming homozygous mutant ( $A_2/A_2$ ) genotypes are sufficiently rare that they can

be ignored, there are eight (non-coalesced) states that must be considered: (i) both samples in a single  $A_1/A_1$  individual; (ii) both samples in a single  $A_1/A_2$  individual; (iii) both samples in separate  $A_1/A_1$  individuals; (iv) one sample in an  $A_1/A_1$  individual and the other on the  $A_1$ -haplotype of an  $A_1/A_2$  individual; (v) one sample in an  $A_1/A_1$  individual and the other on the  $A_2$ -haplotype of an  $A_1/A_2$  individual; (vi) both samples on the  $A_1$ -haplotype of separate  $A_1/A_2$  individuals; (vii) both samples on the  $A_2$ -haplotype of separate  $A_1/A_2$  individuals; and (viii) the samples in separate  $A_1/A_2$  individuals, with one on an  $A_1$ -haplotype and the other on an  $A_2$ -haplotype. A complete model considering all the additional possibilities yields the same results [not shown]. The transition probabilities ( $v_{ij}$ ) among these states, allowing for both sex and asexual reproduction, are given in the supplementary material.

If we assume that sex is common ( $\sigma \sim O(1)$ ) and linkage is tight ( $r \sim O(\xi)$ ), then the mean coalescence time is

$$E[T_{Sex\&Asex}] = 2N \left( 1 - \frac{q}{\left(1 + \frac{\sigma r}{hs}\right)^2} \right) + O(\xi^{-1}) \quad [4]$$

The term in parenthesis represents the reduction in the effective population size due to background selection. This result is equivalent to the classic result (Hudson & Kaplan 1994, 1995; Nordborg *et al* 1996a) but with the effective rate of recombination  $\sigma r$  replacing  $r$  in the previous models, which assumed obligate sex. It is worth noting that partial asexuality ( $\sigma < 1$ ) has an indirect effect on coalescence via background selection (represented by the parenthetical term in [4]), even though its direct effect of asexuality (as represented in [1]) is negligible in magnitude and not shown in this approximation. In this sense, the indirect effects of

asexuality are much more important than the direct effects (unless rates of sex are very low, i.e.,  $(\sigma \sim O(1/N))$ ). The result above is very similar to the result of (ROZE 2014) for partially asexual haploids. When the rate of sex is high, the shared history of homologs within a diploid are very brief relative to the coalescent time so it is unsurprising that the diploid model behaves like a haploid one.

If we assume that sex is rare but not too rare ( $\sigma \sim O(\xi)$ ,  $N\sigma \gg 1$ ) and that linkage is loose ( $r \sim O(1)$ ) then

$$E[T_{Sex\&Asex}] = 2N \left( 1 - \frac{q}{(1 + \frac{\sigma r}{hs})^2} \Psi \right) + O(\xi^{-1}) \quad [5]$$

where

$$\Psi \equiv \frac{2 + \phi(2 + 4r) + \phi^2(1 + 4r^2)}{(1 + \phi)^2} \quad [6]$$

and  $\phi \equiv \sigma/hs$ . Under these assumptions, sex is still sufficiently common that the direct effect of asexual reproduction on coalescent time (via the effect highlighted in eq. (1)) is negligible. However, through the factor  $\Psi$  in [21], the reduced frequency of sex enhances the effect of background selection beyond the effect shown in [4], i.e., reducing  $r$  to  $\sigma r$ . We now turn our attention to biologically interpreting  $\Psi$ .

This additional effect of low sex is clearer when we consider tight linkage ( $r \sim O(\xi)$ ) and even lower sex ( $\sigma \sim O(\xi^2)$ ), then eq. [5] simplifies to

$$E[T_{Sex\&Asex}] = 2N(1 - 2q) + \frac{1}{\sigma} + O(\xi^{-1}) \quad [7]$$

In this approximation, we also see the “direct” effect of sex on increasing the coalescent time,  $1/\sigma$  (technically, [7] gives the coalescence time for two alleles sampled from separate individuals). The reduction in  $N_e$  due to background

selection is captured by the factor  $(1 - 2q)$ . This result is different than NORDBORG *et al.* (1996a) who, based on their result similar to eq. [4], showed that with no recombination the background selection term is  $(1 - q)$ . The difference arises because the NORDBORG *et al.* (1996a) result assumes no recombination but obligate sex. Our result can be understood using the insight of CHARLESWORTH *et al.* (1993) that, without genetic mixing, the effective population size is reduced to mutation-free individuals. The frequency of mutation-free haplotypes is  $1 - q$  but the frequency of mutant-free diploid genotypes is  $1 - 2q$ . We can take [7] to mean that, in diploids with little sex, a focal neutral site is affected by the presence of a deleterious allele on its homolog just as much as it would be by one on its own chromosome, thus accounting for the  $2q$  in place of  $q$ . With somewhat higher rates of sex, the neutral site is more likely to be freed by segregation from its homolog and thus will be less affected by mutations there.

The extent to which a focal site will be affected by mutations on the homolog with partial asexuality is captured by  $\Psi$ . This can be seen most easily when  $r$  is small (and  $\sigma \sim O(\xi)$ ) so that

$$\Psi_{r \sim 0} \approx \frac{2 + 2\phi + \phi^2}{(1 + \phi)^2} \quad [8]$$

We can interpret  $\phi$  as representing the odds that a neutral locus will escape a deleterious allele on the homolog before it is eliminated by selection because  $\phi \equiv \sigma/hs$  represents the persistence time of a deleterious allele ( $1/hs$ ) relative to the waiting time for sex ( $1/\sigma$ ). From the equation above, it is easy to see that  $\Psi \rightarrow 1$  as  $\phi \rightarrow \infty$  (i.e.,  $hs \ll \sigma \ll 1$ ) and  $\Psi \rightarrow 2$  as  $\phi \rightarrow 0$  (i.e.,  $\sigma \ll hs \ll 1$ ). In other words,

background selection will depend as much on mutations on the homolog as on the focal chromosome ( $\Psi \rightarrow 2$ ) if the neutral site is unlikely to escape a deleterious allele on the homolog via sex before that allele is eliminated by selection.

Returning to the full version of  $\Psi$  in eq. [6], we find the somewhat surprising result that the sensitivity of background selection to mutation on the homolog *increases* with recombination, i.e.,  $\partial\Psi/\partial r = 4\phi(1+2r\phi)/(1+\phi)^2 > 0$ . This can be understood as follows. When  $r = 0$ , the focal neutral site is guaranteed to escape the effects of a deleterious allele on the homolog via segregation if sex occurs. However, as  $r$  increases, the focal site is more likely to recombine onto the deleterious homolog during sex rather than escaping it. In fact, for unlinked loci ( $r = 1/2$ ), we have  $\Psi = 2$  (regardless of  $\phi$ ), indicating the focal site is just as strongly affected by mutations on the homolog as on its own chromosome.

In sum, asexual reproduction affects background selection in two ways: (i) via a “recombination effect” (reducing the effective rate of recombination from  $r$  to  $\sigma r$ ), and (ii) via a “segregation effect” (captured by  $\Psi$ , which mediates how much neutral alleles on one haplotype are affected by mutations on the homolog). Both of these effects on background selection occur even when the rate of sex is much higher than is required for a substantial direct effect of sex on coalescent time (i.e.,  $N\sigma \gg 1$ ). As a result of these effects, background selection can be much stronger in facultative asexuals; this is particularly noticeable at larger recombination distances (Fig. 2).

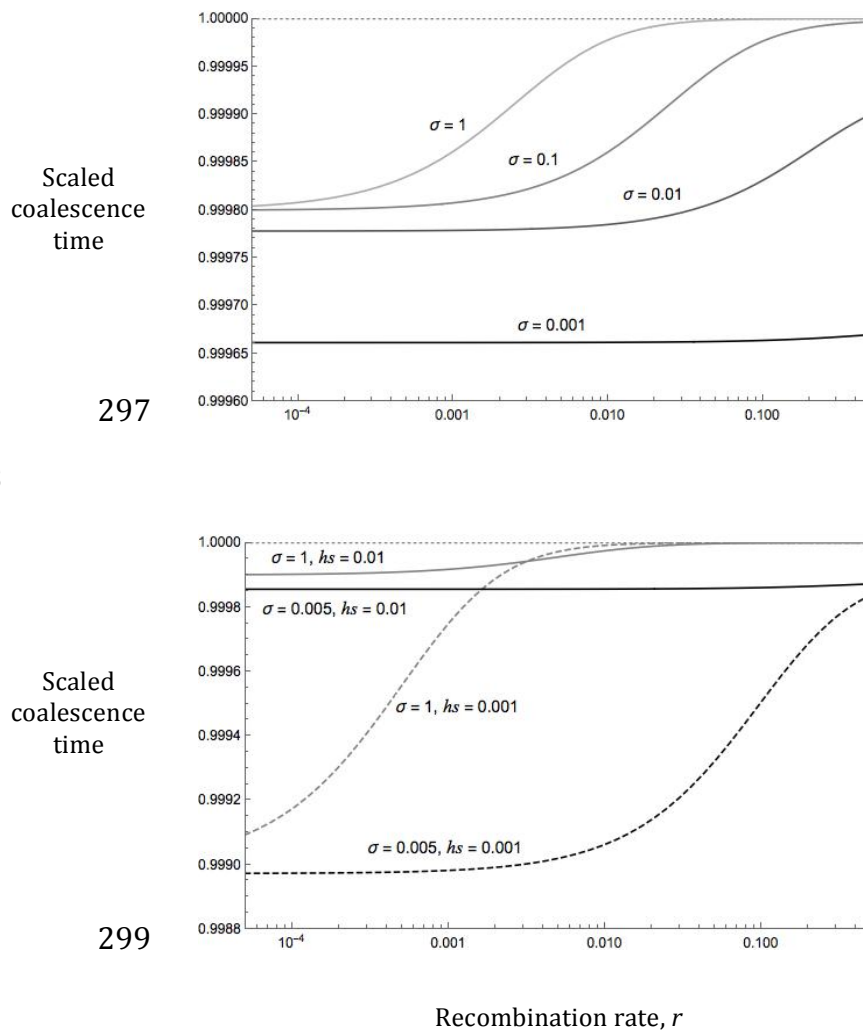


Figure 2. The scaled coalescence time ( $E[T]/2N$  or, equivalently,  $N_e/N$ ) as a function of the recombination distance between the focal neutral locus and the selected locus in the background selection model with partial asexuality. The reduction in coalescence times with low rates of sex is particularly at higher rates of recombination (A). The effects of background selection are larger with weak selection when recombination is low (B). Mutation rate:  $\mu = 10^{-6}$ .

## Extension to multiple loci

Following earlier models (HUDSON and KAPLAN 1995; NORDBOG *et al.* 1996a), we can extend the single-locus background selection results to many loci under the assumption of no disequilibria and no epistasis. Let  $B_i$  be the background selection factor for a single locus (i.e., the term in parenthesis of eq. [5]). Across many loci, the total effect of background selection is  $B = \exp [\sum(1 - B_i)]$ . Using eq. [5], we can calculate

$$B_{Sex\&Asex} \approx \exp \left[ - \int \frac{q}{\left(1 + \frac{\sigma R[x]}{hs}\right)^2} \Psi' dx \right] \quad [9]$$

where the integration is over all positions  $x$  in the genome.  $R[x]$  is the recombination rate between the focal locus and the position  $x$  and  $\Psi'$  is  $\Psi$  as given by [6] but replacing  $r$  with  $R[x]$ .

Although [9] does not reduce to a simple analytical expression applicable to any genomic architecture, it can be evaluated numerically. Moreover, we can decompose the total background selection coefficient into components arising due to strongly linked, loosely linked, and unlinked genes:  $B = B^{TightlyLinked} B^{LooselyLinked} B^{Unlinked}$ . Useful analytical approximations can be obtained for  $B^{TightlyLinked}$  and  $B^{Unlinked}$ .

Here, we use “tightly linked” to refer to sites that are within a distance  $m_{max}$ , measured in Morgans, of the focal site, where  $m_{max}$  is the maximum distance over which the map distance ( $m$ ) is approximately linear with recombination rate. In other words,  $m_{max}$  is the maximum value of  $m$  where  $m \approx r$ ; for typical mapping functions,  $m_{max} < 0.05$ . We will consider a region  $2m_{max}$  wide with the focal site at its



center as “the tightly linked region” and assume deleterious mutations fall uniformly across this region. The background selection coefficient from the tightly linked region is

$$B_{Sex\&Asex}^{TightlyLinked} \approx \exp \left[ -U_{TightlyLinked} \frac{1}{2(\sigma m_{max} + hs)} \right] \quad [10a]$$

where  $U_{TightlyLinked}$  is the diploid mutation rate over the tightly linked region. For the case of  $\sigma = 1$ , this result is the same as eq. [8] of HUDSON and KAPLAN (1995). A much better approximation when  $\sigma < hs$  is

$$B_{Sex\&Asex}^{TightlyLinked} \approx \exp \left[ -U_{TightlyLinked} \frac{\Psi_{r \sim 0}}{2hs} \right] \quad [10b]$$

where  $\Psi_{r \sim 0}$  is given by [8].

Except in genomes with very small maps, most genes will be unlinked ( $r = 1/2$ ) to the focal neutral site. Unlinked loci make a small contribution to the total effect of background selection in species where sex is high. With full sex ( $\sigma = 1$ ) it has been shown by other means (CHARLESWORTH 2012) that

$$B_{FullSex}^{Unlinked} \approx \exp [-4U_{Unlinked}hs] \quad [11]$$

where  $U_{Unlinked}$  is the diploid rate of deleterious mutation over all sites unlinked to the neutral focal site. However, when sex is low ( $\sigma \ll 1$ ), then the background selection coefficient due to unlinked sites based on [5] and [9] is

$$B_{Sex\&Asex}^{Unlinked} \approx \exp \left[ -\frac{4U_{Unlinked}hs}{(\sigma + 2hs)^2} \right] \quad [12]$$

where  $U_{Unlinked}$  is the diploid rate of deleterious mutation over all sites unlinked to the neutral focal site. It is clear from eqs. [10-12], that the effects of background selection are much greater when sex is low, especially the contribution from unlinked loci.

To explore background selection further, we will assume that the focal chromosome exists in the center of a 1 Morgan chromosome and that the total map length of the genome is  $L$  Morgans (where  $L \geq 1$ ). Deleterious mutations are uniformly distributed across the genome, i.e.,  $\mu = \frac{1}{2} U/L$  where  $U$  is the diploid genome-wide mutation rate. This means that  $U_{TightlyLinked} = 2Um_{max}/L$  and  $U_{Unlinked} = U(L - 1)/L$ . The background selection coefficient due to loosely linked ( $m_{max} < r < \frac{1}{2}$ ) sites is

$$B_{Sex\&Asex}^{LooselyLinked} \approx \exp \left[ -2 \int_{m_{max}}^{1/2} \frac{q}{\left(1 + \frac{\sigma R[x]}{hs}\right)^2} \Psi' dx \right] \quad [13]$$

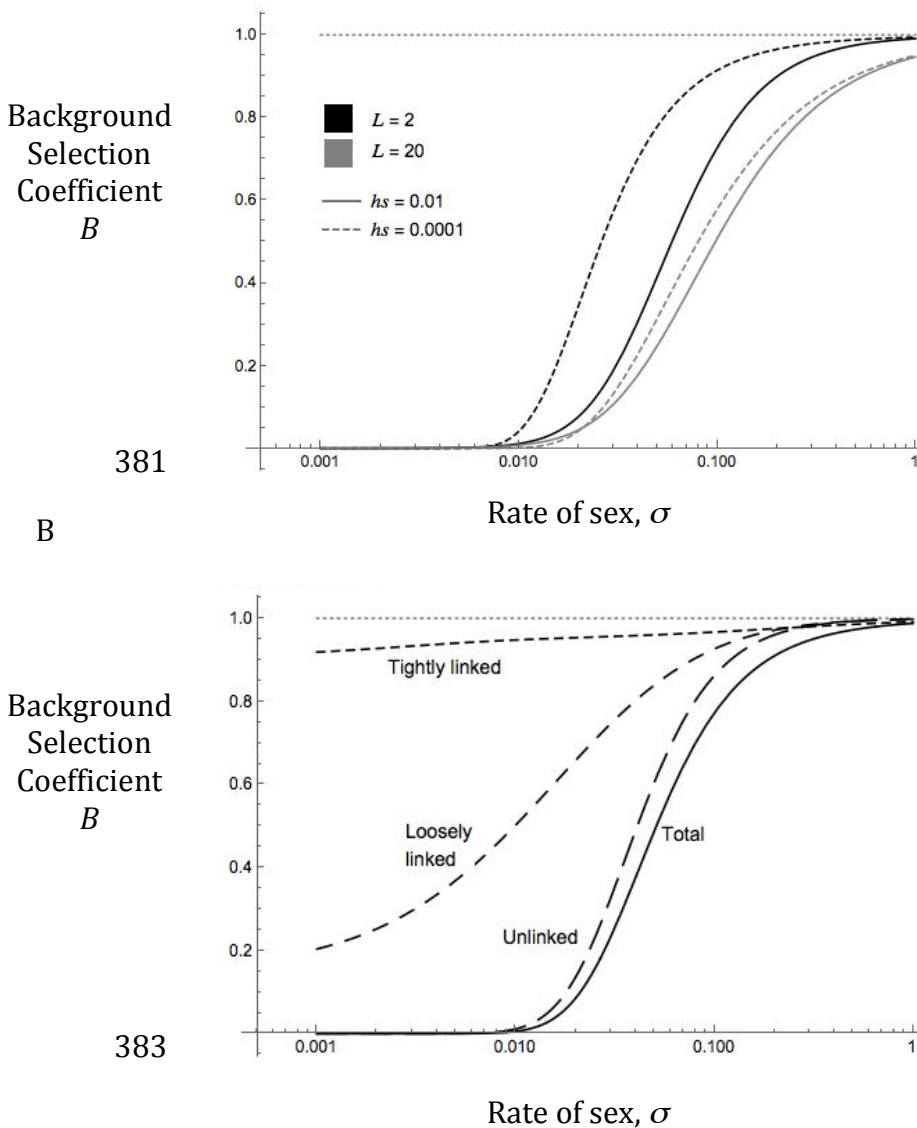
This integral can be expressed in closed form using Haldane's mapping function ( $R[x] = \exp[1 - 2x]/2$ ) but the expression is rather complicated and is left for the supplementary material.

Figure 3 illustrates that the genome-wide effects of background selection can be substantial when  $\sigma < 1$ , even in genomes with long maps (such that gene density is low). If sex is low ( $\sigma < 1\%$ ), effective population size may be reduced by one or more orders of magnitude with relatively mild mutation rates ( $U = 0.1$ ). In a moderate to large population ( $N > 10^4$ ), there would be a negligible direct effect on coalesce time of asexual reproduction with  $\sigma = 10\%$  or  $1\%$ , (Bengtsson 2003; Ceplitis 2003; Hartfield et al 2015), but the indirect effect through background selection can be large. The effect of background selection is more sensitive to the strength of selection in long maps than in short maps, with stronger selection resulting in larger effects of background selection, provided  $\sigma$  is not too low. With lower sex, a larger proportion of total background selection is attributable to

375   unlinked loci (Fig. 3B). This proportion is greater in larger genomes (as more loci  
376   will be unlinked) and when selection is strong, as strongly selected loci do not need  
377   to be closely linked to exert their effects (NORDBORG *et al.* 1996a).

378

379



385 Figure 3. Genome-wide background selection in systems with sexual and asexual  
386 reproduction. (A) The background selection coefficient  $B$  as a function of the rate of  
387 sex considering mutations across the entire genome. Results for genomes with  
388 relative short and long maps ( $L = 2$  vs.  $20$  Morgans) are shown. The longer map  
389 implies lower gene density as mutation rate is held constant ( $U = 0.1$ ); mutations are  
390 assumed to be uniformly distributed across the genome. (B) The background  
391 selection coefficient for different genomic regions (tightly linked, loosely linked, and  
392 unlinked sites) as well as the total; parameters  $L = 10$ ,  $h_s = 0.005$ ,  $U = 0.1$ . Tightly  
393 linked loci make the largest contribution to background selection when rates of sex  
394 are very high but unlinked loci make the largest contribution when sex is rare.

## Simulation Results

The extrapolation to multiple loci used in [9] and in previous work (HUDSON and KAPLAN 1995; NORDBORG *et al.* 1996a; GLÉMIN and RONFORT 2012) relies on the assumption that loci are independent (i.e., no linkage disequilibrium) and that each selected locus is close to its deterministic equilibrium ( $q = \mu/hs$ ). These assumptions become increasingly suspect as the effective rate of recombination declines and as background selection becomes sufficiently strong that  $N_e$  ( $\approx BN$ ) is substantially reduced. When  $BNhs$  is predicted to be low ( $< 1$ ), then we expect that analytical approximations will overestimate the strength of background selection. Under these conditions the analytical approximation is expected to overestimate the reduction in  $N_e$  due to background selection, presumably because when drift is strong then polymorphism is lost from selected sites, i.e.,  $q = 0 < \mu/sh$ , making background selection weaker than predicted. Previous simulation studies have observed that analytical approximations tend to overestimate the strength of background selection when selection is weak or recombination is low (NORDBORG *et al.* 1996a; KAISER and CHARLESWORTH 2009; KAMRAN-DISFANI and AGRAWAL 2014).

To test the analytical approximations presented here, we extended the simulation of (KAMRAN-DISFANI and AGRAWAL 2014) to consider a population where individuals can reproduce asexually with a probability  $\sigma$ , via Wright-Fisher dynamics. We measured  $N_e$  by tracking a linked quantitative trait over time; when an individual reproduced, the value of this trait was perturbed by a value drawn from a normal distribution with mean 0 and variance 1. The steady-state variance of

this trait is equal to the effective population size (LYNCH and HILL 1986; KEIGHTLEY and OTTO 2006).

For a given net mutation rate  $U$  and heterozygote selection coefficient  $sh$ , mutations were initially distributed amongst the population according to mutation-selection balance. Unless stated otherwise, the population then reproduced for  $5N$  generations, to ensure that neutral markers reached a steady state (KEIGHTLEY and OTTO 2006). The population-wide  $N_e$  was then measured at 200 equally spaced intervals for a further  $5N$  generations. The simulation was repeated for 192 burn-ins. The mean  $N_e$  from each time series was determined, and simulation points are calculated as the means of these values along with 95% confidence intervals. In general there is a good match between simulation results and analytical approximations (Fig. 4A). The effects of background selection tend to be overestimated if sex is low ( $\sigma \sim 0.001$ ), where background selection is predicted to be very strong.

Because the analytical approximation for  $B$  is expected to fail as  $BNhs$  becomes small, we examined the correspondence between simulations and the predicted value of  $B$  for different population sizes. As expected, the observed value of  $B$  is much larger than the analytical approximation (i.e., the strength of background selection is overestimated by [9]) when  $N$  is small but the discrepancy declines with larger  $N$  (Fig. 4B, C). This implies that the analytical approximations work well provided  $N$  is ‘sufficiently’ large. Though imperfect, the approximations offer good insight into the effects of background selection. Nonetheless, it is important to remember that the extremely strong effects of background selection

predicted under low rates of sex (e.g.,  $B < 10^{-5}$ ) are unlikely to be realized unless applied to species with a massive census size. Though the most obvious problem is that the analytical approximation overestimates background selection when  $N$  is not sufficiently large, we also find instances where the analytical approximation underestimates the strength of background selection when  $N$  is large (Fig. 4B). This could suggest a weak effect arising from multi-locus associations not captured in the extrapolation of the single-locus analysis to genome-wide background selection given by [9]. This effect does not appear to be sensitive to  $h$  (not shown) and so is unlikely to be a direct result of segregation load.

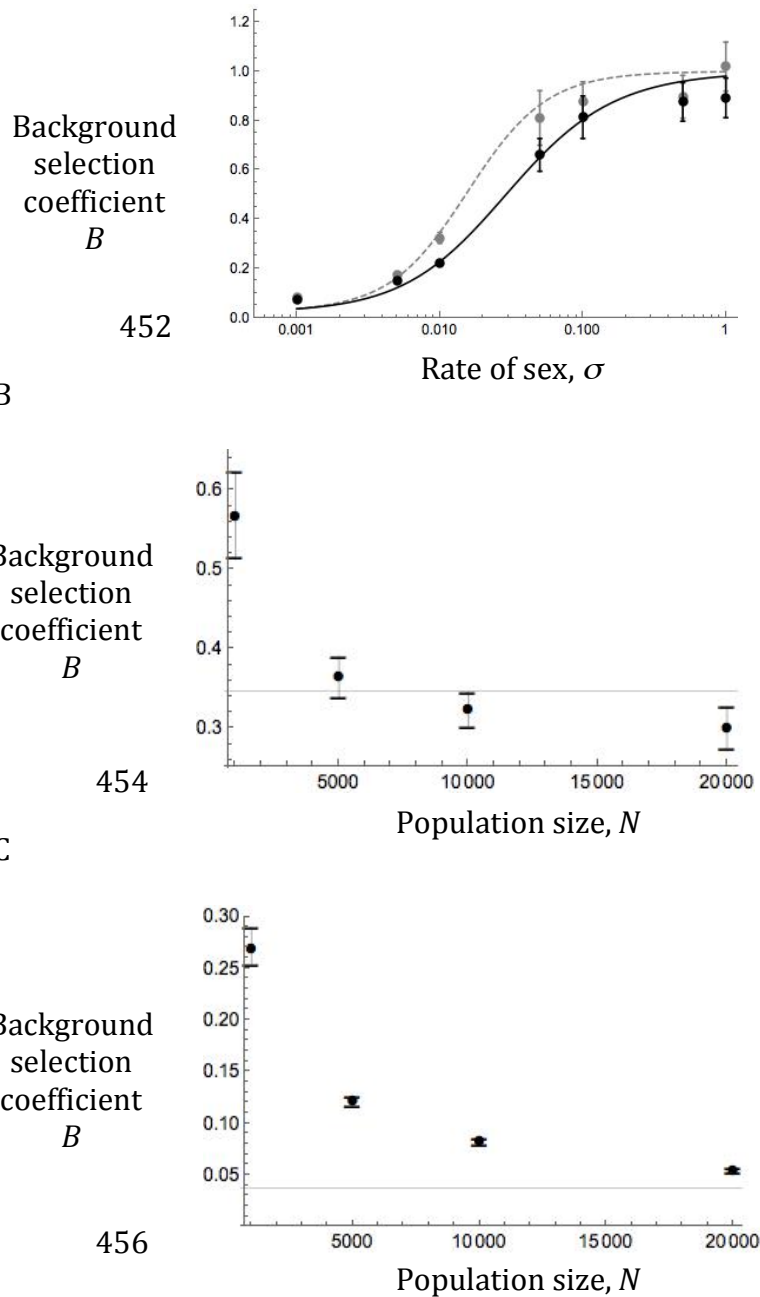


Figure 4: Simulation comparisons of background selection. (A) Simulation results (points) and analytical approximations (lines) of  $B = N_e / N$  as a function of the rate of sex. Parameters are  $U = 0.02$ ,  $s = 0.02$ ,  $h = 0.25$ , and  $L = 1$  (black) or  $L = 10$  (grey). Bars on simulation points represent 95% confidence intervals. (B) Background selection effect for  $\sigma = 0.01$  as a function of the population size. Parameters are  $U = 0.02$ ,  $s = 0.02$ ,  $h = 0.25$ , and  $M = 10$ . The horizontal black line is the theoretical expectation given with Equation 9. (C) Same as (B) but for  $\sigma = 0.001$ .



## Comparison with selfing

Another mode of uniparental reproduction, selfing, also enhances the effect of background selection. As shown by NORDBERG (1997), the coalescent time with selfing is

$$E[T_{Outcross\&Self}] = 2N \frac{1}{1+F} \left( 1 - \frac{\tilde{q}}{\left(1 + \frac{\tilde{r}}{\tilde{t}}\right)^2} \right) \quad [14]$$

Here  $F$  is the inbreeding coefficient which can be written as function of the rate of outcrossing  $F = (1 - o)/(1 + o)$ . The fraction  $1/(1+F)$  is the direct effect of selfing on reducing  $N_e$  (shown in [2] as  $(1 + o)/2$ ), and will not be considered further. The term in parenthesis in [14] is the background selection coefficient. With pure outcrossing the background selection coefficient is the same as the parenthetical term from [4] with  $\sigma = 1$ . Relative to this, selfing changes background selection in two ways. First, it increases the average strength of selection against a deleterious mutation from  $sh$  to  $\tilde{t} = (1 - F)sh + Fs$ , where  $s$  is selection against the mutation in the homozygous state. This also changes the equilibrium frequency of the deleterious allele to  $\tilde{q} = \mu/\tilde{t}$ . More importantly, selfing reduces the effective rate of recombination from  $r$  to  $\tilde{r} = (1 - F)r$ . When selfing rates are high ( $o \ll 1$ ), the background selection coefficient simplifies to

$$B_{Outcross\&Self}^{1-locus} \approx 1 - \frac{\tilde{q}}{\left(1 + \frac{2or}{s}\right)^2} \quad [15]$$

There are three important differences between this result with the equivalent term for background selection in partial asexuals [5]. First, the segregation effect

captured by  $\Psi$  effectively doubles the relevant mutation rate for partial asexuals relative to selfers. Second, a bit of outcrossing in a highly selfing system results in effectively twice as much recombination as the equivalent amount of sex in a highly asexual system (i.e.,  $\sigma$  in [5] vs.  $2\sigma$  in [15]). In sex/asex systems, recombination can only occur with sexual reproduction so the effective recombination rate depends directly on  $\sigma$ . In outcross/self systems, recombination can happen during the production of selfed or outcrossed progeny. However, the only parental genotypes in which recombination is relevant are double heterozygotes, which are created at a rate proportional to the rate of outcrossing,  $o$ . However, once a double heterozygote is created via outcrossing, a relevant recombination event can occur in any future generation via selfing as long as the genotype remains a double heterozygote. Because it takes, on average, two generations of selfing before heterozygosity at a site is lost, there are two generations in which meaningful recombination can occur following each outcrossing event (so we have  $2\sigma$  in [15] but  $\sigma$  in [5]). However, Denis Roze has pointed out that [14] is derived from [4], which assumes tight linkage ( $r \ll 1$ ). He has found (pers. comm.) that the “2” is reduced for loosely linked loci, becoming “1” for unlinked loci. That is, the recombinational difference between selfing and partial asexuals will disappear for unlinked loci.

In addition to these segregation and recombination differences, a third difference arises because the effective strength of selection is greater with selfing than asexual reproduction ( $s$  vs.  $hs$ ). Stronger selection reduces the frequency of deleterious alleles. As there are fewer deleterious backgrounds to avoid in tracing

back a focal neutral site's ancestry, background selection is made weaker. However, stronger selection also means there is less opportunity to recombine away from a deleterious allele before it is removed by selection, enhancing background selection. Overall, stronger selection reduces background selection if the rate of biparental reproduction (and the opportunity for recombinational escape) is high relative to selection.

As a consequence of the three differences between [15] and [5], background selection can be much stronger in partial sexuals than in selfers when the rate of biparental reproduction is low (Fig. 4). However, the reverse may be possible if the net effect of stronger selection in selfers is to increase background selection, which can happen when the rate of biparental reproduction is high relative to selection.

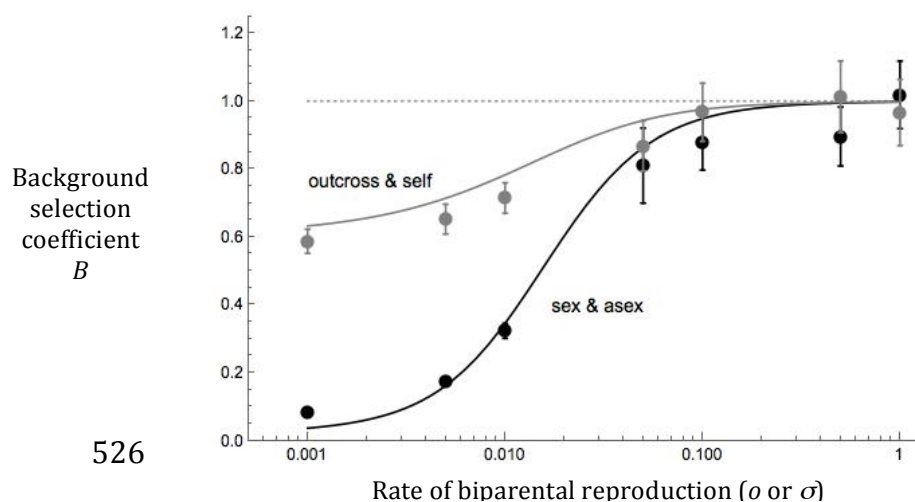


Figure 4. Background selection in systems with different forms of uniparental reproduction (asexuality and selfing). The background selection coefficient  $B$  is shown for systems with sexual and asexual reproduction (black) and systems with outcrossing and selfing (grey). Analytical approximations (lines) are contrasted with simulation results (points; error bars are 95% confidence intervals). The analytical approximation for selfing is based on an extension of [14] and is given in Kamran-Disfani & Agrawal (2014). Parameters:  $U = 0.02$ ,  $s = 0.02$ ,  $h = 0.25$ ,  $L = 10$ , ( $N = 10^4$  for simulations).

## Balancing Selection

With balancing selection, such as heterozygote advantage or negative frequency dependence, two alleles ( $A_1$  and  $A_2$ ) can be maintained indefinitely, i.e., balancing selection prevents normal coalescent processes from operating at the selected site and this affects coalescence at linked sites. Assuming the frequencies of the  $A_1$  and  $A_2$  ( $p$  and  $q$ , respectively) are stable, HUDSON and KAPLAN (1988) showed that closely linked neutral sites have an extended coalescent time under the assumption of obligate sexual reproduction. As mentioned in the previous section, the analysis becomes more complicated when the rate of sex is low because we must use an explicitly diploid perspective rather than classic haploid perspective. Here we cannot assume that the frequency of the  $A_2/A_2$  genotype is negligible. As a result, there are 13 states in which we could find two alleles. In addition to the eight listed in the previous section, the additional five states are: (ix) both samples in a single  $A_2/A_2$  individual; (x) both samples in separate  $A_2/A_2$  individuals; (xi) one sample in an  $A_2/A_2$  individual and the other on the  $A_1$ -haplotype of an  $A_1/A_2$  individual; (xii) one sample in an  $A_2/A_2$  individual and the other on the  $A_2$ -haplotype of an  $A_1/A_2$  individual; and (xiii) one sample in an  $A_1/A_1$  individual and the other sample in an  $A_2/A_2$  individual.

In the previous section we ignored gene conversion for simplicity and because the persistence time of any given copy of the deleterious allele is short, limiting opportunity for gene conversion to be relevant. Here we include gene conversion because  $A_1$  and  $A_2$  are maintained indefinitely by selection so movement between genetic backgrounds via gene conversion may become important over long

time scales, especially when the rate of sex is low so there is little chance for exchange via meiotic recombination. Specifically, we assume that (unbiased) mitotic gene conversion occurs at rate  $\gamma$ . This rate refers to gene conversion events that cover either the selected site or the neutral site but not both. (The model is built with an additional parameter for gene conversion events that cover both sites but such events do not affect the coalescent time.) Throughout, we use “recombination” to refer to (meiotic) cross-over recombination, in contrast to gene conversion.

In this analysis, we assume that the equilibrium frequencies of  $A_1/A_1$ ,  $A_1/A_2$ , and  $A_2/A_2$  genotypes are  $P_{1/1}$ ,  $P_{1/2}$ , and  $P_{2/2}$ . These frequencies are assumed to be stable and to change little in frequency within a generation (i.e., fitness differences at equilibrium are ignored). Without loss of generality, we use  $P_{1/1} = p^2 + C_{A/A}$ ,  $P_{1/2} = 2(pq - C_{A/A})$ , and  $P_{2/2} = q^2 + C_{A/A}$  where  $C_{A/A}$  is the covariance in allelic state. Note  $C_{A/A}$  measures the same genetic property as  $Fpq$  but we use separate symbols because we use the relationship  $F = (1 - o)/(1 + o)$  for species without asexuality because the inbreeding coefficient is determined only by outcrossing rates, to a good approximation, provided selection is not too strong. With obligate sex (and no selfing)  $C_{A/A} = 0$  but when there is partial asexuality,  $C_{A/A}$  can be either positive or negative.

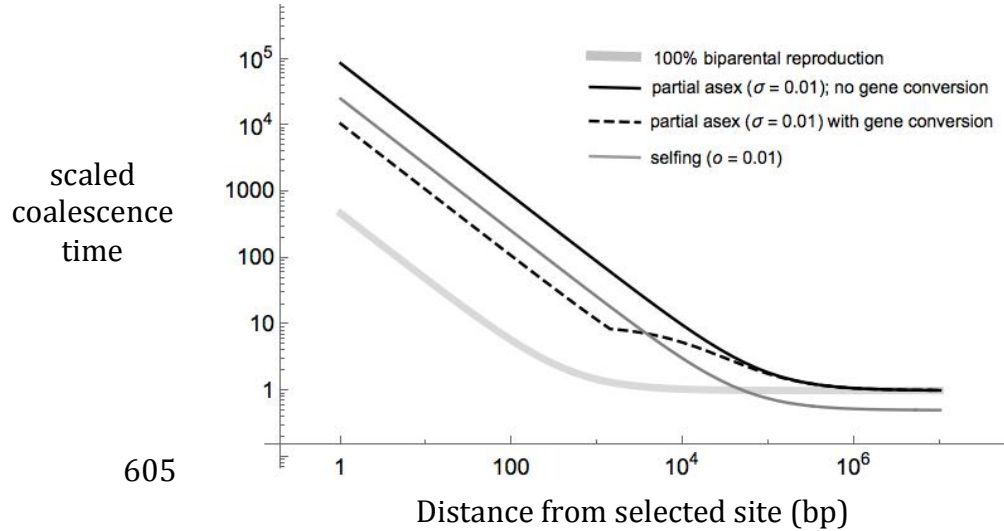
Assuming  $\sigma$  and  $r$  are  $O(\xi)$  and  $\gamma$  and  $1/N$  are  $O(\xi^2)$ , the expected coalescent time of two alleles sampled both from a single  $A_1/A_1$  individual or from two separate  $A_1/A_1$  individuals is found to be

$$E[T_{1/1}] = 2N + \frac{4N(p-q)q}{1+2NP_{1/2}(\sigma r + \gamma)} + O(\xi^1) \quad [21]$$

When both alleles are sampled from the  $A_2$  background, the coalescent time is the same as [21] but with  $p$  and  $q$  reversed. When two alleles are sampled from alternate backgrounds ( $A_1$  and  $A_2$ ), the expected coalescent time

$$E[T_{1/2}] = 2N + \frac{2pq}{P_{1/2}(\sigma r + \gamma)} + O(\xi^{-1}) \quad [22]$$

For the sake of discussion, we will focus on this latter result. In the case of obligate sex ( $\sigma = 1$ ,  $P_{1/2} = 2pq$ ), [22] simplifies to  $E[T_{1/2}] = 2N + 1/(r + \gamma)$ . The coalescent time is extended by the waiting time for the neutral site to move from one background to the other. This longer coalescent time should allow for the accumulation of more neutral variation surrounding sites under balancing selection (Hudson & Kaplan 1988). When the physical distance is small,  $r$  may be very small and this has led to the belief there may be a strong signature of balancing selection when gene conversion is ignored. However, for short physical distances, gene conversion may be much larger than recombination ( $\gamma \gg r$ ) and may play a more important role in coalescence (ANDOLFATTO and NORDBORG 1998; WIUF and HEIN 2000). As implied by [22], gene conversion weakens the signature of balancing selection and this is particularly true in partial asexuals (Fig. 6). At very short physical distances, mutation between  $A_1$  and  $A_2$  provides an important alternate route by which a neutral allele can switch sites, but this is not considered here. Our results apply to physical distances where  $\sigma r + \gamma \gg \mu$ .



**Figure 6. Coalescent times near a site under balancing selection.** The expected time to coalescence (scaled to  $2N$  generations) is shown as a function of the distance (in base pairs) from a site under balancing selection via heterozygote advantage ( $W_{1/1} = W_{2/2} = 1 - s$  and  $W_{1/2} = 1$ ). The population is assumed to be fully sexual and outcrossing (eq. [22],  $\sigma = 1$ ; thick gray line), mostly asexual (eq. [22],  $\sigma = 0.01$ ; black line), or mostly selfing (eq. [24b],  $\sigma = 0.01$ ; thin gray line). We assume  $r = \rho d$  where  $d$  is the distance in base pairs from the selected site. For the partial asexual case, results are shown with and without gene conversion. For gene conversion, the model of Andolfatto and Nordborg (1998) is used:  $\gamma = gd/L$  for  $d < L$  and  $\gamma = g$  for  $d \geq L$  where  $L$  is the length of gene conversion tracts. For the fully sexual case, the results with gene conversion are visually indistinguishable for those with  $\gamma = 0$  for the parameter values used here. The selfing case assumes no gene conversion but gene conversion is not expected to affect coalescence in selfers (see Discussion). The selfing line asymptotes at  $1/2$  rather than 1 because of the direct effect of selfing on coalescence (Fig. 1). These results ignore mutation at the selected site and thus overestimate the coalescence time for extremely tightly linked sites (e.g.,  $d \ll 100$  if  $\mu = 10^{-9}$ , see text). Parameter values:  $s = 0.02$ ,  $N = 10^5$ ,  $\rho = 10^{-8}$  (ASHBURNER 1989),  $g = 10^{-6}$  (SHARP & AGRAWAL, in review),  $L = 1500$  (PRESTON & ENGELS 1996).



As expected, we find in [22] that partial asexuality reduces the effective rate of recombination from  $r$  to  $\sigma r$ . However, there is an additional effect of asexuality that is not explicit in [22]; partial asexuality changes the equilibrium frequency of  $A_1/A_2$  (i.e.,  $P_{1/2} \neq 2pq$ ). The precise nature of this change depends on the mechanism of balancing selection. Consider balancing selection due to heterozygote advantage with symmetrical selection (i.e.,  $W_{1/2} = 1$ ,  $W_{1/1} = W_{2/2} = 1 - s$ ). With obligate sex, at equilibrium  $p = q = 1/2$  and  $P_{1/2} = 1/2$ . With partial asexuality and assuming the rate of sex is low,  $p = q = 1/2$  but  $P_{1/2} = 2(pq - C_{A/A})$  where

$$C_{A/A} \approx -(\sqrt{s^2 + \sigma^2} - \sigma)/4s \quad [23]$$

For example, with  $s = \sigma = 0.01$ ,  $P_{1/2} \approx 0.71$ . In this case, the frequency of heterozygotes (and thus the opportunity for recombination) is increased from 0.5 with full sex to 0.7 with partial sex. Even though  $P_{1/2}$  is higher, this does not cause an increase in the coalescence time because of the 99% reduction in meiotic recombination events when  $\sigma = 0.01$ . Equilibrium genotype frequencies under a more general model of balancing selection are provided in the supplementary material.

NORDBORG (1997) studied balancing selection in species with selfing. He found (ignoring gene conversion) that the expected coalescence time for sites on alternate backgrounds ( $A_1$  and  $A_2$ ) is

$$E[T_{1/2}] \approx 2N \left( \frac{1}{1+F} \right) + \frac{1}{(1-F)r} \quad [24a]$$

Assuming that rate of selfing is high (i.e., weak outcrossing,  $o \ll 1$ ), this becomes

$$E[T_{1/2}] \approx 2N \left( \frac{1+o}{2} \right) + \frac{1}{2or} \quad [24b]$$

Again, the term in parenthesis is the direct effect of selfing. The second term is the increase in coalescent time due to linkage to the target of balancing selection. Contrasting this to the comparable term from the partial asexual result in [22], there are two differences. As noted in the background selection case, a little bit of outcrossing is twice as effective at allowing recombination as a little bit of sex (i.e., the effective rates of recombination are  $2or$  vs.  $\sigma r$ ). Second, in partial asexuals the frequency of heterozygotes cannot be well approximated from the rate biparental reproduction and allele frequencies alone (as it is with selfing). Rather  $C_{A/A}$ , which determines the frequency of heterozygotes through  $P_{1/2}$ , is a function of both the rate of sex and the selection coefficients (as exemplified in [23]).

An important point missing from the comparison above is the likelihood of balancing selection. Consider the model of heterozygote advantage with  $W_{1/2} = 1$ ,  $W_{1/1} = 1 - s$ , and  $W_{2/2} = 1 - \alpha s$  where we assume  $\alpha > 1$  so that  $W_{1/1} > W_{2/2}$ . Ignoring drift, polymorphism is always maintained in sex/asex systems. In contrast, with selfing, the maintenance of polymorphism requires that outcrossing not be too low, specifically that  $o > (\alpha - 1)/(\alpha + 1)$ , otherwise fixation of the more fit homozygote  $A_1/A_1$  is stable (see supplementary material for details). If one homozygote is more fit than another, variation will not be maintained by balancing selection in highly selfing species but will be in species with partial asexuality.

Of course, balancing selection can occur through mechanisms other than heterozygote advantage and in these other cases the contrast between partial asexuality and sex will be somewhat different. For example, consider a simple model of frequency dependent selection where the fitness of each genotype is

negatively related to its frequency as follows  $W_{1/1} = 1 - (p - \alpha)s$ ,  $W_{2/2} = 1 + (p - \alpha)s$ , and  $W_{1/2} = 1$ . Regardless of the reproductive system, at equilibrium  $p = \alpha$  and the polymorphism is stable (i.e., in contrast to heterozygote advantage, there is no difference in the likelihood of balancing selection in this model between partial asexuality and selfing systems; see supplementary material). However with partial asexuality, there is no excess of heterozygotes ( $C_{A/A} \approx 0$ ), meaning that coalescence is only affected by the recombination effect (i.e., the reduction from  $r$  to  $\sigma r$ ) and there is no segregation effect. With other forms of negative frequency dependent selection (supplementary material), it is possible for there to be a deficit of heterozygotes with partial asexuality ( $C_{A/A} > 0$ ), decreasing the opportunity for recombination beyond the change from  $r$  to  $\sigma r$ , thereby further extending coalescent time.

## Discussion

We have examined how linked selection, either background or balancing selection, alters coalescence for pairs of alleles in partial asexuals. As expected, the lower level of genetic mixing in partial asexuals results in much stronger effects of linked selection. In diploids, partial asexuality affects coalescence both via segregation and recombination. The most obvious effect is through a reduction in the effective rate of recombination. In the case of background selection, the segregation effect can be understood as mediating how much a focal site is affected by deleterious alleles occurring on the homolog in addition to those on its own chromosome. If sex is low, it is equally affected by mutations on either chromosome, effectively doubling the

mutation rate. In the case of balancing selection, the segregation effect is manifest in the way that partial asexuality affects the equilibrium frequency of heterozygotes at the selected locus, which regulates the opportunity for a neutral site to move between alternative selected haplotypes via recombination.

Ignoring linked selection, the direct effect of partial asexuality on coalescence is negligible unless the rate of sex is extremely small (i.e., on the order of  $1/N$ ; (BENGTTSSON 2003; CEPLITIS 2003; HARTFIELD *et al.* 2015)). In contrast, there are strong effects of partial asexuality through background selection even when rates of sex are much larger than  $1/N$ . Large reductions in  $N_e$  through background selection are expected under realistic mutation rates ( $U \geq 0.1$ ) if sex is below 10%.

When sex is high, only those mutations that are tightly physically linked make a substantial contribution to background selection. Across the genome of a highly sexual species, variation in gene density resulting in variation in the local deleterious mutation rate leads to variation in the strength of background selection, which contributes to variation in neutral diversity (CHARLESWORTH and CAMPOS 2014; ELYASHIV *et al.* 2014). These local differences may be detectable at the within-gene level (i.e., stronger background selection at the center of genes than on the edges, (LOEWE and CHARLESWORTH 2007; ZENG and CHARLESWORTH 2011)). When sex is low, deleterious alleles across the whole genome contribute to background selection. In fact, the majority of background selection experienced by a focal site is due to unlinked loci (Fig. 3B) but this is simply because unlinked sites are much more numerous. Even when sex is low, background selection is stronger from tightly linked sites than unlinked ones (Fig. 2B) so one might still expect to find differences

in neutral diversity associated with gene density. However, the relationship between background selection and  $r$  is much weaker when sex is low, so gene density-diversity patterns are likely to be severely muted.

Because background selection can be so strong with uniparental inheritance, it may be more important than demographic history in determining within-population diversity. As has recently been discussed with respect to selfers (BRANDVAIN *et al.* 2013; BARRETT *et al.* 2014), transitions to uniparental inheritance can be associated with population bottlenecks but it can be difficult to infer whether reduced diversity in an evolutionarily young population is due to a small number of founders or the strong effects of background selection.

The analytical approximations not only quantify how much coalescence times are altered by the reproductive system, but they are also biologically interpretable. Hence these derivations provide insight into how segregation, recombination, and selection interact to affect average coalescence time. However, the approach used here cannot be used to make other, more detailed, predictions about  $n$ -sample genealogies. For example, although the major effect of background selection can be interpreted as a reduction in  $N_e$  (as done here), it has been shown that background selection also alters the shape of genealogies with relatively long external branches, leading to negative values of Tajima's  $D$  (CHARLESWORTH *et al.* 1995; ZENG and CHARLESWORTH 2011; NICOLAISEN and DESAI 2012). Presumably, these effects would be magnified by low rates of sex.

The reduction in the effective rate of recombination with uniparental inheritance should make it much easier to detect balancing selection. Work by

NORDBORG *et al.* (1996b) and NORDBORG (1997) showed that, in selfers, this is because (i) the physical distance over which the effects of balancing selection are realized is expanded and (ii) there is lower diversity within each selected allelic class because of stronger background selection. These arguments also apply to partial asexuals but there are differences between these alternative modes of uniparental reproduction. First, there is a higher effective rate of recombination with selfing than with partial asexuality for the same rate of uniparental reproduction ( $2\sigma$  vs.  $\sigma$ , as discussed in the Results), which will make signatures of balancing selection weaker in selfers than partial asexuality (Fig. 6). There are also differences in the frequency of heterozygotes in partial asexuals compared to selfers but this likely plays a more minor role, conditional on the selected polymorphism being maintained. However, with heterozygote advantage, the conditions for a stable polymorphism are much broader in partial asexuals than selfers.

Another difference between partial asexuals and selfers is the importance of gene conversion. The frequency of heterozygotes puts an upper limit on the rate of exchange between allelic backgrounds via any genetic mechanism. Selfing directly reduces the frequency of heterozygotes, so gene conversion is not expected to be an important process (though it was not formally included in the model of (NORDBORG 1997)). In partial asexuals, heterozygotes can be very common even if sex is rare (especially if there is heterozygote advantage). Through asexual descent, a lineage can persist in heterozygous state for many generations, providing numerous opportunities for movement of linked sites between backgrounds via gene conversion. As implied by [22], gene conversion becomes the primary force erasing

the signature of balancing selection when  $\sigma r \ll \gamma$ , which must be true if sex is sufficiently rare. Because of this, the signature of balancing selection may be weaker in partial asexuals than selfers if the rate of biparental reproduction is very small (Fig. 6), even though the reverse is expected at less extreme levels of uniparental reproduction because of the difference in effective rates of recombination ( $2\sigma r$  vs.  $\sigma r$ ).

Here we explored how background and balancing selection affect coalescence times but we did not attempt to examine selective sweeps. In sexual populations, beneficial alleles spreading to fixation can drastically reduce coalescence times and diversity levels at closely linked sites (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989; STEPHAN *et al.* 1992; BARTON 1998; HERMISSE and PENNINGS 2005) and are believed to play an important role in shaping genomic patterns of diversity, at least in some species (ELYASHIV *et al.* 2014; WILLIAMSON *et al.* 2014). As with other forms of linked selection, we expect that the signature of selective sweeps will be more dramatic in systems with high rates of uniparental reproduction because of their reduced effective rates of recombination. For example, in the highly selfing *C. elegans*, sweeps are believed to have reduced variation across entire chromosomes (ANDERSEN *et al.* 2012). As with background and balancing selection, we speculate the “recombination effect” will not be the whole story in partial asexuals. With low rates of sexes, a “segregation effect” is likely because, under asexuality, the selected diploid genotypes are inherited intact (rather than their haploid gametes), altering the trajectory of the selected allele from that in the standard sexual (or haploid)

model. The full consequences of partial asexuality with respect to selective sweeps remain a challenge for future work.

Though the consequences of sweeps are likely to be larger in systems with uniparental inheritance, sweeps are less likely to occur in such systems. Both theory (GLÉMIN and RONFORT 2012; KAMRAN-DISFANI and AGRAWAL 2014) and data (HOUGH *et al.* 2013; BURGARELLA *et al.* 2015) show that reductions in  $N_e$  due to background selection render selection less efficient, reducing the probability of fixation of beneficials, in selfers compared to outcrossers. A qualitatively similar effect is expected in partial asexuals.

A number of interesting and important diploid species reproduce predominantly asexually but these have been studied less by evolutionary geneticists than selfing species. Though there are similarities between partial asexuality and selfing, it is misleading to consider these two forms of uniparental inheritance as conceptually interchangeable. As population genomic data becomes easier to obtain, the hope is that such data will allow us to study the evolutionary consequences of different forms of uniparental reproduction and, ideally, gain insight into the evolutionary forces acting on reproductive mode. Understanding the coalescent properties of systems with uniparental inheritance is key to interpreting population genomic data and much work remains, especially with respect to partial asexuality.



810    **Acknowledgements:**

811    We thank Stephen I. Wright for extensive discussion and Denis Roze for sharing his  
 812    unpublished result on background selection in selfers. This work was supported by  
 813    the Natural Sciences and Research Council of Canada (AFA) and by a Marie Curie  
 814    International Outgoing Fellowship, grant number MC-IOF-622936 project SEXSEL  
 815    (MH).

816

## Literature Cited

- ANDERSEN E. C., GERKE J. P., SHAPIRO J. A., CRISSMAN J. R., GHOSH R., BLOOM J. S., FÉLIX M.-A., KRUGLYAK L., 2012 Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nature Genetics* **44**: 285–290.
- ANDOLFATTO P., NORDBOG M., 1998 The effect of gene conversion on intralocus associations. *Genetics* **148**: 1379-1399.
- ANDRÉS A. M., HUBISZ M. J., INDAP A., TORGERSON D. G., DEGENHARDT J. D., BOYKO A. R., GUTENKUNST R. N., WHITE T. J., GREEN E. D., BUSTAMANTE C. D., CLARK A. G., NIELSEN R., 2009 Targets of balancing selection in the human genome. *Mol Biol Evol* **26**: 2755–2764.
- ASHBURNER M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, New York.
- BARRETT S. C. H., ARUNKUMAR R., WRIGHT S. I., 2014 The demography and population genomics of evolutionary transitions to self-fertilization in plants. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **369**: 20130344–20130344.
- BARTON N. H., 1998 The effect of hitch-hiking on neutral genealogies. *Genet Res* **72**: 123–133.
- BENGTSOON B., 2003 Genetic variation in organisms with sexual and asexual reproduction. *J Evol Biol.* **16**: 189-199.
- BRANDVAIN Y., SLOTT T., HAZZOURI K. M., WRIGHT S. I., COOP G., 2013 Genomic Identification of Founding Haplotypes Reveals the History of the Selfing Species *Capsella rubella*. *PLoS Genet* **9**: e1003754.
- BURGARELLA C., GAYRAL P., BALLENGHIEN M., BERNARD A., DAVID P., JARNE P., CORREA A., HURTREZ-BOUSSÈS S., ESCOBAR J., GALTIER N., GLÉMIN S., 2015 Molecular evolution of freshwater snails with contrasting mating systems. *Mol Biol Evol*: msv121.
- CEPLITIS A., 2003 Coalescence times and the Meselson effect in asexual eukaryotes. *Genet Res.* **82**: 183-190
- CHARLESWORTH B., 2012 The effects of deleterious mutations on evolution at linked sites. *Genetics* **190**: 5–22.
- CHARLESWORTH B., CAMPOS J. L., 2014 The relations between recombination rate and patterns of molecular variation and evolution in *Drosophila*. *Annu Rev Genet* **48**: 383–403.
- CHARLESWORTH B., MORGAN M. T., CHARLESWORTH D., 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289-1303.

852 CHARLESWORTH D., CHARLESWORTH B., MORGAN M. T., 1995 The pattern of neutral  
853 molecular variation under the background selection model. *Genetics* **141**: 1619-  
854 1632.

855 ELYASHIV E., SATTATH S., HU T. T., STRUSTOVSKY A., McVICKER G., ANDOLFATTO P., COOP G.,  
856 SELLA G., 2014 A genomic map of the effects of linked selection in *Drosophila*.  
857 arXiv **q-bio.PE**.

858 GLÉMIN S., RONFORT J., 2012 Adaptation and maladaptation in selfing and outcrossing  
859 species: new mutations versus standing variation. *Evolution* **67**: 225-240.

860 HARTFIELD M., WRIGHT S. I., AGRAWAL A. F., 2015 Coalescent times and patterns of  
861 genetic diversity in species with facultative sex: effects of gene conversion,  
862 population structure and heterogeneity. *bioRxiv*.

863 HERMISSON J., PENNINGS P., 2005 Soft sweeps molecular population genetics of  
864 adaptation from standing genetic variation. *Genetics* **169**: 2335-2352.

865 HEY J., 1991 A multi-dimensional coalescent process applied to multi-allelic selection  
866 models and migration models. *Theor Popul Biol* **39**: 30-48.

867 HOUGH J., WILLIAMSON R. J., WRIGHT S. I., 2013 Patterns of selection in plant genomes.  
868 *Annu Rev Ecol Evol Syst* **44**: 31-49.

869 HUDSON R. R., KAPLAN N. L., 1988 The coalescent process in models with selection and  
870 recombination. *Genetics* **120**: 831-840.

871 HUDSON R. R., KAPLAN N. L., 1994 Gene trees with background selection. In: Golding GB  
872 (Ed.), *Non-neutral Evolution*, New York, pp. 140-153.

873 HUDSON R. R., KAPLAN N. L., 1995 The coalescent process and background selection.  
874 *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **349**: 19-23.

875 HUGHES A. L., NEI M., 1988 Pattern of nucleotide substitution at major  
876 histocompatibility complex class I loci reveals overdominant selection. *Nature*  
877 **335**: 167-170.

878 KAISER V. B., CHARLESWORTH B., 2009 The effects of deleterious mutations on  
879 evolution in non-recombining genomes. *Trends Genet.* **25**: 9-12.

880 KAMAU E., CHARLESWORTH D., 2005 Balancing selection and low recombination affect  
881 diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*. *Curr*  
882 *Biol* **15**: 1773-1778.

883 KAMRAN-DISFANI A., AGRAWAL A. F., 2014 Selfing, adaptation and background selection  
884 in finite populations. *J Evol Biol* **27**: 1360-1371.

885 KAPLAN N. L., DARDEN T., HUDSON R. R., 1988 The coalescent process in models with

886 selection. *Genetics* **120**: 819–829.

887 KAPLAN N. L., HUDSON R. R., LANGLEY C. H., 1989 The “hitchhiking effect” revisited.  
888 *Genetics* **123**: 887–899.

889 KEIGHTLEY P. D., OTTO S. P., 2006 Interference among deleterious mutations favours  
890 sex and recombination in finite populations. *Nature* **443**: 89–92.

891 LOEWE L., CHARLESWORTH B., 2007 Background selection in single genes may explain  
892 patterns of codon bias. *Genetics* **175**: 1381–1393.

893 LYNCH M., HILL W. G., 1986 Phenotypic Evolution by Neutral Mutation. *Evolution* **40**:  
894 915.

895 MAYNARD SMITH J., HAIGH J., 1974 The hitch-hiking effect of a favourable gene. *Genet*  
896 *Res* **23**: 23–35.

897 MCVICKER G., GORDON D., DAVIS C., GREEN P., 2009 Widespread genomic signatures of  
898 natural selection in hominid evolution (MW Nachman, Ed.). *PLoS Genet* **5**:  
899 e1000471.

900 NICOLAISEN L. E., DESAI M. M., 2012 Distortions in genealogies due to purifying  
901 selection. *Mol Biol Evol* **29**: 3589–3600.

902 NORDBORG M., 1997 Structured coalescent processes on different time scales.  
903 *Genetics* **146**: 1501–1514.

904 NORDBORG M., DONNELLY P., 1997 The coalescent process with selfing. *Genetics* **146**:  
905 1185–1195.

906 NORDBORG M., CHARLESWORTH B., CHARLESWORTH D., 1996a The effect of recombination  
907 on background selection. *Genetical Research* **67**: 159–174.

908 NORDBORG M., CHARLESWORTH B., CHARLESWORTH D., 1996b Increased levels of  
909 polymorphism surrounding selectively maintained sites in highly selfing species.  
910 *Proc R Soc Lond B* **263**: 1033–1039.

911 PRESTON C. R., ENGELS W. R., 1996 P-element-induced male recombination and gene  
912 conversion in *Drosophila*. *Genetics* **144**: 1611–1622.

913 RICHMAN A. D., UYENOYAMA M. K., KOHN J. R., 1996 Allelic diversity and gene genealogy  
914 at the self-incompatibility locus in the solanaceae. *Science* **273**: 1212–1216.

915 ROZE D., 2014 Selection for sex in finite populations. *J Evol Biol* **27**: 1304–1322.

916 SHARP N. P., AGRAWAL A.F. *in review*. Low genetic quality alters key dimensions of the  
917 mutational spectrum.

918 STEPHAN W., WIEHE T. H. E., LENZ M. W., 1992 The effect of strongly selected  
919 substitutions on neutral polymorphism: Analytical results based on diffusion  
920 theory. *Theor Popul Biol* **41**: 237–254.

921 WAKELEY J., 2008 *Coalescent Theory: An Introduction*. Roberts & Company,  
922 Greenwood Village, CO.

923 WILLIAMSON R. J., JOSEPHS E. B., PLATTS A. E., HAZZOURI K. M., HAUDRY A., BLANCHETTE M.,  
924 WRIGHT S. I., 2014 Evidence for widespread positive and negative selection in  
925 coding and conserved noncoding regions of *Capsella grandiflora*. *PLoS Genet* **10**:  
926 e1004622.

927 WIUF C., HEIN J., 2000 The coalescent with gene conversion. *Genetics* **155**: 451–462.

928 ZENG K., CHARLESWORTH B., 2011 The joint effects of background selection and genetic  
929 recombination on local gene genealogies. *Genetics* **189**: 251–266.

930