Rate of Adaptation in Sexuals and Asexuals: A Solvable Model of the Fisher–Muller Effect

Su-Chan Park*†1 and Joachim Krug†

*Department of Physics, The Catholic University of Korea, Bucheon 420-743, Republic of Korea, and †Institut für Theoretische Physik, Universität zu Köln, 50937 Köln, Germany

ABSTRACT The adaptation of large asexual populations is hampered by the competition between independently arising beneficial mutations in different individuals, which is known as clonal interference. In classic work, Fisher and Muller proposed that recombination provides an evolutionary advantage in large populations by alleviating this competition. Based on recent progress in quantifying the speed of adaptation in asexual populations undergoing clonal interference, we present a detailed analysis of the Fisher–Muller mechanism for a model genome consisting of two loci with an infinite number of beneficial alleles each and multiplicative (nonepistatic) fitness effects. We solve the deterministic, infinite population dynamics exactly and show that, for a particular, natural mutation scheme, the speed of adaptation in sexuals is twice as large as in asexuals. This result is argued to hold for any nonzero value of the rate of recombination. Guided by the infinite population result and by previous work on asexual adaptation, we postulate an expression for the speed of adaptation in finite sexual populations that agrees with numerical simulations over a wide range of population sizes and recombination rates. The ratio of the sexual to asexual adaptation speed is a function of population size that increases in the clonal interference regime and approaches 2 for extremely large populations. The simulations also show that the imbalance between the numbers of accumulated mutations at the two loci is strongly suppressed even by a small amount of recombination. The generalization of the model to an arbitrary number \( L \) of loci is briefly discussed. If each offspring samples the alleles at each locus from the gene pool of the whole population rather than from two parents, the ratio of the sexual to asexual adaptation speed is approximately equal to \( L \) in large populations. A possible realization of this scenario is the reassortment of genetic material in RNA viruses with \( L \) genomic segments.

Copyright © 2013 by the Genetics Society of America
doi: 10.1534/genetics.113.155135
Manuscript received July 8, 2013; accepted for publication August 7, 2013
Supporting information is available online at http://www.genetics.org/lookup/suppl/di0.1534/genetics.113.155135/-/DC1.
1Corresponding author: Department of Physics, The Catholic University of Korea, 43 Jibong-ro, Wonmi-gu, Bucheon 420-743, Republic of Korea.
E-mail: spark0@catholic.ac.kr

The evolutionary advantage of sex remains one of the most intriguing puzzles in evolutionary biology (Kondrashov 1993; de Visser and Elena 2007; Otto 2009). Many hypotheses explaining why sexual reproduction is widespread in nature despite apparent disadvantages such as the twofold cost of sex have been suggested (Maynard Smith 1978). Well-known examples are the deterministic mutation hypothesis (Kondrashov 1988), the Fisher–Muller (FM) mechanism (Fisher 1930; Muller 1932; Crow and Kimura 1965), and Muller’s ratchet (Muller 1964; Felsenstein 1974), to name only a few. These three hypotheses are applicable when the fitness landscape in question has certain specific features. Specifically, the deterministic mutation hypothesis requires deleterious mutations to be synergistically epistatic, while the Fisher–Muller mechanism as well as Muller’s ratchet can explain the advantage of sex if epistasis is negligible.

Theoretical analyses of the effect of epistasis on the speed of Muller’s ratchet have concluded that it practically stops operating when epistasis is synergistic (Charlesworth et al. 1993; Kondrashov 1994; Jain 2008). Furthermore, recent experimental analyses of empirical fitness landscapes seem to indicate that a particularly strong form of epistasis termed sign epistasis (Weinreich et al. 2005) is quite common (Weinreich et al. 2006; de Visser et al. 2009; Franke et al. 2011; Szendro et al. 2013b). Sign epistasis generally implies that the fitness landscape is rugged. On a rugged fitness landscape, sex can be detrimental, even without taking into account the twofold cost of sex, in that sexual populations, unlike the corresponding asexual populations, cannot escape from local fitness peaks (Crow and Kimura 1965; Eshel and Feldman 1970; de Visser et al. 2009; Park and Krug 2011).

Although research on empirical fitness landscapes has been growing substantially in recent years, it is still practically
infeasible to reliably determine genotypic fitness on a genome-wide scale (but see Koyos et al. 2012). Because of the small sizes of most empirical fitness landscapes that have so far been constructed experimentally, the implications of sign epistasis for long-term evolution remain unclear. At the same time experimental evidence in favor of the FM mechanism has also accumulated (Colegrave 2002; Cooper 2007). For these reasons further quantitative analysis of the advantage of sex in the absence of epistasis remains a worthwhile endeavor, and we pursue this approach in the present contribution.

The essence of the FM mechanism is the competition between independently arising beneficial mutations, termed clonal interference, which slows down the adaptation of large asexual populations (Gerrish and Lenski 1998; Miralles et al. 1999; Wilke 2004; Kim and Orr 2005; Park and Krug 2007; Fogle et al. 2008; Sniegowski and Gerrish 2010; Schiffels et al. 2011). The concept of clonal interference has played an important role in interpreting the behavior observed in laboratory selection experiments (Lenski et al. 1991; Lenski and Travisano 1994; Barrick et al. 2009) and has also been invoked in explaining the population-size dependence of evolutionary predictability in rugged fitness landscapes (Jain et al. 2011; Szendro et al. 2013a). Although in its original formulation clonal interference theory neglects the occurrence of secondary beneficial mutations within a growing clone (Gerrish and Lenski 1998; Gerrish 2001), in general the coexistence of multiple beneficial mutations cannot be neglected in large populations (Park and Krug 2007). In the following we therefore use the term clonal interference in a wider sense than originally conceived, in that two clones with different numbers of beneficial mutations can compete with each other for fixation.

Much recent theoretical work has focused on obtaining accurate quantitative estimates of the speed of adaptation in the presence of clonal interference for the simple situation of an unlimited supply of beneficial mutations that act independently on fitness, without epistatic interactions (see Park et al. 2010 for review). It turns out that the population dynamics in this regime is well described by a traveling wave moving at constant speed along a one-dimensional fitness space. The traveling wave picture was first established for the case when the beneficial selection coefficient is the same for all mutations (Tsimring et al. 1996; Rouzine et al. 2003; Desai and Fisher 2007; Brunet et al. 2008; Rouzine et al. 2008) and recently extended to the more realistic case of selection coefficients drawn from a continuous effect size distribution (Good et al. 2012; Fisher 2013); see also Bürger (1999) and references therein for a traveling wave picture of adaptation in changing environments.

In natural populations it is unlikely that the traveling wave picture persists forever. Apart from the assumed absence of epistatic interactions, two main features lead to a breakdown of this picture in long-term evolution. First, a fluctuating environment generally makes the fitness landscape change with time. In a time-dependent situation it is problematic to compare absolute fitnesses of two individuals living on different landscapes and, accordingly, adaptation is measured through the relative fitness increase or its time-integrated form termed fitness flux (Mustonen and Lässig 2010). Second, even if the fitness landscape remains constant for a very long time, the indefinite supply of beneficial mutations appearing at constant rate cannot be a good approximation in the real world. For example, in long-term evolution experiments the speed of adaptation usually slows down (Lenski and Travisano 1994; Barrick et al. 2009), which is attributed to the decreasing supply of beneficial mutations. In this context, the house-of-cards model, in which fitness values are assigned randomly to genotypes, could provide a more realistic description (Kingman 1978; Park and Krug 2008). In the framework of this model one cannot, however, explain the advantage of sex, because the fitness of a recombinant genotype is uncorrelated with the parental fitnesses and therefore beneficial mutations cannot accumulate through recombination.

Although the nonepistatic model with an infinite supply of beneficial mutations is of limited validity, it can provide a reasonable approximation when a population undergoes a severe environmental change, as is often the case at the beginning of an evolution experiment. At the same time this setting is conceptually simple and allows for detailed (if approximate) mathematical analysis. In this article, we therefore build upon the recent line of work on asexual populations undergoing clonal interference and add to it a minimal yet realistic recombination scheme. Specifically, we consider a sexual population model with two genetic loci, each of which can acquire infinitely many beneficial mutations. For simplicity we assume that epistasis is absent both between and within loci. Upon reproduction, the offspring receives one locus from each parent with probability $r$ and both loci from a single parent with probability $1 - r$. A possible biological realization of this kind of facultatively sexual reproduction is the assortment of genetic material in RNA viruses with two genomic segments, where the parameter $r$ reflects the probability of co-infection and is governed by the multiplicity of infection, the ratio of viruses to the number of infected cells (Simon-Loriere and Holmes 2011). In this context it is natural to consider the generalization of the model to $L$ loci, which is described in Discussion.

We first analyze the infinite population dynamics of the two-locus model, obtaining exact expressions for the speed of adaptation in the limiting cases of zero and maximal recombination rates (asexuals vs. obligate sexuals). When the selection coefficient of beneficial mutations is the same at both loci and at most one mutation may occur per generation and individual, the speed of adaptation for obligate sexuals is twice that of asexuals, a result that we argue holds for any positive recombination rate. Based on this observation we conjecture that for finite populations the speed of adaptation in sexuals is approximately equal to the sum of the speeds of the two loci, each of which receives half of the supply of beneficial mutations. Denoting the speed of adaptation by $v_s$ for sexuals and by $v_a$ for asexuals, and the
some of the assumptions of our model, describing in particular. In comparison of Equation 1 to rate for very large populations (Park et al. 2010), there is a twofold advantage of sex in this regime. Second, the precise theoretical estimates for the speed of adaptation in asexuals that have been developed in recent work translate through Equation 1 into explicit expressions for the sexual speed of adaptation in our model. In Discussion we address the consequences of relaxing some of the assumptions of our model, describing in particular a possible extension of the model to more than two loci, and place our work into the context of related studies.

Models

We consider a sexual or asexual population of haploid individuals in discrete generations. The population size is denoted by N and assumed to be constant. As a reproduction scheme we employ the Wright–Fisher model (Fisher 1930; Wright 1931), the prototypical model of discrete, nonoverlapping generations. Since our main concern is how recombination affects the speed of adaptation, we assume that all mutations are beneficial. This naturally leads us to study evolution in the framework of the infinite-sites model (Kimura 1969); otherwise back mutations of beneficial mutations, which are deleterious by definition, should appear with nonzero probability. Furthermore, we assume no epistasis among mutations, which is reflected by the multiplicative fitness assignment. As a minimal model with the above properties, we study an evolving population with only two loci under selection. Each locus is assumed to have infinitely many sites. We assume an initially homogenous population and the fitness of the initial genome, or wild type, is set to unity.

In line with the assumption of multiplicative fitness effects, the fitness of an individual that has n1 mutations at locus i compared to the wild type is \( \exp(n_1 s_1 + n_2 s_2) \). Without loss of generality we take \( s_2 \geq s_1 \). Note that two genotypes with the same number of mutations at each locus are not necessarily the same although both have the same fitness \( \exp(s_1 n_1 + s_2 n_2) \). Since we are interested only in how fast mean fitness increases and not in the genealogy, all genotypes with the same number of mutations at each locus are treated as if they were the same.

The population evolves in the following way. Let \( f_t(n_1, n_2) \) denote the frequency of all genotypes with \( n_1 \) mutations at the first locus and \( n_2 \) mutations at the second locus at generation \( t \). At \( t = 0 \), the population is homogeneous with \( f_0(0, 0) = 1 \). By selection, the frequency at generation \( t + 1 \) on average will change to be

\[
f_{t+1}(n_1, n_2) = \frac{e^{(r(n_1+s_1)n_1+s_2n_2)} f_t(n_1, n_2)}{\bar{w}_t},
\]

where

\[
\bar{w}_t = \sum_{n_1, n_2} e^{f_t(n_1+s_1n_1+s_2n_2)} f_t(n_1, n_2)
\]

is the average fitness of the population at generation \( t \).

Mutation can also change the frequency of genotypes. The probability that an offspring is hit by \( m_1 \) mutations at the first locus and \( m_2 \) mutations at the second locus is denoted by \( g_0(m_1, m_2) \). Here we implicitly assume that the mutation probability is not affected by the genetic background. To be concrete, \( g_0(0, 0) \) is the probability that neither locus is mutated, \( g_0(1, 0) \) is the probability that a mutation occurs at the first locus, but not at the second, and so on. In most of our analysis, we assume that \( g_0(0, 0) + g_0(1, 0) + g_0(0, 1) = 1 \), which reflects that only single-site mutations can occur. The frequency change due to both selection and mutation is

\[
f_{t+1}(n_1, n_2) = \sum_{m_1, m_2 \geq 0} g_0(n_1 - m_1, n_2 - m_2) f_t(m_1, m_2)
\]

\[
= \sum_{m_1, m_2 \geq 0} g_0(n_1 - m_1, n_2 - m_2) \frac{e^{(r(n_1+s_1)n_1+s_2n_2)}}{\bar{w}_t}
\]

\[
\times f_t(m_1, m_2),
\]

where \( g_0(x, y) \) with at least one negative argument should be understood to be 0. We further assume that mutation does not have any preference for a certain locus; that is, \( g_0(m_1, m_2) = g_0(m_2, m_1) \) for any pair of \( m_1 \) and \( m_2 \).

After selection and mutation, two randomly chosen parents mate and beget an offspring. Let \( R(n_1, n_2|k_1, k_2; l_1, l_2) \) denote the probability that the resulting progeny of two individuals with respective genotypes \( (k_1, k_2) \) and \( (l_1, l_2) \) has the genotype \( (n_1, n_2) \). To be specific, we set \( 0 \leq r < 1 \)

\[
R(n_1, n_2|k_1, k_2; l_1, l_2) = \begin{cases} (1-r)/2 & \text{if } n_1 = k_1, n_2 = k_2, \\ (1-r)/2 & \text{if } n_1 = l_1, n_2 = l_2, \\ r/2 & \text{if } n_1 = k_1, n_2 = l_2, \\ r/2 & \text{if } n_1 = l_1, n_2 = k_2, \\ 0 & \text{otherwise,} \end{cases}
\]

which means that with probability \( 1 - r \) the two loci of the offspring in question are inherited solely from a single parent, which is selected with probability \( 1/2 \) and with probability \( r \) the offspring inherits one locus from one parent and the other from the other parent. When \( r = 0 \), an offspring inherits all genotypes from a single parent, so we call the case with \( r = 0 \) asexuals. On the other hand, when \( r = 1 \), an offspring inherits alleles from both parents, so we call the
case with \( r = 1 \) obligate sexuals. In this sense, the case with \( 0 < r < 1 \) can be regarded as facultatively sexual populations.

Since the probability that the randomly chosen parents have genotypes \((k_1, k_2)\) and \((l_1, l_2)\) is \(f_t^{k_1 l_1}(k_1, k_2)f_t^{l_1 l_2}(l_1, l_2)\), the mean frequency after selection, mutation, and recombination is

\[
f_t^{11}(n_1, n_2) = \sum_{k_1, k_2, l_1, l_2} R(n_1, n_2|k_1, k_2; l_1, l_2)f_t^{k_1 l_1}(k_1, k_2)f_t^{l_1 l_2}(l_1, l_2) = (1 - r)f_t^{n_1}(n_1, n_2) + f_t^{s(1)}(n_1)f_t^{s(2)}(n_2),
\]

where

\[
f_t^{s(1)}(n_1) = \sum_{n_2} f_t^{n_1}(n_1, n_2), \quad f_t^{s(2)}(n_2) = \sum_{n_1} f_t^{n_2}(n_1, n_2),
\]

are marginal frequency distributions of genotypes after the selection and mutation steps with \( n_1 \) mutations at locus 1 and \( n_2 \) mutations at locus 2, respectively.

Finally, the actual population distribution at generation \( t + 1 \) is determined by multinomial sampling using \( f_t^{s(1)}(n_1)f_t^{s(2)}(n_2) \) in Equation 6 with the restriction that the population size is \( N \). For simulations, we employ the algorithm explained by Park and Krug (2007) (see also Park et al. 2010 for simulations of extremely large populations).

The speed of adaptation, or shortly speed, is defined as the rate of increase of the log mean fitness,

\[
v = \lim_{t \to \infty} \left( \frac{\ln w_t}{t} \right),
\]

where \((\ldots)\) denotes an average over independent realizations of evolution with the same parameters. In the following, we mainly focus on the dependence of speed on parameters such as the population size, the mutation probability per generation, the selection coefficient of a single mutation, and the recombination probability.

Results

Infinite populations

Although the infinite population limit cannot be reached in real biological populations for the model we are studying (Park et al. 2010), it does provide some insight into the adaptation dynamics of finite populations. Furthermore, the deterministic nature of the infinite population dynamics renders an analytic approach feasible. We therefore begin our discussion with the evolutionary dynamics of infinite populations. Detailed derivations and generalizations of the results presented here can be found in Appendix A.

As shown in Appendix A, the advantage of sex in infinite populations depends on the exact form of mutation probability distribution \( g_0 \). However, as demonstrated later in Discussion, the form of \( g_0 \) does not affect the speed of populations with biologically relevant size as long as the mutation probability is small. In the following we employ the simple mutation scheme

\[
g_0(0, 0) = 1 - U, \quad g_0(1, 0) = g_0(0, 1) = U/2,
\]

which does not allow for multiple-site mutations. In this case, the speed for asexuals, \( v_a \), and for obligate sexuals, \( v_s \), are found to be (see Appendix A)

\[
v_a = \max(s_1, s_2) = s_2, \quad v_s = s_1 + s_2.
\]

This result can be understood as follows: for obligate sexuals \((r = 1)\), the two loci are unlinked and, thus, each locus evolves independently with mutation probability \( U/2 \). Since, regardless of the actual value of \( U \neq 0 \), the contributions from each locus are \( s_1 \) and \( s_2 \), respectively, the total speed \( v_s \) is the sum of these two. For asexuals, clonal interference prohibits accumulation of the weaker beneficial effect \( s_1 \), so the speed is determined solely by the larger beneficial effect \( s_2 \). Equation 10 is also valid when \( s_1 = s_2 \). In this case, \( v_s \) is twice as large as \( v_a \), that is, a twofold advantage of sex, which is the maximum effect of sex in the two-locus model. When we study the adaptation dynamics of finite populations, we set \( s_1 = s_2 = s \) to maximize the advantage of sex.

Although we found the speed exactly only for the cases \( r = 0 \) and \( r = 1 \), we now argue that the asymptotic speed does not depend on \( r \) provided \( r > 0 \) for any mutation scheme. Let \( \ell_1(t) \) \((\ell_2(t)) \) denote the maximum number of mutations at locus 1 \((\text{locus 2})\) accumulated up to generation \( t \):

\[
\ell_1(t) = \max \left\{ n \left| \sum_{m} f_t(n, m) \neq 0 \right. \right\}, \quad \ell_2(t) = \max \left\{ m \left| \sum_{n} f_t(n, m) \neq 0 \right. \right\}.
\]

This definition can be used for finite populations as well and is closely related to the lead of the fitness distribution considered in the traveling wave approach to asexual adaptation (Desai and Fisher 2007; Park et al. 2010; Fisher 2013). Within our general mutation scheme with homogeneous initial conditions, \( \ell_2(t) = M t \) for infinite populations, where \( M \) is the largest possible number of sites that can be mutated at one locus in a single mutation event. Hence the frequency \( f_t(M t, M t) \) of genotypes with \( M t \) mutations at each locus at generation \( t \) is nonzero due to recombination, although it can be extremely small.

Now assume that the speed \( v_s(r = \infty) \) for \( 0 < r < 1 \) is strictly smaller than \( M(s_1 + s_2) \). Then, with time \( t \), the ratio of the detectable largest fitness to the mean fitness increases as \( \exp((M(s_1 + s_2) - v_s) t) \). Thus, at some \( t \), the relative fitness of the genotypes with \( M t \) mutations at each locus to the mean fitness becomes extremely large, which eventually results in an abrupt increase of frequency of these genotypes in one generation. Accordingly, \( \bar{w}_t \) becomes of the order of \( \exp(M(s_1 + s_2) t) \), and in the long run the speed becomes \( M(s_1 + s_2) \) for any \( r > 0 \).

In the above discussion, we argued that the speed does not depend on \( r \) once \( r \) is nonzero. On the other hand, if \( r \) is
very small, the whole population behaves almost like an asexual population for quite some time. Hence, the abrupt jump of fitness mentioned above should be observable. To see this phenomenon, we studied the deterministic evolution numerically, using the mutation scheme of Equation 9 with \( U = 0.1 \) and \( s_1 = s_2 = 0.02 \). In Figure 1, we show how the mean fitness behaves with time for \( r = 0, r = 10^{-9} \), and \( r = 1 \). Even for the minute recombination rate of \( r = 10^{-9} \), the mean fitness closely follows the \( r = 1 \) curve, however, with some oscillations. To elucidate the origin of this behavior we need to consider how the frequency distribution changes with time.

In the asexual case \((r = 0)\) the frequency distribution over the number of mutations is well described by a Gaussian (Park et al. 2010). Furthermore, the frequency distribution of the obligately sexual population with \( r = 1 \) should also be well described by a Gaussian, because the generating function is just the product of two generating functions of asexual evolution (see Equation A10). However, for \( 0 < r < 1 \), the Gaussian may not be a good approximation. In Figure 2 we depict the time evolution of the frequency distribution for \( r = 10^{-9} \). Clearly the frequency distribution cannot be approximated by a Gaussian traveling wave. Moreover, the shape of the distribution changes with time, which implies that there is no time-independent steady state. Rather, the distribution behaves like a “breathing traveling wave” in that the behavior seen in Figure 2 repeats periodically. In Supporting Information, File S1, one can find an animation showing the breathing traveling wave. The time when two peaks become comparable in Figure 2 corresponds to the abrupt jump of mean fitness alluded to above. Further mathematical analysis of this phenomenon seems interesting, but we do not pursue it here because it is hardly observable in real, finite populations.

**Finite populations**

We mentioned before that many analytic approaches have been developed to find an expression for the speed of adaptation in large asexual populations (Rouzine et al. 2003; Desai and Fisher 2007; Brunet et al. 2008; Rouzine et al. 2008). Park et al. (2010) summarized these developments and compared simulation results with the proposed analytic expressions. The approximation of Rouzine et al. (2008) turned out to be quite accurate in a wide range of parameters. The only disadvantage of this approach is that the speed is obtained as an implicit function of \( N \) (see below). In this section, we find a mathematical formula for the speed of adaptation in sexual populations, using both the suggested formula for asexuals and the results for the infinite population dynamics in the previous section.

For an infinite population, as shown in Appendix A, the precise form of the mutational probability distribution \( g_0(k_1, k_2) \) affects the speed. However, for plausible values of the mutation rate and the selection coefficient such infinite population effects become observable only for unrealistically large populations (Park et al. 2010); see Discussion for a detailed argument. In the following we therefore use Equation 9 and set \( s_1 = s_2 = s \) for the reasons mentioned previously. This implies that at most one mutation can occur per individual in each generation, and all mutations have the same selective effect \( s \).

We begin with a discussion of the speed for asexual populations. As was illustrated by Park et al. (2010), the speed for the asexual version of our model \((r = 0)\) is well approximated by the implicit equation

\[
\ln N \approx \frac{v_{RBW}}{2s^2} \left( \ln v_{RBW} + 1 \right) - \ln \frac{s^3 U}{v_{RBW} \ln(v_{RBW}/(Us))},
\]

where the subscript \( a \) in \( v_a \) refers to the asexual population, the superscript RBW refers to the authors of Rouzine et al. (2008), and \( e \approx 2.718182 \) is the base of the natural logarithm. Since the approximation of the fitness distribution by a continuous traveling wave was used to derive Equation 12, it should not be surprising that the discrepancy between theory and simulation becomes relatively large when the size of population is small enough to realize the strong-selection weak-mutation (SSWM) regime, where the population is mostly monomorphic. Based on this observation,
there is room for improvement of the approximation in an ad hoc way as follows: First we note that the first term in Equation 12 is dominant when the speed is high and the second term is dominant when the speed is low. Thus, when the population size is small, we can neglect the first term. In the SSWM regime, two consecutive fixations of beneficial mutations can be considered independent, so the speed can be estimated as the mean number of fixations per generation times the selection coefficient of the fixed mutation. Since the fixation probability of a beneficial mutation with selection coefficient $s$ is $\sim 2s$ and all beneficial mutations have the same effect in our model, the speed in the SSWM regime is $v_a = NU \times 2s \times s = 2NUs^2$. Using the speed in the SSWM regime, we modify Equation 12 as

$$\ln N \approx \frac{v_a}{2s^2} \left( \ln \frac{v_a}{eUs} + 1 \right) + \ln \frac{v_a}{2s^2U},$$

(13)

which keeps the large speed behavior unchanged and enforces the SSWM result for small speeds. In Figure 3 we show that Equation 13 provides a more accurate approximation to the speed obtained from simulations with $U = 10^{-6}$ and $s = 0.01$ than does Equation 12.

Now we move on to the speed of sexual populations. At first, let us start from the case of $r = 1$ whose infinite population limit allows for an exact solution. As we show in Appendix A, the evolutionary dynamics of an infinite population with $r = 1$ can be viewed as the independent evolution of each locus with the marginal mutation probability $g_0(k)$. That is, we can divide the evolutionary dynamics into two independent asexual populations with reduced mutation probability and the speed of the sexual population is obtained by simply adding the speeds of these two virtual asexual populations. Within the mutation scheme given by Equation 9 with selection coefficients $s_1 = s_2 = s$ this implies that

$$v_s(r = 1, U) = 2v_a(U/2)$$

(14)

for sufficiently large populations. Interestingly, Equation 14 is trivially valid in the SSWM regime where the speed is linear in $U$ and $v(r, U) \approx 2NUs^2$ irrespective of $r$. Since Equation 14 accurately estimates the speed for very small and very large populations, it is likely that Equation 14 is a good approximation for any population size. Indeed, as we show in Figure 4, $v_s(r = 1, U)$ is well approximated by twice $v_a(U/2)$ for any population size. The parameters we have used in these simulations are $U = 10^{-6}$ and $s = 0.01$. With the help of Equation 13, we may thus approximate the speed $v_s$ of the obligately sexual population as

$$\ln N \approx \frac{v_s}{4s^2} \left( \ln \frac{v_s}{eUs} + 1 \right) + \ln \frac{v_s}{2s^2U}.$$  

(15)

It is clear that for $0 < r \ll 1$ there should be a regime where Equation 14 cannot approximate the speed accurately. To see this deviation, we simulated populations with various $r$ (Figure 4). It turns out that Equation 14 is still a good approximation for $r \approx 10^{-2}$. In particular, the speed for $r = 0.1$ is hardly discernible from that for $r = 1$ for all population sizes. The deviation starts to be significant for $r = 10^{-3}$. For comparison, we also plot $v_s(r = 0, U)/v_a(U/2)$ or equivalently $v_s(U)/v_a(U/2)$ in Figure 4, which should approach 1 in the infinite population limit (Park et al. 2010). For $N \approx 10^6$, where $NU \ln(Ns)$ becomes larger than 1, $v_s(r, U)/v_a(U/2)$ starts to increase, although very slowly, and $v_s(r, U)$ becomes significantly larger than $v_a(U)$. Note that for asexual populations clonal interference sets in around $NU \ln(Ns) \sim 1$ (Wilke 2004; Park et al. 2010). That is, as soon as clonal interference becomes relevant, even a small amount of recombination leads to a significant speedup of adaptation, in agreement with the FM mechanism.

To display the FM effect more clearly, we depict $v_s(r, U)/v_a(U)$ versus $N$ in Figure 5. The fact that the ratio $v_s(r, U)/v_a(U)$ continues to rise monotonically with $N$ for all cases with $r > 0$ in Figure 4 and Figure 5 is consistent with the twofold advantage predicted by the infinite population analysis.
When measuring the speed of adaptation in our simulations, a useful consistency check was provided by the Guess relation

$$v = Us + \left( \frac{1}{N} \sum_{i=1}^{N} (\chi_i - 1) \ln \chi_i \right)_{\text{stat}},$$  \hspace{1cm} (16)

where $\chi_i$ is the relative fitness of the $i$th individual in the infinite time limit,

$$\chi_i = \lim_{t \to \infty} \frac{w_i(t)}{w(t)},$$  \hspace{1cm} (17)

and $\langle \cdot \rangle_{\text{stat}}$ signifies an average over the stationary measure of $\chi_i$. Equation 16 was originally established for asexual populations undergoing discrete generation Wright-Fisher dynamics (Guess 1974a,b). In Appendix B, we prove that the relation holds for sexuals as well, and in Figure 6, we numerically confirm its validity. The two terms on the righthand side of Equation 16 represent the increase in population fitness due to mutation and selection, respectively. Recombination affects the speed of adaptation only indirectly through its effect on the relative fitnesses $\chi_i$. Note that the Guess relation should hold even if one uses discrete-time, overlapping generation models such as the Moran model.

Finally, we analyze the difference in the number of beneficial mutations acquired by the two loci. We quantify this difference as

$$V = \lim_{t \to \infty} \frac{\langle (\ell_1(t) - \ell_2(t))^2 \rangle}{t},$$  \hspace{1cm} (18)

where $\ell_1$ and $\ell_2$ are defined in Equation 11. We refer to $V$ as the mutation number imbalance (MNI). To discern the MNI of asexuals from that of sexuals, we add subscripts $a$ and $s$, for asexual and sexual populations, respectively. In infinite populations each locus accumulates the same number of mutations; hence this study is meaningful only for finite populations.

For asexual populations, an approximation for $V$ can be obtained by comparing the origination processes at the two loci, which count the mutations that are present in some individuals of the population at time $t$ and that are destined to eventually go to fixation (Gillespie 1993, 1994; Park and Krug 2007). Denoting the number of such mutations at locus $i$ by $k_i(t)$, we assume that (1) the difference between $k_i(t)$ and the lead $\ell_i(t) \geq k_i(t)$ remains bounded in the long time limit, (2) the total number of mutations $M_i(t) = k_1 + k_2$ in the origination process increases at the same rate as the mean number of mutations, $M_i(t) \approx \nu_i t / s$ for large $t$, and (3) each new mutation appearing in the origination process chooses one of two loci with equal probability. Assumptions 1 and 2 reflect the existence of a steady state and have been verified in simulations (Park and Krug 2007), and assumption 3 is a consequence of the symmetry between the two loci. By assumption 3, the probability that there are $m_1$ mutations at locus 1 and $m_2 = M_i(t) - m_1$ mutations at locus 2 is given by

$$P(m_1, t) \approx \left( \frac{M_i(t)}{m_1} \right)^{1/2} \frac{M_i(t)}{2}.$$  \hspace{1cm} (19)

Since the mean of $m_1$ is $\langle m_1 \rangle = M_i/2$ and its variance is $\langle (m_1 - \langle m_1 \rangle)^2 \rangle = M_i/4$, we can calculate $V_a$ invoking the assumption 1, as

$$V_a \approx 4 \frac{\langle (M_i(t) - 2m_1)^2 \rangle}{t} = 4 \frac{\langle (m_1 - \langle m_1 \rangle)^2 \rangle}{s} \frac{V_a}{s}.$$  \hspace{1cm} (20)

In Figure 7, we compare $V_a$ to $\nu_a / s$ for $U = 10^{-6}$ and $s = 0.01$, which shows an excellent agreement.

Recombination changes the behavior of the MNI substantially. As can be seen in Figure 7, once the population size is in the regime of clonal interference the MNI decreases abruptly, and then remains almost constant for a wide range of population sizes. Even a small amount of recombination

---

**Figure 5** Ratio of sexual to asexual speed of adaptation, $\nu_a / \nu_s(U)$, as a function of population size $N$ on a semilogarithmic scale. Recombination rates are $r = 10^{-5}$, $10^{-4}$, $10^{-3}$, $10^{-2}$, $10^{-1}$, and 1 from bottom to top, and $U = 10^{-6}$ and $s = 0.01$ are used as before. As in Figure 4, the two data sets for $r = 1$ and $r = 0.1$ are hardly discernible.

**Figure 6** Numerical verification of the Guess relation. The figure compares the logarithmic mean fitness divided by generation time, $\langle \ln \bar{w} \rangle / t$, to the right-hand side of Equation 16. The data are obtained from simulations with $r = 10^{-3}$, $U = 10^{-6}$, and $s = 0.01$. For $t \to \infty$, both curves intersect the ordinate at the same point, which equals the asymptotic speed of adaptation.
efficiently equalizes any major fitness difference between the two loci by creating competitively superior recombinants in which both loci have high fitness.

### Discussion

The Fisher–Muller mechanism for the evolutionary advantage of sex is based on the slowing down of asexual adaptation due to clonal interference, which is alleviated by the recombination of high fitness genotypes. While much recent theoretical work has been devoted to quantifying the speed of adaptation in asexuals, the speedup that can be achieved through recombination has been explicitly addressed in only a few studies (see below). In the present article we take a step in this direction by providing a detailed analysis of a simple, yet biologically meaningful model in which recombination occurs between two loci, each of which can harbor an unlimited number of linked beneficial mutations. Our analysis shows that the advantage of sex becomes significant in the parameter regime where clonal interference plays an important role in asexual populations. In our two-locus model, the adaptation speed of sexual populations is about twice as large as that of the corresponding asexual populations for a wide range of recombination rate. In the remainder of this section we discuss the robustness of our results to relaxing some of the assumptions in our model, in particular the neglect of multiple and recurrent mutations. We then describe a possible extension of the model to L loci and discuss its relevance to the adaptation of RNA viruses with multiple genetic segments. Finally, we briefly compare our findings to related previous work.

### Multiple-site mutations

In most of the analysis and simulations presented above we have assumed that only single-site mutations can occur in an individual each generation. Since mutations are replication errors that may occur at multiple sites in an independent fashion, a more realistic assumption would be that the probability for n mutations to arise in one individual is of order $U^n$, where $U$ is the probability of a single-site mutation. In the following we argue that allowing for multiple-site mutations does not significantly affect our results for the speed of adaptation for any biologically plausible population size, provided $U$ is small.

In the SSWM regime, $NU \ll 1$, multiple-site mutations obviously cannot contribute to the adaptation dynamics and Equation 1 remains valid. On the other hand, in the infinite population limit the speed of adaptation strongly depends on the form of the mutation probability $g_0(m_1, m_2)$ (see Equations A16 and A17). To be concrete, we adopt the mutation scheme of Equation A19, which allows for mutations at up to two sites, with two-site mutations occurring with probability $U^2$. The infinite population analysis then predicts that $\nu_s = \nu_a$, hence Equation 1 must break down beyond some characteristic population size $N_c$.

![Figure 7](https://academic.oup.com/genetics/article/195/3/941/5935482)

**Figure 7** The mutation number imbalance (MNI) $V$ vs. population size $N$ for $r = 0$ (open inverted triangle), $10^{-5}$ (open square), $10^{-4}$ (solid triangle), $10^{-3}$ (open triangle), $10^{-2}$ (solid circle), $10^{-1}$ (open circle), and 1 (solid square) from top to bottom. Other parameter values are $U = 10^{-6}$ and $s = 0.01$. The symbols for $r \approx 10^{-4}$ essentially overlap. For comparison, the analytic prediction $V^* = \nu_s/\nu_a$ for asexuals is drawn as a line.

A first guess about $N_c$ invokes the criterion for the onset of clonal interference. Since clonal interference among single-site mutations becomes important when $NU \ln N \approx 1$ (Gerrish and Lenski 1998; Wilke 2004; Park et al. 2010), clonal interference among clones with two-site mutations would become important when $NU^2 \ln N \approx 1$. Thus, for $U = 10^{-6}$ as assumed in our simulations, the effect of multiple-site mutations should be observable for $N \gg 10^6$. To check the validity of this argument, we simulated the model for $N = 10^{20}$ and $U = 10^{-6}$ using the mutation scheme of Equation A19 and compare the results to those presented previously assuming that only single-site mutations are possible; see Figure 8. Contrary to the above expectation, no detectable difference is observed. In fact, we could not observe any significant difference even for $N = 10^{100}$, even though in that case about $10^{98}$ double mutants occur in every generation (results not shown).

The reason for the failure of the above criterion is that multiple-site mutations can affect the speed of adaptation only if they occur among the offspring of the fittest individuals in the population. Within the traveling wave picture of asexual adaptation, these individuals reside in the so-called stochastic edge, which governs the rate of advance of the entire population (Desai and Fisher 2007; Brunet et al. 2008; Rouzine et al. 2008; Good et al. 2012; Fisher 2013), while mutations occurring in the bulk of the traveling wave are wasted by clonal interference. If the total number of offspring of the stochastic edge class per generation is much smaller than $U^{-2}$, a mutant offspring of the edge class is most likely to have a single-site mutation and, accordingly, single-site mutations should play a dominant role in the advance of the stochastic edge.

To find $N_c$, consider a large asexual population such that the selection coefficient of the fittest class, $s$, relative to the mean fitness is large and loss of the stochastic edge by genetic drift is unlikely. If this is not the case, the edge almost always starts from a single individual with few offspring and, in turn, multiple-site mutations cannot affect the speed
for the reason given above. When only single-site mutations can occur, \( \bar{s} = \ln(1/U) \) for an infinite population and the frequency of individuals in this maximum fitness class with \( t \) mutations is of order \( \exp[-\ln^2 U/(2s)] \) (Park et al. 2010). Thus for a population with size \( N \approx \exp[\ln^2 U/(2s)] \) the fittest class is occupied by at least one individual at all times and the traveling wave reaches the deterministic speed limit \( v_a = s \); for this finite population \( \bar{s} \) is also \( \ln(1/U) \). The mean number of offspring of an individual in the fittest class is of order \( e^s = 1/U \), so that on average one of the offspring will gain an additional mutation, securing the advance of the wave at maximum speed. Correspondingly, when double mutations are allowed and occur at rate \( U^2 \), the number of individuals in the fittest class investigated above must be of order \( 1/U \) to ensure that one double mutant can be created with high probability from this class. We therefore conclude that multiple-site mutations will affect the speed of adaptation only if

\[
N \gtrsim N_c = U^{-1} \exp[\ln^2 U/(2s)].
\] (21)

Since \( N_c \approx 10^{4150} \) for \( U = 10^{-6} \) and \( s = 0.01 \), multiple site mutations cannot change the outcome for any biologically reasonable population size. This implies that, in contrast to the infinite population model, the dynamics of finite populations are remarkably robust with regard to changes in the mutation scheme.

Although the above conclusion has been arrived at only by analyzing asexual populations, multiple-site mutations in sexual populations cannot affect the speed for any biologically relevant population size because the fittest class of each locus still has a small number of individuals; see also Figure 8 for numerical support.

**Finite number of sites**

We next discuss the implications of relaxing our assumption that each of the two loci carries an infinite number of sites at which beneficial mutations can occur. If the number of sites is finite, there is a nonzero probability that the same site will be hit multiple times. Two cases must be distinguished. If a beneficial mutation that was previously lost by genetic drift or clonal interference arises a second time, its effect will not be different from that of a new mutation in the infinite sites model, and in that sense such recurrent mutations are already accounted for in our analysis. On the other hand, if a site at which a beneficial mutation has been fixed is hit again, it constitutes a deleterious mutation. As long as such events are rare, the deleterious mutations will quickly be purged by natural selection. However, in the long run this leads to a depletion of the (finite) supply of beneficial mutations and causes the rate of adaptation to slow down in sexuals as well as asexuals, a regime that is beyond the scope of our study.

When the number of sites is finite, the Fisher–Muller effect thus gives rise to a transient advantage of sex that has been studied quantitatively by Kim and Orr (2005). They find that the speedup due to recombination is maximal when all beneficial mutations have the same selective strength and becomes less pronounced when different mutations have different strengths. Within our infinite-sites model this aspect could be addressed by allowing for a distribution of mutational effects instead of a single selection coefficient \( s \).

**More than two loci**

It is natural to surmise that the factor of 2 arises in our model because we are considering two loci and that the speed increase should in general be proportional to the number of loci. Indeed, this turns out to be true if we use a “communal” recombination scheme where the gene of each locus is collected from the whole population rather than from the two parents, once the genome of the offspring is constructed by recombination. In a three-locus model with the communal recombination scheme, the frequency distribution of next generation is sampled from

\[
f_t^{(n)}(n_1, n_2, n_3) = \left(1 - r\right)f_t^{(0)}(n_1, n_2, n_3) + r f_t^{(1)}(n_1) f_t^{(2)}(n_2) f_t^{(3)}(n_3),
\] (22)

where \( f_t^{(i)}(n_i) \) is the marginal frequency distribution for having \( n_i \) mutations at locus \( i \) after the (deterministic) selection and mutation steps (compare to Models). It is a straightforward extension of the calculation in Appendix A to show that the infinite population dynamics for \( r = 1 \) is again divided into three independent evolutions of each locus with marginal mutation probabilities just as in the two-locus case. In general, if we consider a model system with \( L \) loci within the communal recombination scheme mentioned above, the evolution is a superposition of \( L \) independent evolutions at each locus, and there is an \( L \)-fold advantage of sex in the infinite population.

To see if this \( L \)-fold advantage persists for finite populations, we performed simulations of the three-locus model. As in the two-locus case, we expect that

---

Figure 8 Mean logarithmic fitness (\( \ln \bar{w}_t \)) vs. time \( t \) in the presence (symbols) and absence (lines) of two-site mutations. The mutation schemes employed in the two cases are given in Equation A19 and Equation 9, respectively. The population size is \( N = 10^{20} \) and the probability for a single mutation is \( U = 10^{-6} \). The two data sets are indistinguishable, which implies that multiple-site mutations do not play any role.
for sufficiently large \( r \). Indeed, we observe that the simulations are consistent with Equation 23 for a wide range of parameter values. In Figure 9, we depict \( v_s(r, U)/v_a(U/3) \) as a function of \( N \) for \( U = 1.5 \times 10^{-6} \) and \( s = 0.01 \) with varying \( r \). As in the two-locus model, the advantage of sex becomes significant when clonal interference is important in the corresponding asexual populations.

We also studied the mutation number imbalance in the three-locus model. We slightly modify the definition of the MNI as the difference between the maximum and minimum numbers of accumulated mutations at all loci, which reduces to the definition of Equation 18 for the two-locus model. In Figure 10, we depict the MNI for the three-locus model with \( U = 1.5 \times 10^{-7} \) and \( s = 0.01 \) for various \( r \). Like the MNI of the two-locus model, the MNI for the asexuals increases with \( N \) while slowly decreasing for sexuals. Again, the qualitative difference between sexuals and asexuals becomes significant in the regime where clonal interference is important.

**Genetic reassortment in RNA viruses**

The communal recombination scheme described above arises naturally in RNA viruses with \( L \) genomic segments, which are reassorted during the co-infection of a single cell by several viruses (Simon-Loriere and Holmes 2011). Since the degree of reassortment can be controlled via the multiplicity of infection, this class of systems offers the opportunity to test hypotheses concerning the evolutionary advantage of recombination through the direct comparison between sexual and asexual populations (Chao 1900; Miralles et al. 1999; Poon and Chao 2004). Of particular interest in the context of our work is a study by Turner and Chao (1998), which aimed to test the Fisher–Muller mechanism by measuring the rate of fitness increase for the \( \phi 6 \) bacteriophage in the presence and absence of reassortment. Surprisingly, the asexual populations were found to adapt faster because a possible advantage of sexuals is more than offset by an additional cost due to intrahost competition during co-infection. If this complication could be avoided through an appropriate experimental design, RNA viruses would provide a suitable framework for experimentally testing the predictions of this article.

**Relation to previous studies and outlook**

The quantitative analysis of the speed of evolution of sexual populations compared to that of asexual populations, when both evolve on the same nonepistatic fitness landscape with the same beneficial mutation rate per genome, has a long history (Crow and Kimura 1965, 1969; Maynard Smith 1968, 1971, 1976, 1978; Felsenstein 1974; Kim and Orr 2005). When the number of accessible beneficial mutations is finite, the relevant quantity is the time for all beneficial mutations to be fixed. In this context, Maynard Smith (1971) analyzed the fixation time for sexual and asexual populations evolving on a fitness landscape with \( L \) loci under selection. Each locus has two alleles, one of which confers a beneficial fitness effect in a nonepistatic fashion. For sexual populations, the linkage among loci was assumed weak. Using a rough approximation, Maynard Smith (1971) argued that the time for completing evolutionary changes in asexual populations is \( L \) times longer than that in sexual populations for sufficiently large population size, which implies an \( L \)-fold advantage of sex similar to what we found in our study (see also Maynard Smith 1976). While the analysis by Maynard Smith (1971) is fairly crude and (as conceded by the author) actually not consistent with the simulation results presented in the same article, the conclusion that the advantage of sex becomes stronger with an increasing number of loci under selection is in qualitative agreement with our results, as well as with the related work of Kim and Orr (2005).
Recent studies of the speed of sexual populations in the context of the FM mechanism have mostly focused on the case with an infinite supply of beneficial mutations (Neher et al. 2010; Rouzine and Coffin 2010; Weissman and Barton 2012), exploiting the mathematical progress in treating the spreading of beneficial mutations as a Gaussian traveling wave (Tsimring et al. 1996; Rouzine et al. 2003; Desai and Fisher 2007; Rouzine et al. 2008; Park et al. 2010; Good et al. 2012; Fisher 2013). Rouzine and Coffin (2010) studied how recombination speeds up adaptation when there is standing variation of beneficial mutations. Neher et al. (2010) studied the speed of adaptation of large facultatively sexual populations, starting from a monomorphic state. Similar to our results, Neher et al. (2010) found a regime of intermediate recombination rates where the speed increases logarithmically with population size, however, with a prefactor that varies quadratically with r. Although Rouzine and Coffin (2010) and Neher et al. (2010) investigated the adaptation dynamics of sexual populations with an (effectively) infinite supply of beneficial mutations, their results cannot be directly compared to ours. This is because the model genomes of Rouzine and Coffin (2010) and Neher et al. (2010) assume no or weak linkage between beneficial mutations, whereas in our model mutations in the same locus are tightly linked. Stated differently, unlike our model, which allows for an infinite number of possible beneficial alleles per locus, each locus in the models cited above has only two possible alleles. A related study with an explicit genetic map was recently presented by Weissman and Barton (2012). In future work, it may be of interest to consider models in which the number of linked sites per locus, the number of loci and the rate and mode of recombination can all be varied independently, and the different limiting cases considered in these earlier studies and in this work can be explored in a unified setting.

The two-locus genome considered in this article can be viewed as a simple example of a modular genomic architecture, where recombination occurs between modules but not within a module. Watson et al. (2011) have pointed out that such a modular structure induces a strong benefit for sexual reproduction when there is sign epistasis within the modules and different modules contribute independently to fitness. Another promising avenue for future research would therefore be to extend our approach to include a tunable degree of epistatic interactions within the loci. Following Watson et al. (2011), such interactions should affect not only the speed of adaptation but also the set of genotypes that can be reached at all by the population.

Acknowledgments

We thank Pleuni Pennings and an anonymous reviewer for helpful comments on an earlier version of the manuscript. Support by Deutsche Forschungsgemeinschaft within SFB 680 Molecular Basis of Evolutionary Innovations is gratefully acknowledged. In addition, S.-C.P. acknowledges the support by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant No. 2011-0014680) and the Catholic University of Korea, Research Fund, 2012.

Literature Cited


Good, B. H., I. M. Rouzine, D. J. Balick, O. Hallatschek, and M. M. Desai, 2012 Distribution of fixed beneficial mutations and the
Appendix A: Infinite Population Dynamics for Asexuals \( (r = 0) \) and Obligate Sexuals \( (r = 1) \)

When the population size is infinite, the frequency of genotypes with \( n_i \) mutations at locus \( i \) at generation \( t + 1 \), \( f_{t+1}(n_1, n_2) \), is equal to \( f_t(n_1, n_2) \) as given in Equation 6 due to the law of large numbers. For the deterministic dynamics, the method of (moment) generating functions has been successfully applied to models with nonestatic fitness landscapes (Johnson 1999; Maia et al. 2003; Park and Krug 2007), and we employ this method in this appendix.

Let \( F_t(z_1, z_2) \) denote the generating function for the frequency distribution at generation \( t \), which is defined as

\[
F_t(z_1, z_2) = \sum_{n_1, n_2} z_1^{n_1} z_2^{n_2} f_t(n_1, n_2). \tag{A1}
\]

Since the fitness landscape is multiplicative, the mean fitness at generation \( t \) can be found from \( F_t \) through

\[
\bar{w}_t = F_t(e^{x_1}, e^{x_2}). \tag{A2}
\]

Likewise, we introduce the generating function for \( f_t^i \) in Equation 2, which is obtained from \( F_t \) according to

\[
f_t^i(z_1, z_2) = \sum_{n_1, n_2} z_1^{n_1} z_2^{n_2} f_t^i(n_1, n_2) = \frac{F_t(e^{x_1}, x_2 z_2)}{F_t(e^{x_1}, e^{x_2})}. \tag{A3}
\]

Since \( f_t^i \) in Equation 4 is the convolution of \( g_0 \) and \( f_t^i \), the generating function for \( f_t^i \) is the product of \( F_t^i \) and \( G_0(z_1, z_2) \), where \( G_0 \) is the generating function for mutation probability \( g_0 \) defined as

\[
G_0(z_1, z_2) = \sum_{k_1, k_2} z_1^{k_1} z_2^{k_2} g_0(k_1, k_2). \tag{A4}
\]

Using that \( f_{t+1} \) is the same as \( f_t^i \) for infinite populations, we obtain an iterative evolution equation for \( F_t \) that reads

\[
F_{t+1}(z_1, z_2) = (1 - r) F_t^i(z_1, z_2) + r F_t^i(z_1, z_2) F_t^i(z_1, 1) + r G_0(z_1, z_2) F_t^i(z_1, 2) F_t^i(z_1, 1)
\]

\[
= (1 - r) G_0(z_1, z_2) F_t^i(z_1, e^{x_1} z_2) G_0(z_2) F_t^i(z_1, 2) F_t^i(z_1, 1), \tag{A5}
\]

where \( F_t^i(z_1, z_2) = G_0(z_1, z_2) F_t^i(z_1, 1, 2) \) is the generating function of \( f_t^i \) and

\[
\tilde{g}_0(k) = \sum_m g_0(k, m) = \sum_m g_0(m, k). \tag{A7}
\]

Note that we are using the symmetry \( g_0(k_1, k_2) = g_0(k_2, k_1) \) introduced earlier, but the generalization to asymmetric \( g_0 \) is straightforward.

For \( r = 0 \), Equation A5 can be solved by iterating backward until \( t = 0 \), that is (Park and Krug 2007),

\[
F_t(z_1, z_2) = G_0(z_1, z_2) F_{t+1}(z_1, z_2) F_t(z_1, 2) F_t(z_1, 1)
\]

\[
= G_0(z_1, z_2) G_0(z_1, z_2) F_{t+1}(z_1, z_2) G_0(z_1, z_2) F_t(z_1, 2) F_t(z_1, 1)
\]

\[
= \frac{F_0(z_1, z_2) F_t(z_1, 2) F_t(z_1, 1)}{F_0(z_1, z_2)} \prod_{r=0}^{t-1} \frac{G_0(z_1, z_2) F_t(z_1, 2) F_t(z_1, 1)}{G_0(z_1, z_2) F_t(z_1, 2) F_t(z_1, 1)}.
\]

(A8)

where we have used \( F_0(z_1, z_2) = 1 \) for the homogeneous initial condition. Thus the mean fitness at generation \( t \) is

\[
\bar{w}_t = G_0(e^{x_1}, e^{x_2}). \tag{A9}
\]

For \( r = 1 \), Equation A5 suggests that each locus evolves independently and, in turn, that the generating function is the product of two functions, such as

\[
F_t(z_1, z_2) = F_t^1(z_1) F_t^2(z_2), \tag{A10}
\]

which can be considered the absence of linkage between two locus, or linkage equilibrium. With the above ansatz, we can find an evolution equation for \( \tilde{F}_i(z) \) \((i = 1 \text{ or } 2)\) from Equation A5,

\[
\tilde{F}_{i+1}(z) = \tilde{G}_0(z) \tilde{F}_i(z e^{x_i}), \tag{A11}
\]

which is exactly the evolution equation for an asexual population with marginal mutation probability \( \tilde{g}_0 \). Hence the solution of Equation A11 is

\[
\tilde{F}_i(z) = \prod_{r=0}^{t-1} \frac{\tilde{G}_0(z e^{x_i})}{\tilde{G}_0(z e^{x_i})}, \tag{A12}
\]

where we have again used the homogeneous initial condition \( \tilde{F}_0(z) = 1 \). One can easily check that Equation A10 with \( \tilde{F}_i(z) \) in Equation A12 actually solves Equation A5 for \( r = 1 \) by substitution. Hence the mean fitness at generation \( t \) for \( r = 1 \) is

\[
\bar{w}_t = F_t(e^{x_1}, e^{x_2}) = G_0(e^{x_1}) G_0(e^{x_2}). \tag{A13}
\]

One should note that the ansatz Equation A10 successfully gives the exact solution because the homogeneous initial condition satisfies Equation A10, but the speed does not depend on the initial condition as long as the maximum number of existing mutations at \( t = 0 \) is finite.
From Equations A9 and A13, we deduce the speed of adaptation as

\[ v_a = v(r = 0, N = \infty) = \lim_{t \to \infty} \frac{\ln G_0(e^{x^t}, e^{x^t})}{t} \]  
(A14)

\[ v_s = v(r = 1, N = \infty) = \lim_{t \to \infty} \frac{\ln \tilde{G}_0(e^{x^t}) + \ln \tilde{G}_0(e^{x^t})}{t} \]  
(A15)

where subscripts a and s stand for asexuals and (obligate) sexuals, respectively. Since the arguments of \( G_0 \) in Equation A14 and of \( \tilde{G}_0 \) in Equation A15 increase exponentially, the speed is fully determined by the largest possible fitness effect due to a single mutation event. Thus,

\[ v_a = \text{Max}\{n_1 s_1 + n_2 s_2 | g_0(n_1, n_2) \neq 0\}, \]  
(A16)

\[ v_s = M(s_1 + s_2), \]  
(A17)

where \( M \) is the largest possible number of sites mutated at one locus in a single mutation event,

\[ M = \text{Max}\{n|g_0(n) \neq 0\}. \]  
(A18)

Since, by definition, \( M \) is the maximum of all possible \( n_1 \) and \( n_2 \) with \( g_0(n_1, n_2) \neq 0 \), \( v_s \) cannot be smaller than \( v_a \). Thus, sex is at least not detrimental, although it may have no effect depending on the form of \( g_0 \). For example, if single mutations occur with probability \( U \) and double mutations involving both loci with probability \( U^2 \), corresponding to

\[ g_0(0, 0) = 1 - U - U^2, \quad g_0(1, 0) = g_0(0, 1) = \frac{U}{2}, \]  
(A19)

then \( v_s = v_a = s_1 + s_2 \). On the other hand, if double mutations are forbidden and

\[ g_0(0, 0) = 1 - U, \quad g_0(1, 0) = g_0(0, 1) = \frac{U}{2}, \]  
(A20)

we have \( v_s = s_1 + s_2 \geq v_a = s_2 \) (recall that we assume \( s_2 \geq s_1 \)). Hence the effect of sex significantly depends on the form of \( g_0 \) in the infinite population limit. If \( s_2 > s_1 \) (strict inequality) and if \( g_0 \) is as in Equation 9, beneficial mutations occurring at locus 1 do not contribute to the speed of an asexual population. This can be understood in the framework of clonal interference as the “wasting” of weaker beneficial mutations by the competition with stronger mutations.

Appendix B: Guess Relation in the Presence of Recombination

In this appendix, we show that for evolution on multiplicative, nonepistatic fitness landscapes the Guess relation (Equation 16) is valid even in the presence of recombination.

Let \( w_i(t) \) be the fitness of the \( i \)th individual at generation \( t \), \( w(t) \) the mean fitness of the population, \( w(t) = \sum_i w_i(t)/N \), and \( X_i(t) \) the relative fitness of \( i \)th individual, \( X_i(t) = w_i(t)/w(t) \). We assume that \( X_i(t) \) approaches a well-defined steady state as \( t \) goes to infinity. We take each individual to be characterized by a genome with \( L \) loci, each of which has infinitely many sites. The contribution of a locus to fitness is denoted by \( z_n \) (\( n = 1, \ldots, L \)) and the fitness of an individual with such a genome is \( w = \prod_{n=1}^L z_n \). In the following, \( z_n \) will be called CFn, meaning the Contribution to Fitness of the nth-locus. If a mutation hits the nth locus, CFn changes from \( z_n \) to \( z'_n = z_n v_n \), where \( v_n \) is drawn from a given probability distribution that may vary from locus to locus but does not depend on \( z_n \) or the generation. If \( v_n \) is larger (smaller) than 1, the mutation is beneficial (deleterious). In this appendix, the explicit form of the probability distribution for \( v_n \) does not need to be specified.

We use the vector notation \( z = (z_1, \ldots, z_L) \) for fitness vectors with \( L \) elements. Assume that there are \( N \) individuals and the CFn of individual \( i \) is \( z_i \). The corresponding fitness vector is denoted by \( z_i \). At first, we calculate the expected mean log-fitness at generation \( t + 1 \) assuming that \( X_i(t) \) and \( z_i(t) \) are given.

By selection, the probability density that the fitness vector of an offspring is \( z \) is

\[ f_i(z) = \frac{N X_i(t)}{N} \delta(z - z_i), \]  
(B1)

where \( \delta(x) \) is the \( L \)-dimensional Dirac delta function. Let \( g(v) \) be the probability density that a mutation event changes the CFn of an offspring by \( v_n(z_n \to z'_n = z_n v_n) \) for all \( n \)'s. Then due to mutation, the expected frequency becomes

\[ f_m(z) = \int dv d\mathbf{v} \delta(z - \mathbf{z} \otimes \mathbf{v}) g(\mathbf{v}) f_i(\mathbf{z}'), \]  
(B2)

where \( \mathbf{z}' \otimes \mathbf{v} \) denotes the vector with elements \( z'_n v_n \).

Next we consider recombination. Let \( R(z_1, z_2) \) be the probability density that offspring resulting from the recombination of two parents with fitness vectors \( z_1 \) and \( z_2 \) has fitness \( z \). In general, we can write \( R \) in the form

\[ R(z|z_1, z_2) = \sum_S p(S) \prod_{n=1}^L \delta(z_n - z_{S(n)}.n) \]  
(B3)

where \( S \) runs over all possible outcomes of recombination and \( p(S) \) is the probability of this event. Here \( S(n) = 1 \) (2) if locus \( n \) is inherited from parent 1 (2); hence the total number of possible outcomes is \( 2^L \). Then the final probability density becomes

\[ f(z) = \int dz_1 dz_2 R(z|z_1, z_2) f_m(z_1) f_m(z_2). \]  
(B4)

Now we calculate the expected log-fitness of a randomly chosen individual at generation \( t + 1 \) for given \( w_i(t) \). Since
the probability density that a randomly chosen individual at generation \( t + 1 \) has fitness vector \( z \) is \( f(z) \) given in Equation B4 and the corresponding fitness is \( w = \prod_{n=1}^{i} z_n \), the quantity we want to calculate is

\[
I = \int dw \ln w \, \text{prob}(w) = \int dz \ln \left( \prod_{n=1}^{i} z_n \right) f(z), \quad (B5)
\]

where \( \text{prob}(w) \) is the probability density that an individual has fitness \( w \) at generation \( t + 1 \) for given \( w_i(t) \). Guess (1974b) showed that as long as there is a well-defined steady state the speed \( v \) can be calculated as

\[
v = \left\langle \ln \frac{W(t + 1)}{W(t)} \right\rangle = \left\langle \ln \frac{W_1}{W_2} \right\rangle = \langle I \rangle - \langle \ln W_2 \rangle, \quad (B6)
\]

where \( W_1 \) (\( W_2 \)) is the fitness of a randomly chosen individual at generation \( t + 1 \) (\( t \)) and \( \langle \ldots \rangle \) signifies the average over the steady-state distribution. Loosely speaking, the above relation can be understood as follows: Since the dynamics is symmetric under permutations of the population index, the steady state must have this permutation symmetry as well. Hence the expected log-mean fitness at steady state should be the same as the expected log-fitness of a randomly chosen individual. Moreover, at stationarity the speed can be calculated from the difference of log-fitness between two consecutive generations. Thus Equation B6 follows.

From Equations B1, B2, B3, and B4, we get

\[
I = \int dz \ln \left( \prod_{n=1}^{i} z_n \right) \frac{X_i(t)}{N} \int dv_i dv_j g(v_i) g(v_j) d\ln \left( \prod_{n=1}^{i} z_n \right) R(z_i \otimes v_i, z_j \otimes v_j)
\]

\[
= \sum_{i,j} \frac{X_i(t)}{N} \int dv_i dv_j g(v_i) g(v_j) \int d\ln \left( \prod_{n=1}^{i} z_n \right) R(z_i \otimes v_i, z_j \otimes v_j)
\]

\[
= \sum_{i,j} \mathbb{P}(S) \frac{X_i(t)}{N} \int dv_i dv_j g(v_i) g(v_j) \int d\ln \left( \prod_{n=1}^{i} z_n \right) \left[ \ln \left( \sum_{a=1}^{i} z(a) \right) + \ln \left( \sum_{a=1}^{i} v(a) \right) \right]
\]

\[
= \sum_{S} \mathbb{P}(S) \left( \sum_{i,j} \frac{X_i(t)}{N} \sum_{n=1}^{i} \ln \left( \sum_{a=1}^{i} z(a) \right) + \int dv_i dv_j g(v_i) g(v_j) \int d\ln \left( \prod_{n=1}^{i} z_n \right) \right)
\]

where \( S(n) \) in the subscript of \( z \) and \( v \) can be either \( i \) or \( j \). Since

\[
\sum_{j} \frac{X_i(t)}{N} \ln(z_{i,n}) = \sum_{j} \frac{X_i(t)}{N} \ln(z_{j,n})
\]

\[
\int dv_i dv_j g(v_i) g(v_j) \ln v_{i,n} = \int dv_i dv_j g(v_i) g(v_j) \ln v_{j,n}
\]

\[
= \int dv g(v) \ln v_i,
\]

the summation and integral in the parentheses of the last line of Equation B7 do not depend on \( S \). Due to the normalization \( \sum S \mathbb{P}(S) = 1 \), we get

\[
I = \sum_{i} \frac{X_i(t)}{N} \ln(X_i(t)) - \langle \ln V \rangle + \langle \ln V \rangle + \ln \bar{w}(t), \quad (B10)
\]

where \( V \) is the total effect of fitness increase by a single mutation event, \( V = \prod_i v_{n,i} \), and \( \prod_z z_n = x_i(t) \bar{w}(t) \) was used. Hence the speed becomes

\[
v = \langle I \rangle - \left( \sum_{i} \frac{1}{N} \mathbb{P}(S) \ln w_i \right) = \langle \ln V \rangle + \left( \frac{1}{N} \sum_i \chi_i - 1 \right) \ln \chi_i.
\]

where \( \chi_i \) is the relative fitness of the individual \( i \) at steady state. Note that the formula does not depend on the explicit form of the recombination operator. If we use Equation 9 for the mutation scheme, \( \langle \ln V \rangle = Us \).

The application of the above procedure to the Moran model is straightforward. Hence, the Guess relation is valid for discrete time models regardless of recombination, once the fitness landscape is multiplicative.
Rate of Adaptation in Sexuels and Asexuels: A Solvable Model of the Fisher–Muller Effect

Su-Chan Park and Joachim Krug
File S1

Animation of a breathing traveling wave in frequency space

\( n_1 \) and \( n_2 \) stands for the number of mutations at locus 1 and locus 2, respectively. Genotypes with frequency larger than \( 10^{-5} \) are shown.