ORIGINAL ARTICLE

GENOME STRUCTURE AND THE BENEFIT OF SEX

Richard A. Watson,[1,2] Daniel M. Weinreich,[3] and John Wakeley[4]

[1]Natural Systems group, School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom

[2]E-mail: raw@ecs.soton.ac.uk

[3]Department of Ecology and Evolutionary Biology and Center for Computational Molecular Biology, Brown University, Box G-W, Providence, Rhode Island 02192

[4]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138

We examine the behavior of sexual and asexual populations in modular multipeaked fitness landscapes and show that sexuals can systematically reach different, higher fitness adaptive peaks than asexuals. Whereas asexuals must move against selection to escape local optima, sexuals reach higher fitness peaks reliably because they create specific genetic variants that "skip over" fitness valleys, moving from peak to peak in the fitness landscape. This occurs because recombination can supply combinations of mutations in functional composites or "modules," that may include individually deleterious mutations. Thus when a beneficial module is substituted for another less-fit module by sexual recombination it provides a genetic variant that would require either several specific simultaneous mutations in an asexual population or a sequence of individual mutations some of which would be selected against. This effect requires modular genomes, such that subsets of strongly epistatic mutations are tightly physically linked. We argue that such a structure is provided simply by virtue of the fact that genomes contain many genes each containing many strongly epistatic nucleotides. We briefly discuss the connections with "building blocks" in the evolutionary computation literature. We conclude that there are conditions in which sexuals can systematically evolve high-fitness genotypes that are essentially unevolvable for asexuals.

KEY WORDS: Epistasis, models/simulations, molecular evolution, population genetics, sex.

As alleles at different loci are separated from one another by recombination, sexuals can select on these alleles individually, whereas in asexual populations selection acts on whole genomes (Fisher 1930; Muller 1932). For example, if one individual carries one beneficial allele and another individual carries another beneficial allele (at a different locus) then in an asexual population these two genotypes will be in competition, and one must ultimately fix while the other is lost. In contrast, in a sexual population, we can view selection on each allele as acting almost independently of the other, there is no competition between alleles at different loci, and both beneficial alleles may fix (Fisher 1930; Muller 1932; Hill and Robertson 1966; Kim and Orr 2005; Neher et al. 2010). Separating selection on good genetic material from selection on poor genetic material (e.g., a beneficial mutation from the background it arose on, or an unmutated background from a deleterious mutation) in this manner enables a sexual population to approach a local adaptive peak in a fitness landscape more rapidly than an asexual population, or maintain higher mean fitness at mutation–selection equilibrium (Kondrashov 1988; Huynen 1996; Desai et al. 2007). In this article, we expand on this effect to show that in modular multipeaked fitness landscapes it can also enable sexuals to systematically reach different, higher fitness adaptive peaks than asexuals.

The effect we investigate depends on separating selection on different alleles in exactly the same way as described above, but introduces the idea that alleles differ not by single mutations but by several mutations some of which may be individually deleterious. The significance of viewing the allele of a gene as a composite or "module" in this manner is that despite their essentially particulate behavior under recombination and obvious functional integration, the set of nucleotides that constitute the allele of a gene nonetheless present multiple independent mutational sites. That is, spontaneous point mutation operates on nucleotides individually whereas sexual recombination manipulates the set of tightly linked nucleotides within a gene in combination (and hence respecting its functional unity). Accordingly, evolution by natural selection in asexual populations can only follow fitness increases created by point mutations whereas sexual populations can additionally follow fitness increases created by allelic substitutions that may otherwise require several simultaneous point mutations (see Discussion). This might be particularly pertinent when diverse alleles evolved in different gene pools are being brought together in hybrid zones, but in this article our models work with a single population (see Discussion).

Other studies assessing the ability of sexual and asexual populations to traverse fitness valleys intrinsically depend on stochastic effects that move contra to selective gradients (famously, Wright 1932). This seems inevitable. The very notion of a local adaptive peak is that movement away from the peak is contra to selection and therefore limited, although it may occur with somewhat different probabilities in sexuals and asexuals (Michalakis and Slatkin 1996; Weinreich and Chao 2005; de Visser et al. 2009). In contrast, we investigate a scenario in which sexuals find high-fitness peaks in an adaptive landscape without requiring genetic drift (i.e., "deterministically," Phillips 1996) or changes in allele frequencies that move contra to selection (e.g., Weinreich and Chao 2005), despite the fact that, on the same landscape, asexuals become routinely trapped at local optima. Specifically, the possibility of substituting several mutations simultaneously, as a functional module, creates a scenario in which sexuals may "skip over" the intervening fitness valley, avoiding the need to maintain the individual mutations in genetic contexts in which they would be deleterious.

Consideration of genes as modules in this manner is motivated by the following observations: (1) Recombination rates between genes are higher than recombination rates between nucleotides within a gene. (2) Epistatic interactions between nucleotides in the same gene are likely to be strong and/or numerous and create local optima in the "intragenic" adaptive landscape.

The first observation is straightforward: Because nucleotides in different genes are further apart on the genome than nucleotides within a gene, sexual recombination destroys genetic linkage between the former more rapidly than the latter (this must be the case unless recombination rates are so unbiologically high that all nucleotides recombine freely). It does not assume that intergenic regions are longer than genic regions (or that intergenic regions exist at all) but this fact further amplifies the difference.

The second observation is also natural given the protein-coding role of genes. That is, given that the proper functioning of a gene depends on intricate details of sequence-dependent protein shape and binding affinities, as well as fundamental biophysical properties (DePristo et al. 2005), it is not surprising that a significant amount of epistasis among the nucleotides of a gene is found (Whitlock et al. 1995). We refer to this as intragenic epistasis. Several recent empirical studies confirm that intragenic epistasis occurs (DePristo et al. 2005; Poon and Chao 2005; Poelwijk et al. 2006; Weinreich et al. 2006). Moreover, these studies show that this epistasis includes sign epistasis, where a mutation that is beneficial in one genetic context is deleterious in another, which is necessary and sufficient for the removal of selectively accessible trajectories (Weinreich et al. 2005; de Visser et al. 2009; Rowe et al. 2010). In the β-lactamase gene, sign epistasis among five mutations dramatically reduces the number of selectively accessible trajectories (Weinreich et al. 2006). Stabilizing selection on protein-folding stability (DePristo et al. 2005) implies a large number of locally optimal alleles for every protein-coding gene. Direct evidence of multiple optima in the "intragenic fitness landscape" (see also nucleotide sequence space, Weinreich et al. 2005, protein space, Maynard Smith 1970, and molecular landscape, Gillespie 1984), causing replicate mutational trajectories to reach alleles that differ at multiple sites and differ significantly in fitness, has been shown in several studies (Poelwijk et al. 2006; Lozovsky et al. 2009; Rowe et al. 2010).

Note that other models for the benefit of sex have studied the influence of epistasis extensively (Eshel and Feldman 1970; Felsenstein 1974; Kondrashov 1988; Charlesworth 1990; Barton 1995; Feldman et al. 1997; Otto and Feldman 1997; Barton and Charlesworth 1998; Otto and Lenormand 2002; Keightley and Otto 2006). Epistasis is centrally implicated in the benefit of sex in finite populations (Feldman et al. 1997; Otto and Feldman 1997; Otto and Barton 2001; Barton and Otto 2005), and is also central to Kondrashov's well-known model for the benefit of sex in infinite populations (Kondrashov 1988). Consequently, empirical studies of such epistasis are also popular (Elena and Lenski 1997; West et al. 1998; Kishony and Leibler 2003; Segrè et al. 2005). However, these works concern epistasis between recombining loci, or intergenic epistasis. This cannot produce the effect we address in this article because when intergenic epistasis creates local optima it constrains sexuals as well as asexuals. Moreover, these works generally use only magnitude epistasis, and other restricted forms of epistasis, that cannot create local optima (Kondrashov and Kondrashov 2001; Weinreich et al. 2005; but see de Visser et al. 2009). In contrast, our emphasis is on epistatic subsets of

mutations that do not recombine (and thus mask their individually deleterious effects when substituted together as a unit).

Together these observations highlight an intrinsically modular structure to natural genomes in the sense of a correspondence between physical linkage and epistatic dependencies deriving simply from the fact that genomes are composed of multiple genes each composed of multiple nucleotides. The main intuition of this investigation is therefore that sexuals, by producing variants that change multiple nucleotide sites simultaneously, have the potential to escape local optima that trap asexuals. It is of course possible in principle that a high mutation rate could change many nucleotides simultaneously. But importantly, because the nucleotides within the allele of a gene have been previously subject to selection, the substitution of one allele for another under recombination creates a nonarbitrary genetic change in many nucleotides simultaneously that is highly improbable in asexuals (i.e., requires several specific simultaneous mutations not merely a high mutation rate).

Intuitively, the potential for sexual recombination to create new combinations of "modules" that have each already been subject to selection seems likely to increase the possibility of discovering high-fitness genotypes compared to mutation alone which creates arbitrary genetic variation. This intuition is well-known in the genetic algorithm literature (Holland 1975, 2000; Goldberg 1989), a field of engineering optimization inspired by analogy with evolution by natural selection. Specifically, the "building block hypothesis" asserts that the genetic algorithm with sexual recombination will perform better than a mutation-only algorithm, when it does, because of its ability to select on and recombine building-blocks (Mitchell et al. 1992; Watson 2006). Building-blocks are, loosely speaking, tightly linked subsets of genetic material that are especially high in fitness. Although, initial attempts to verify the building-block hypothesis foundered (Mitchell et al. 1992; Forrest and Mitchell 1993), and subsequent work used various building-block structures with only loose biological analogs (Watson 2004, 2005, 2006; Jansen and Wegener 2005), recent work has shown a principled distinction between the abilities of sexual and asexual genetic algorithms to find fit genotypes using a very simple building-block structure (Watson and Jansen 2007). The underlying principles of the latter work (see also Watson 2005; Watson et al. 2006) form the basis of the present study—but the present study reconceives the genes themselves as building blocks each containing multiple mutations.

EXTENDING THE FISHER/MULLER EFFECT TO INCORPORATE INTRAGENIC EPISTASIS

The Fisher/Muller model assumes free recombination between genes and no intergenic epistasis. Given no intergenic epistasis, the fitness of a genotype, G, with L genes is: $w(G) = \prod_{i=1}^{L} \omega_{a_i}$, where ω_{a_i} is the fitness of the allele, a_i, at the ith gene.

To clarify our assumptions, it is useful to consider the minimal and simplified case of two (or more) genes each containing just two nucleotide sites. Using the (haploid) genome representation of four sites, *abcd*, let *ab* represent two nucleotide sites within gene-1 and likewise *cd* represent gene-2. The recombination rate between sites *a* and *b* is low or zero, likewise the recombination rate between *c* and *d*, whereas we assume free recombination between *b* and *c*. Thus *ab* and *cd* form tightly linked pairs. Following the same pattern, the epistasis between *a* and *b* is assumed to be high, likewise *c* and *d*, but the epistasis between all other pairs of loci is low or zero. Thus *ab* and *cd* also form strongly epistatic pairs.

In general, especially when the number of sites per recombining locus is large (Watson and Jansen 2007) the effect we explore depends on the assumption that recombination among sites in different genes occurs at a higher rate than recombination between sites in the same gene. The simplest physical model to explore the consequences of such linkage modularity assumes no recombination between the sites within each pair but assumes free recombination between these pairs. Hence we refer to the combination of mutations in these genes as an "allele" of a single recombining "locus." In this manner we retain the appropriate level of description to connect with the Fisher/Muller model—that is, both models involve the advantage of independent selection on beneficial alleles at different recombining loci. But, unlike the Fisher/Muller model, there are four alleles for each gene: for example, the first gene has the alleles *ab*, *Ab*, *aB*, *AB*. More generally, we consider multigene genomes with alleles *xy*, *Xy*, *xY*, *XY*.

Each ω_{a_i} is then a function of the mutations at the two sites within that gene. Let the ancestral allele have fitness 1, and the two single mutant alleles have fitness $1 + s_{Xy}$ and $1 + s_{xY}$, and the double mutant $1 + s_{XY}$. (Defining all genes identically in this manner produces a somewhat "unbiological" regularity in the appearance of the fitness landscape in Figure 1, but this simplification is immaterial to the result). We suppose that when a change in environment occurs, mutations at these two sites are beneficial with different magnitudes of effect ($s_{Xy} > 0$, $s_{xY} > 0$, $s_{Xy} \neq s_{xY}$). Without loss of generality, let $s_{Xy} < s_{xY}$. We thus refer to *Xy* as the "inferior" allele (even though it is still beneficial compared to the ancestral allele) and *xY* as the "superior allele" (although we have not specified the fitness of the *XY* allele yet).

Our model hinges on the presence of epistasis within each gene but we wish to make as few assumptions as possible about this intragenic epistasis. Accordingly, we define the intragenic epistasis by varying the single parameter, s_{XY}, which permits us to model all four classes of intragenic epistasis (below). When there is no intragenic fitness epistasis, $s_{XY} = (1 + s_{Xy})(1 + s_{xY}) - 1 = s_{xY} + s_{Xy} + s_{Xy}s_{xY}$, otherwise intragenic epistasis is present. However, subtle deviations in s_{XY} do not necessarily change the fitness
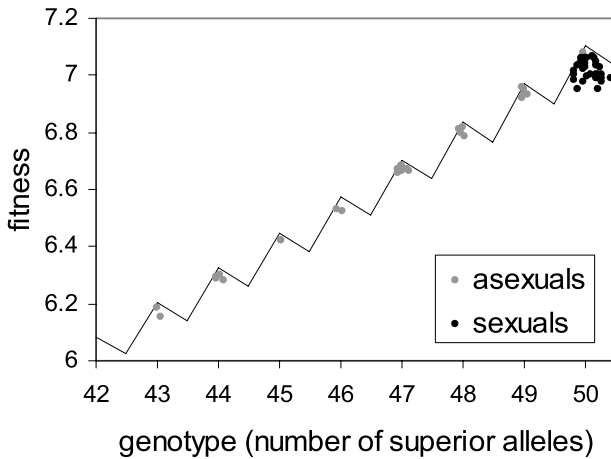
Figure 1. The distribution of genotypes found by 30 independent runs ($\mu = 10^{-5}$, $N = 10^5$) overlaid on the local optima of the fitness surface. Each dot indicates the genotype to which the population converged in one run of the simulation. The exact position of each dot is scattered slightly to facilitate distinction (populations are actually converged on the nearest local optimum, see Methods). All but one run of the asexual population becomes trapped on an inferior genotype that is locally optimal. But all 30 runs of the sexual population converge to the globally optimal genotype. This figure shows that $P_{sex} \approx 1$ and $P_{asex} \approx 0$ for these parameters. Note that there are actually L different genotypes with exactly 49 good alleles, and $L(L − 1)$ with 48, etc. so this one-dimensional classification of points in the fitness landscape should be interpreted with care.
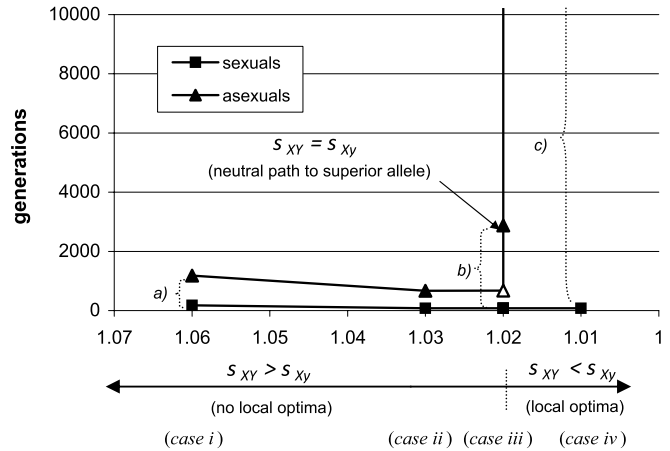

Figure 2. Mean time to convergence to the fittest genotype as the amount of intragenic epistasis is varied by changing the value of s_{XY} ($N = 10^5$, $L = 50$). When $s_{XY} > s_{Xy}$ (i.e., without local optima) an asexual population reaches an optimal genotype in all runs, albeit ≈ 9 times more slowly than sexuals (For both sexuals and asexuals $s_{XY} = 1.06$ takes approximately twice as long as $s_{XY} = 1.03$ only because the optimal alleles, in this case XY, are two mutations away from the ancestral allele, xy, whereas at $s_{XY} = 1.03$ the optimal allele is xY). The speed advantage (a) is due to the conventional Fisher/Muller advantage of sex, but there is no long-term consequence of sex with respect to genotypes discovered. In contrast, for $s_{XY} < s_{Xy}$ (i.e., with local optima), the asexuals become trapped on local optima (Fig. 1), and on average asexuals take longer to reach the optimum genotype than the 30,000 generations used in these simulations (c)—in such cases their time is recorded as 30,000 generations. At the point $s_{XY} = s_{Xy} = 1.02$, the Xy allele admits no beneficial mutations but is not a strict local optimum because it admits a neutral mutation to XY, indirectly enabling access to the superior allele, xY (b). For asexuals two additional points either side of the point $s_{XY} = s_{Xy}$ are examined: $s_{XY} = 1.020001$ (hollow marker) has an average time of 676 generations, $s_{XY} = 1.02$ of 2,874 generations (arrowed), and $s_{XY} = 1.019999$ (not shown) takes time greater than the 30,000 generations simulation limit.

ranking of the alleles, that is, which alleles are preferred by selection (Weinreich et al. 2005). Here we are particularly interested in four regions of intragenic epistasis, produced by varying s_{XY}, each of which affords different selectively accessible routes between alleles via mutation: (1) No epistasis, ($s_{xy} < s_{Xy} < s_{xY} < s_{XY}$). (2) Sign epistasis (Weinreich et al. 2005), $s_{xy} < s_{Xy} < s_{xY} < s_{xY}$. Here the fitness effect of a $x \to X$ mutation, is either beneficial or deleterious depending on whether it is in the background of y or Y, respectively. Thus the superior allele xY can be reached either directly from xy, or via Xy and XY. (3) Conditional neutrality, $s_{xy} < s_{Xy} = s_{XY} < s_{xY}$, at the transition to the region of multiple optima. Here the fitness effect of the $y \to Y$ mutation is either beneficial or neutral depending on whether it is in the background of x or X, respectively. Thus the same routes as case *ii* are available, but the latter now involves a neutral mutation. (4) Multiple optima, $s_{xy} < s_{XY} < s_{Xy} < s_{xY}$. Here $s_{XY} < \min(s_{Xy}, s_{xY})$ causes the Xy and xY alleles to be locally optimal; thus, if the Xy allele is found first, there is no selectively accessible path to the superior xY allele via single point mutations.

Our main investigation, Figure 1, uses case 4 (and Fig. 2 illustrates all cases). When local optima are present, a transition from Xy to xY, though yielding an increase in fitness, either requires a specific two-point mutation changing both sites simultaneously,

or a sequence of two separate mutations where the first is deleterious. If we suppose that this fitness valley between Xy and xY creates an evolutionary impasse then once a population (sexual or asexual) has fixed an inferior, Xy, allele at any locus further fitness improvement at that locus cannot occur (see Appendix A for expected waiting times to cross this fitness valley). However, we will see that the probability of arriving at such an impasse in any locus is higher for asexuals than sexuals because, whereas asexuals are likely to fix inferior alleles due to linkage with superior alleles at other loci, sexuals are unlikely to fix inferior alleles because selection at each locus behaves approximately independently of other loci as per the Fisher/Muller model. It should be clear that what is new here is that we are examining the influence of intragenic epistasis on this model while maintaining the assumption of no

intergenic epistasis as is usual for the Fisher/Muller model. (For comparison, results for the Fisher/Muller model with no intragenic epistasis are also shown, Figure 2, and with both intragenic and intergenic epistasis together, Appendix B).

Finally, as per the Fisher/Muller model, we do not want to assume that diverse alleles are present in the population from the outset and so we commence all simulations with the population converged (at all genes) to the same genetic sequence (i.e., *xy*).

To assess the different abilities of sexuals and asexuals to find high fitness genotypes in this model, we address conditions where sexual populations find optimal genotypes reliably, and then measure the frequency with which asexuals fail to find these optimal genotypes. We then assess how this result is sensitive to the type of intragenic epistasis present and various other parameters of the model.

Methods

We used individual-based simulations to explore sexual and asexual populations separately. Where appropriate we use similar parameters to Kim and Orr (2005): recurrent mutation rate per site $\mu = 10^{-5}$, selection coefficient $s = 0.02$ ($s_{Xy} = s$, $s_{XY} = 2s$) no intergenic epistasis. A haploid individual-based simulation applies Wright/Fisher reproduction and selection taking full account of all finite-population stochastic sampling, then mutation, then recombination in discrete generations. The initial population is converged to the ancestral, *xy*, allele at all loci. Free recombination is applied (at interlocal positions) without interference. Departing from Kim and Orr, we use a range of intragenic epistasis, specifically, $0.01 \leq s_{XY} \leq 0.06$, between the two mutations within each gene. All datapoints in Figure 2 are an average of 30 independent simulations.

Figures indicate the alleles/genotypes attained at convergence of the population or the number of generations for the population to converge to the optimal genotype: Because mutation is recurrent, convergence is taken to mean that greater than 99% (rather than 100%) of the individuals have identical genotypes. Where the maximum fitness genotype is not attained, simulations are run for 30,000 generations (no change in the maximum fitness genotype discovered is observed after 3,000 generations). Although our discussion disregards double mutations, this possibility (or other means of crossing the minimal fitness valley, see Appendix A) is not excluded from our simulations.

Results

Using intragenic epistasis that creates two locally optimal alleles (specifically, $s_{XY} = 0.01$, $s_{Xy} = 0.02$, $s_{xY} = 0.04$), we investigated the probability, *P*, that a population will ultimately fix the superior

xY alleles, rather than the inferior *Xy* alleles, in sexual and asexual populations. In simulations modeling a genotype with only one locus (i.e., one two-site gene, no recombination) we found that with these parameters the superior allele is fixed in all 100 of 100 independent runs, that is, $P \approx 1$, for sexual and asexual populations. This simply means that both alleles arise in the population and the superior allele fixes first, excluding the inferior allele (Kim and Orr 2005). However, in simulations with $L = 50$ loci (averaged over 30 independent runs), we found that the probability of fixing the superior allele at any one locus is decreased in asexuals: $p_{asex} = 0.92$. This arises because selection for the superior allele at one locus cannot be separated from selection for the inferior allele at other loci therefore causing the inferior allele to be fixed at some loci even though the superior allele at those loci exists in the population. Thus, assessing the probability that a population fixes the superior allele at all loci, *P*, (i.e., the probability that the fittest genotype is found) we find that $P_{asex} \ll 1$; only 1 of 30 runs of asexuals found the fittest genotype (Fig. 1). In contrast, the probability of fixing the superior allele at any one locus is unaffected by selection at other loci in sexuals, $P_{sex} \approx 1$, and accordingly, sexuals therefore found the fittest genotypes (with the superior allele at all 50 loci) in all 30 runs, $P_{sex} \approx 1$ (Fig. 1).

Figure 2 shows the time to find the fittest genotypes as a function of intragenic epistasis. To the left of the point $s_{XY} = s_{Xy}$, where there are no local optima, a speed advantage of sexuals is seen (as per the usual Fisher/Muller effect). But to the right of this point, where inferior beneficial alleles create local optima, asexuals are unable to attain the fittest genotypes even in the long term. Accordingly, these results show that sexuals can attain high-fitness genotypes that are inaccessible to asexuals, and that this is due to the type of intragenic epistasis present (thereby departing from the Fisher/Muller effect).

Analysis and Discussion

We start by considering the fixation probabilities of the superior and inferior alleles in the single locus case (a single locus containing two mutational sites). The rates at which the two alternate single-mutation alleles arise by mutation are equal, therefore the ratio of superior alleles that fix will be determined by the relative rates at which the two alleles are lost. Under strong selection weak mutation (SSWM) assumptions (Gillespie 1984) (i.e., $s \gg 1/N \gg \mu$) where each new mutation is either fixed or lost from the population before the next mutation occurs, a higher value of *P* might be expected for two alleles with selection coefficients s_{sup} and s_{inf} when fixation probability is correlated with fitness (Weinreich et al. 2006), that is, $P = s_{sup}/(s_{sup} + s_{inf})$. For superior alleles with twice the fitness increase of inferior alleles, that

is, $s_{sup} = 2s$ and $s_{inf} = s$, as used here, $P = 2s/(s + 2s) = 2/3$. However, when different alleles segregate simultaneously (i.e., outside SSWM assumptions) P can be greater than $s_{sup}/(s_{sup} + s_{inf})$ because of the relative speeds with which competing alleles may fix—the fixing of one causing the other to be lost—rather than the independent probabilities of loss. Specifically, because the superior allele fixes faster than the inferior allele, the chance that the superior allele arises and fixes before an inferior allele can fix is high even if the inferior allele was already segregating at the time when the superior allele appeared (Gerrish and Lenski 1998; Kim and Orr 2005; Desai et al. 2007). Hence $P \approx 1$ when modeling a genotype with only one locus (which is necessarily the same for sexual and asexual scenarios).

For multilocus systems, the probability that a sexual population will ultimately fix the superior allele at a given locus is approximately as per the single locus case because under free recombination each locus behaves approximately independently. But the ability of asexuals to fix the superior beneficial alleles at each locus is significantly depressed for large L due to the conventional Fisher/Muller effect—that is, selection on a genotype containing a superior allele at one locus causes fixation of an inferior allele at another locus. Specifically, in an asexual population it is genotypes rather than alleles that compete for fixation because all loci are linked, thus competition between alleles at one locus is interfered with (Huynen 1996; Kim and Orr 2005; Desai et al. 2007) by competition between alleles at other loci. Accordingly, whereas P_{sex} remains approximately 1 for large L, $P_{asex} < 1$ for large L (Appendix B). In general, the probability, P, that the superior allele is found at all L loci is $P = p^L$. Therefore, given that we observe $p_{sex} \approx 1$ and $p_{asex} < 1$, we also observe $P_{sex} \approx 1$ but $P_{asex} \approx 0$--that is, asexuals cannot reliably attain genotypes with superior alleles at all loci.

As per the conventional Fisher/Muller model (also Felsenstein 1974; Kim and Orr 2005), the benefit of sex shown here involves the ability of sexual populations to use beneficial alleles that arise in parallel in the population. But the consequences of losing beneficial alleles that arise in parallel is quite different in the results shown here compared to a model in which there is no intragenic epistasis. Here we are not merely concerned with the probability that a beneficial allele is lost but also with the consequent effect that a beneficial allele becomes permanently selectively inaccessible via mutation. Put simply, when there is no intragenic epistasis there is no constraint on the selective accessibility of beneficial alleles—so if a beneficial allele is lost on one occasion, there is nothing to stop it being found again later under recurrent mutation (see $s_{XY} > s_{Xy}$ region in Fig. 2). But when intragenic epistasis restricts selectively accessible trajectories, a population that loses the opportunity to access that allele from one genetic background may be unable to access it from another. Specifically, in this case, if the superior allele is

lost from the population and the inferior allele fixes, there is no longer any selectively accessible trajectory to reach the superior allele—i.e., although the superior allele was selectively accessible from the ancestral allele, intragenic epistasis makes it selectively inaccessible from the inferior allele. In contrast, in the conventional Fisher/Muller model, any beneficial allele that is lost remains selectively accessible because there is no intragenic epistasis, indeed there is no distinction between mutational sites and recombining loci, so alleles at a locus are mutational neighbors and there cannot be any intragenic epistasis. Our results thus show a scenario in which intragenic epistasis permanently "locks in" the adaptive fate of the population in the sense that the competition between beneficial alleles at different loci in asexuals, that is, the Fisher/Muller effect, has the consequence of permanently preventing access to fit genotypes (Figs. 1 and 2) rather than just slowing it down. This contrasts with prior models of clonal interference (Huynen 1996; Kim and Orr 2005; Desai et al. 2007) showing that asexual reproduction slows down the accumulation of beneficial alleles as expected from the Fisher/Muller model (Fisher 1930; Muller 1932; Hill and Robertson 1966; Felsenstein 1974; Gerrish and Lenski 1998; Neher et al. 2010).

Figure 2 shows a strong contrast in the effect given different assumptions about intragenic epistasis. Where there are no local optima present, the speed advantage of sexuals is seen, as per Fisher/Muller, but where the inferior alleles are locally optimal, asexuals are unable to attain the fittest genotypes even in the long term. Given that $s_{XY} < s_{Xy}$ creates numerous local optima in the fitness landscape, the discontinuity in the asexuals curve should not be surprising—the presence of local optima, of course, makes discovery of the globally optimal genotype difficult. The truly striking feature of this figure is not that asexuals are trapped by intragenic local optima but that sexuals are effectively unhindered. Note that in the conventional form of the Fisher/Muller model, the two alleles at each locus (the ancestral allele and the mutant allele) are assumed to be mutational neighbors. It therefore cannot be the case that there is any restricting epistasis involved that might preclude a selectively accessible trajectory to the fittest alleles at each locus and hence the fittest genotypes. In the conventional model, and whenever there is no restriction on evolutionary paths created by intragenic epistasis, the attainment of fit genotypes may be slow because we have to wait for mutations to accumulate serially and any that are discovered in parallel (in different genotypes) are wasted. But here, when intragenic epistasis creates locally optimal alleles, if beneficial alleles that are discovered in parallel are wasted (as they are in asexuals) it does not matter how long we wait for those mutations to occur again serially because the intermediate mutations required to reach those alleles are no longer beneficial.

These results therefore depend on the presence of local optima in the intragenic fitness landscape because this causes

asexuals, unlike sexuals, to become "stuck." When multiple local optima are present, varying the mutation rate, population size, intergenic epistasis, selection coefficients and the number of loci has very little effect of interest (Appendix B). Specifically, sexuals succeed (i.e., exhibit $P_{sex} = 1$) for population sizes as small as $N = 1000$ (default 100,000), for mutation rates as low as $\mu = 10^{-7}$ (default, 10^{-5}), for $N\mu$ from 0.01 to 0.5 (default, $N\mu = 1$), for any number of loci tested (default, $L = 50$), for intergenic epistasis values in the range $\varepsilon = [0.5, 1.5]$ (defined in Appendix B) (default $\varepsilon = 1$) and for the selection coefficients $s = 0.02$ to $s = 1$ (default $s = 0.02$). In contrast, asexuals fail to show a reasonable (greater than 1%) possibility of finding the fittest genotypes with any of the population sizes, mutation rates, any $N\mu < 10$, all intergenic epistasis values, and all selection coefficients tested. However, as expected, if the number of loci, L, is small then the number of local optima, 2^L, may also be small enough that asexuals (using the default $N = 10^5$ and $Nu = 1$) do not necessarily become trapped (Fig. B2). This, and the case in which intragenic epistasis does not produce local optima (Fig. 2, left), are the only cases examined where asexuals find the fittest genotype. Note that the presence of intergenic sign epistasis (not examined in these studies), where the genetic background at one locus may alter which allele was the superior allele at some other locus (Weinreich et al. 2005), will interfere with the ability of both sexuals and asexuals to find the fittest genotypes.

Overall, these results support an expectation that sexual recombination will be selectively advantageous; showing that it can enable a population to avoid or escape local optima in nucleotide sequence space. Specifically, sexuals avoid local optima in the sense that sexual populations do not converge on a locally optimal genotype as asexual populations do. Sexuals escape local optima in the sense that even if every individual in the population has a locally optimal genotype (i.e., all single-nucleotide changes are deleterious), a sexual population can nonetheless converge on the global optimum because recombinative substitutions (changing more than one nucleotide simultaneously) can still permit fitness increases if the epistasis that creates these multiple optima is modular.

Although the presence of local optima in intragenic epistasis as used in this model is supported empirically as discussed above, surprisingly, we find that this is not strictly necessary to see a new effect. Specifically, it might be expected that the effect we have shown will not occur if pathways of neutral mutations (Huynen 1996; Fontana and Schuster 1998) remove local optima, that is, facilitate escape from would-be optima via neutral pathways to superior genotypes. It is all the more interesting then that Figure 2 shows a considerable new effect without local optima. Specifically, in the case in which $s_{XY} = s_{Xy}$ (Fig. 2), although asexuals do reach the optimal genotype in this scenario we see a sharp increase in the time taken to do so. The fact that such "pseudo

optima" (genotypes that admit no immediate beneficial mutations but permit access to superior genotypes after one or more neutral mutations have been accumulated) significantly retard asexuals much more than in the conventional Fisher/Muller effect shows that the effect does not strictly require local optima. It is, more correctly, the lack of selectively supported trajectories that defines the onset of the effect. This result makes the effect strongly relevant in cases of widespread intragenic neutrality in molecular evolution (Fontana and Schuster 1998), broadening the conditions in which the effect will be seen.

Clearly, evolving populations can, with limitations, cross small fitness valleys. The key distinction we refer to when talking of "attainable" and "unattainable" in this article is that the former only requires evolutionary changes that are supported by selection and are therefore rapid, whereas the latter requires evolutionary changes that are not supported by selection (although this can occur deterministically, Phillips 1996, in large populations, Weinreich and Chao 2005). In the case of multiple local optima the distinction is then quite intuitive—that is, "unattainable" means multiple specific point mutations must occur simultaneously to find a fitter genotype directly by mutation (other means of valley-crossing, Weinreich and Chao 2005, are considered in Appendix A). Simple analysis (see appendix) shows that only in cases in which some transitions are required that are not supported by selection, can there be a principled significant difference between sexuals and asexuals in the expected time to reach the fittest genotype. Specifically, the possibility of a population moving against selection to escape a local optimum has expected time approximately proportional to $1/N\mu^2$, and the expected time to reach the fittest allele if it is available via a neutral path is approximately proportional to $1/\mu$. These are contrasted with times approximately proportional to $1/N\mu s$ for sexuals who avoid these scenarios by substituting superior alleles from other genotypes. Moreover, this is the minimum distinction caused by this effect because we are using a scenario in which the superior allele differs from the locally optimal inferior allele at only two sites. If the mutational distance from a locally optimal allele to the nearest allele of higher fitness is large, then escaping the locally optimal allele becomes exponentially unlikely with the width of the valley (Watson and Jansen 2007). But even when a genotype of higher fitness is only two point mutations away from the locally optimal allele, such an escape is biologically implausible.

Note that immediate (and ongoing) selective benefits are necessary for the maintenance of sexual recombination in a population as studied in models with modifier loci (Feldman et al. 1997 Otto and Feldman 1997; Keightley and Otto 2006). The long term consequences of sex studied here, such as an ability of sexual populations to ultimately find higher fitness genotypes than asexual populations, do not speak directly to the issue of faster adaptation nor immediate benefits. However, the long-term

consequences of sex observed in this case are coincident with an immediate benefit of sex due to the usual Fisher/Muller effects in finite populations.

Clearly the results shown here depend on the ability of the population to generate allelic diversity (as modeled), or perhaps to exploit standing genetic variation. Additional simulation results (see also Watson and Jansen 2007) show that an island model with low interdeme migration significantly broadens the conditions on N and μ where sexuals can find high-fitness genotypes that asexuals cannot because these conditions increase the likelihood of alternate alleles being maintained in the population. This produces new significance for population structure in evolution (Otto and Lenormand 2002), and also illustrates that hybridization (Macken and Perelson 1989; Rieseberg et al. 2003) and lateral gene transfer may produce genotypes that could not be produced by a single sexual (or asexual) population. Relatedly, the results also depend on conditions in which multiple genes (with strong intragenic epistasis) are under simultaneous selection. How often this might be biologically relevant is unknown but may also be more likely in cases of hybridization where different populations have been adapting for some time to different environmental conditions.

Note that although Wright's shifting balance theory (Wright 1932; Phillips 1996) also concerns subdivided populations and escape from local optima, it is fundamentally dependent on genetic drift and the probability of evolutionary trajectories that move contra to selective gradients. Shifting balance theory is also not fundamentally concerned with the differing behavior of sexual and asexual populations nor the interaction of subdivision with the Fisher/Muller effect. In contrast, the ability of sexuals to reach high-fitness genotypes in the effect we illustrate in this article does not require any subpopulation to move against selective gradients, as discussed above, and is not available to an asexual population even if subdivided.

RELATED EPISTASIS MODELS
There are a number of other works whose partial overlap with the present model deserves clarification. The presence of strong intragenic epistasis and weak or no intergenic epistasis, as we assume, creates a sequence-based fitness landscape that is partially correlated (Perelson and Macken 1995). Interestingly, the block model of fitness defined in Perelson and Macken's paper concerns a qualitative difference between epistasis at two different scales as in our model, where epistasis within blocks is random and epistasis between blocks is absent. However, Perelson and Macken's model concerns multiple blocks within one locus and the recombination of blocks is therefore not addressed. If intragenic recombination were considered then such an epistatic structure within loci would make the effect we model relevant to the creation of chimeric proteins (Meyer et al. 2003). This would

be consistent with Drummond et al's (2005) empirical observation that intragenic recombination is less likely to be disruptive than (intragenic) mutation.

The schema hypothesis (Voigt et al. 2002) discusses an analogous model of block-like epistatic structure, with recombination between blocks, where the blocks are intragenic domains. As Martin et al. (2005) point out, the schema hypothesis can be seen as an intragenic analogue of the complexity hypothesis of Jain et al. (1999). That is, both hypotheses assert that recombination is favored when least disruptive and discuss the idea that some fragments of genetic material will be more robust, more "modular," than others with respect to transfer into different backgrounds or replacement by different sequences. This modularity concept refers to the sparseness of dependencies between the gene and other genes in the complexity hypothesis, and the sparseness of dependencies between the intragenic domain and other intragenic domains in the schema hypothesis. Both hypotheses therefore argue for the significance of modular epistatic structure in natural genomes and the significance of recombination between modules. Interestingly, the physical linkage structure of intragenic subdomains is supported by the intron/exon structure of eukaryotic genes. At the opposite physical scale, some evidence suggests that linkage in haplotype blocks is supported by recombination hot spots (Przeworski 2005; see also Neher and Shraiman 2009), although the functional significance of blocks at the larger scale that might benefit from this linkage pattern is not clear.

However, Jain et al. and Martin et al. focus on minimizing the disruption of recombination, as in the Voigt et al. model, not the possibility that recombination can produce new genotypes that asexual populations cannot. If we merely assume that intramodule dependencies (intraschema dependencies in Voigt et al, intragenic dependencies here) are stronger than intermodule dependencies this is sufficient to produce a selective preference for recombination that respects module boundaries over recombination that disrupts modules (as Voigt et al., Jain et al., and Martin et al, suggest). But it should be remembered that asexuals do not disrupt modules at all—so this reasoning does not in itself describe any benefit to sex. In short, such reasoning is not about a benefit of sex but about minimizing the deleterious effects of sex.

Kouyos et al. (2006) use a model of epistasis built from multiple disjoint pairwise interactions, just as we do, but that work assumes free recombination between all mutational sites and therefore does not allow alleles to be substituted reliably as units.

Our modelling approach benefits from cross-fertilization with evolutionary computation (Holland 1975; Watson 2005; Watson 2006; Watson and Jansen 2007) where consideration of multilocus models with complex epistasis is common. Several evolutionary computation models have shown a principled distinction in the accessibility of fit genotypes for sexual and

asexual populations (e.g., Spears 1992, 2004; Culberson 1995; Shapiro and Prügel-Bennett 1997; Rogers and Prügel-Bennett 2001; Watson 2001, 2004; Jansen and Wegener 2005; Watson 2006, 2005). However, although some of these models have used various forms of building-block structure they have only loose biological analogues—see Watson and Jansen (2007) for discussion. Nonetheless, a modular or building-block form of epistasis is familiar in evolutionary computation (Holland 1975; Watson 2005; Watson 2006; Watson and Jansen 2007) and the effect we model here depends on genetic modularity in the sense that units of physically linked genetic material correspond to units with complex epistatic dependencies (Felsenstein 1974; Charlesworth 1990; Watson 2006; Watson and Jansen 2007). This structure may be present at super- or sub- genetic scales also (e.g., gene complexes [Garcia-Fernandez 2005] or exons, respectively), but the genes themselves are an obvious form of such structure (i.e., the nucleotides within a gene are both tightly physically linked and exhibit complex epistasis) that has been abstracted-away in prior models that do not model recombining alleles as multisite structures with internal epistasis (see below).

The concatenated deceptive trap models (Deb and Goldberg 1992) have a very similar structure to that modeled in this article, but there are important (although somewhat subtle) differences in the underlying assumptions adopted that have prevented the ability to show a principled distinction between the abilities of sexual and asexual populations to find fit genotypes (Watson and Jansen 2007). Relatedly, it should be noted that computer science theory can be used to formalize the "can evolve"/"cannot evolve" distinction further by expressing the expected time to find the fittest genotype as a function of the mutational separation between the locally optimal inferior allele and the superior allele (Watson and Jansen 2007). Also, although tight physical linkage within genes is not essential in the two-site alleles modeled in the current article, the tightness of linkage between sites within genes becomes very important when the mutational separation of locally optimal beneficial alleles is larger (Watson and Jansen 2007).

Finally, we find it useful to describe asexual populations, that follow fitness increases created by point mutations, using trajectories in "genotype sequence space" (Wright 1932; Gillespie 1984; Provine 1986; Weinreich et al. 2005) whereas sexual populations, that can additionally follow fitness increases created by allelic substitutions, can also move in "allele frequency space" (Fisher 1930; Wright 1931; Haldane 1932; Provine 1986; Weinreich et al. 2005). (Neher and Shraiman, 2009, address a similar distinction when they discuss how selection effectively transitions from alleles to genotypes given different epistatic assumptions). We suggest that conflating these two conceptually different (but systematically related) spaces has been the main reason why the effect we model has been overlooked in previous models. In particu-

lar, any model that assumes alleles at a locus are the mutational neighbors of one another, a very widespread assumption in population genetic models, makes this mistake because in such models recurrent mutation offers the same genetic moves as allelic substitutions and therefore the effect we have shown is eliminated. Likewise, any model that disregards the genetic map, or uses only two sites/loci where the map is degenerate (i.e., affords only one ordering of positions), cannot express the modularity we address and will preclude the benefit of sex we have illustrated.

Conclusions

Many hypotheses regarding the benefit of sex and recombination concern differences in the rate at which beneficial mutations are accumulated but not differences in the genotypes that are ultimately attainable. That is, in prior models asexuals can always reach genotypes as good as (or sometimes better, de Visser et al. 2009) than those found by sexuals, although perhaps more slowly. Prior models for the benefit of sex, including those that address (intergenic) epistasis, suffer from the assumption that the alleles at a locus differ by only a single nucleotide mutation—which makes the accessibility of fit genotypes identical for sexuals and asexuals. These models thereby overlook a simple idea—that sexuals are not restricted by intragenic local optima in the same way that asexuals are because sexuals can obtain alternate alleles from other individuals by recombination. Here, we show that intragenic epistasis can prevent asexuals from finding high-fitness genotypes that are nonetheless found quickly and reliably by sexuals. Our simulations use a modular epistasis structure inspired by intragenic epistasis rather than intergenic epistasis. A significant consequence of sexual recombination in natural populations may thus arise from even the most ubiquitous architecture of genomes: the fact that genomes contain functionally integrated and physically particulate genes, each composed of thousands of epistatic nucleotides.

The inability of evolution by natural selection to escape from a local adaptive peak is one of the basic consequences of the gradualist framework (Watson 2006). However, this article shows that what is a local optimum for one type of population is not a local optimum for another—sexuals can escape local optima that trap asexuals without moving against selective gradients because substituting one allele for another skips over inferior mutational intermediates. This requires us to revise our naïve attachments to what is evolvable and what is not evolvable under natural selection, and to realize that what is unevolvable to asexuals may be evolvable to sexuals. These findings show that what Sewell Wright described as "the central problem of evolution," the presence of local fitness peaks, can create an evolutionary impasse for asexuals but nonetheless be trivially evaded by sexuals.

LITERATURE CITED

Barton, N. H., and S. P. Otto. 2005. Evolution of recombination due to random drift. Genetics 169:2353–2370.

Barton, N. H., and B. Charlesworth. 1998. Why sex and recombination? Science 281:1986–1990.

Barton, N. H. 1995. A general model for the evolution of recombination. Genet. Res. 65:123.

Charlesworth, B. 1990. Mutation selection balance and the evolutionary advantage of sex and recombination. Genet. Res. 55:199–221.

Culberson, J. C. 1995. Mutation-crossover isomorphism and the construction of discriminating functions. Evol. Computation 2:279.

Deb, K., and D. E. Goldberg. 1992. Analyzing deception in trap functions. Pp. 93–108 in D. Whitley, ed. Foundations of genetic algorithms 2. Morgan Kaufmann, San Mateo, CA.

DePristo, M. A., D. M. Weinreich, and D. L. Hartl. 2005. Missense meanderings in. sequence space: a biophysical view of protein evolution. Nat. Rev. Genet. 6:678–687.

Desai, M. M., D. S. Fisher, and A. W. Murray. 2007. The speed of evolution and maintenance of variation in asexual populations. Curr. Biol. 17:385–394.

de Visser, J. A. G. M., S.-C. Park, and J. Krug. 2009. Exploring the effect of sex on empirical fitness landscapes. Am. Nat. 174:S15–S30.

Drummond, D. A., J. J. Silberg, C. O. Wilke, M. M. Meyer, and F. H. Arnold. 2005. On the conservative nature of intragenic recombination. Proc. Natl. Acad. Sci. USA 102:5380.

Elena, S. F., and R. E. Lenski. 1997. Test of synergistic interactions between deleterious mutations in bacteria. Nature 390:395.

Eshel, I., and M. W. Feldman. 1970. On the evolutionary effect of recombination. Theor. Popul. Biol. 1:88–100.

Feldman, M. W., S. P. Otto, and F. B. Christiansen. 1997. Population genetic perspectives on the evolution of recombination. Annu. Rev. Genet. 30:261–295.

Felsenstein, J. 1974. The evolutionary advantage of recombination. Genetics 78:737–756.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford Univ. Press, Oxford.

Fontana, W., and P. Schuster. 1998. Shaping space: the possible and the attainable in RNA genotype-phenotype mapping. J. Theor. Biol. 194:491–515.

Forrest, S., and M. Mitchell. 1993. What makes a problem hard for a genetic algorithm? Some anomalous results and their explanation. Machine Learning 13:285–319.

Garcia-Fernandez, J. 2005. The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. 6:881–892.

Gerrish, P. J., and R. E. Lenski. 1998. The fate of competing beneficial mutations in an asexual population. Genetica 102–103:127–144.

Gillespie, J. H. 1984. Molecular evolution over the mutational landscape. Evolution 38:1116–1129.

Goldberg, D. E. 1989. Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Reading, MA.

Haldane, J. B. S. 1932. A mathematical theory of natural selection. Part VIII. Metastable populations. Proceedings of the Cambridge Philosophical Society 26:137–142.

Hill, W. G., and A. Robertson. 1966. The effects of linkage on limits to artificial selection. Genet. Res. 8:269–294.

Holland, J. H. 1975. Adaptation in natural and artificial systems. Univ. Michigan Press, Ann Arbor, MI.

Holland, J. H. 2000. Building blocks, cohort genetic algorithms, and hyperplane-defined functions. Evol. Computation 8:373.

Huynen, M. A. 1996. Exploring phenotype space through neutral evolution. J. Mol. Evol. 43:165–169.

Jansen, T., and I. Wegener. 2005. Real royal road functions—where crossover provably is essential. Discrete Appl. Math. 149:111.

Jain, R., M. C. Rivera, and J. A. Lake. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proc. Natl. Acad. Sci. USA 96:3801.

Keightley, P. D., and S. P. Otto. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. Nature 443:89–92.

Kim, Y., and H. A. Orr. 2005. Adaptation in sexuals vs. asexuals: clonal interference and the Fisher-Muller model. Genetics 171:1377–1386.

Kishony, R., and S. Leibler. 2003. Environmental stresses can alleviate the average deleterious effect of mutations. J. Biol. 2:14.

Kondrashov, A. S. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336:435–441.

Kondrashov, F. A., and A. S. Kondrashov. 2001. Multidimensional epistasis and the disadvantage of sex. Proc. Natl Acad. Sci. USA 98:12089–12092.

Kouyos, R. D., S. P. Otto, and S. Bonhoeffer. 2006. Effect of varying epistasis on the evolution of recombination. Genetics 173:589.

Lozovsky, E. R., T. Chookajorn, K. M. Brown, M. Imwong, P. J. Shaw, S. Kamchonwongpaisan, D. E. Neafsey, D. M. Weinreich, and D. L. Hartl. 2009. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. Proc. Natl. Acad. Sci. USA 106:12025–12030.

Macken, C. A., and A. S. Perelson. 1989. Protein evolution on rugged landscapes. Proc. Natl. Acad. Sci. USA 86:6191.

Martin, D. P., E. Van der Walt, D. Posada, and E. P. Rybicki. 2005. The evolutionary value of recombination is constrained by genome modularity. PLoS Genet. 1:e51.

Maynard Smith, J. 1970. Natural selection and the concept of protein space. Nature 225:563–564.

Meyer, M. M., J. J. Silberg, C. A. Voigt, J. B. Endelman, S. L. Mayo, Z.-G. Wang, and F. H. Arnold. 2003. Library analysis of SCHEMA-guided protein recombination. Protein Science 12:1686.

Michalakis, Y., and M. Slatkin. 1996. Interaction of selection and recombination in the fixation of negative-epistatic genes. Genet. Res. 67:257.

Mitchell, M., S. Forrest, and J. H. Holland. 1992. The royal road for genetic algorithms: fitness landscapes and GA performance. Pp. 245–254 in F. J. Varela and P. Bourgine, eds. Procs. of first European conference on artificial life. MIT Press, Cambridge, MA.

Muller, H. J. 1932. Some genetic aspects of sex. Am. Nat. 66:118–138.

Neher, R. A., and B. I. Shraiman. 2009. Competition between recombination and epistasis can cause a transition from allele to genotype selection. Proc. Natl. Acad. Sci. USA 106:6866–6871.

Neher, R. A., B. I. Shraiman, and D. S. Fisher. 2010. Rate of adaptation in large sexual populations. Genetics 184:467–481.

Otto, S. P., and N. H. Barton. 2001. Selection for recombination in small populations. Evolution 55:1921.

Otto, S. P., and M. W. Feldman. 1997. Deleterious mutations, variable epistatic interactions, and the evolution of recombination. Theor. Popul. Biol. 51:134–147.

Otto, S. P., and T. Lenormand. 2002. Resolving the paradox of sex and recombination. Nat. Rev. Genet. 3:252–261.

Perelson, S., and C. A. Macken. 1995. Protein evolution on partially correlated landscapes. Proc. Natl. Acad. Sci. USA 92:9657.

Phillips, P. C. 1996. Waiting for a compensatory mutation: phase zero of the shifting-balance process. Genet. Res. 67:271.

Poelwijk, F. J., D. J. Kiviet, and S. J. Tans. 2006. Evolutionary potential of a duplicated repressor-operator pair: simulating pathways using mutation data. PLoS Comput. Biol. 2:e58.

Poon, A., and L. Chao. 2005. The rate of compensatory mutation in the bacteriophage φX174. Genetics 170:989–999.

Provine, W. B. 1986. Sewall Wright and evolutionary biology. Univ. of Chicago Press, Chicago.

Przeworski, M. 2005. PERSPECTIVES GENETICS: motivating hotspots. Science 310:247.

Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, and C. Lexer. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1189–90.

Rogers, A., and A. Prügel-Bennett. 2001. A solvable model of a hard optimisation problem. Pp. 207–221 in L. Kallel, B. Naudts, and A. Rogers, eds. Proceedings of theoretical aspects of evolutionary computing. Springer, Berlin.

Rowe, W., M. Platt, D. C. Wedge, P. J. Day, D. B. Kell, and J. Knowles. 2010. Analysis of a complete DNA-protein affinity landscape. J. R. Soc. Interface 7:397–408.

Segrè, D., A. Deluna, G. M. Church, and R. Kishony. 2005. Modular epistasis in yeast metabolism. Nat. Genet. 37:77.

Shapiro, J. L., and A. Prügel-Bennett. 1997. Genetic algorithms dynamics in two-well potentials with basins and barriers. Pp. 101–116 in R. K. Belew and M. D. Vose, eds. Foundations of genetic algorithms 4. Morgan Kaufmann, San Francisco.

Spears, W. M. 1992. Crossover or mutation? Pp. 221–237 in D. Whitley, ed. Foundations of genetic algorithms 2. Morgan Kaufmann, San Mateo, CA.

———. 2004. Evolutionary algorithms: the role of mutation and recombination. Springer-Verlag, Berlin.

Voigt, C. A., C. Martinez, Z. G. Wang, S. L. Mayo, and F. H. Arnold. 2002. Protein building blocks preserved by recombination. Nat. Struct. Biol. 9:553.

Watson, R. A. 2001. Analysis of recombinative algorithms on a non-separable building block problem. Pp. 69–89 in W. N. Martin and W. M. Spears, eds. Foundation of genetic algorithms 6. Morgan Kaufmann, San Francisco, CA.

———. 2004. A Simple two-module problem to exemplify building-block assembly under crossover. Pp. 161–171 in X. Yao, J. A. Bullinaria, J. Rowe, P. Tiňo, and A. Kabán, eds. Parallel problem solving from nature – PPSN VIII. Springer, Berlin (LNCS 3242).

———. 2005. On the unit of selection in sexual populations. Pp. 895–905 in Advances in artificial life, eighth European conference (ECAL 2005). Springer, Berlin (LNCS 3630).

———. 2006. Compositional evolution: the impact of sex, symbiosis and modularity on the gradualist framework of evolution. MIT Press, Cambridge, MA.

Watson, R. A., and T. Jansen. 2007. A building-block royal road where crossover is provably essential. GECCO '07: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation 2:1452–1459.

Watson, R. A., D. Weinreich, and J. Wakeley. 2006. Effects of intra-gene fitness interactions on the benefit of sexual recombination. Biochem. Soc. Trans. 34:560–561.

Weinreich, D. M., and L. Chao. 2005. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. Evolution 59:1175.

Weinreich, D. M., N. F. Delaney, M. A. DePristo, and D. L. Hartl. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312:111–114.

Weinreich, D. M., R. A. Watson, and L. Chao. 2005. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution 59:1165–1174.

West, S. A., A. D. Peters, and N. H. Barton. 1998. Test for epistasis between deleterious mutations. Genetics 149:435.

Whitlock, M. C., P. C. Phillips, F. B. Moore, and S. J. Tonsor. 1995. Multiple fitness peaks and epistasis. Annu. Rev. Ecol. Syst. 26:601–630.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

———. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In Proceedings of the VI International Congress of Genetics 1:356–366.

Associate Editor: S. Nuismer

Appendix A
EXPECTED WAITING TIMES TO CROSS A FITNESS VALLEY

To quantify the distinction between trajectories that are and are not supported by selection, we calculate the expected time, T, to reach the superior allele from the ancestral allele at a single locus for different classes of intragenic epistasis, as defined by different values of s_{XY}. Our analysis considers just one locus with no recombination between the two sites involved. This is analyzed for all four cases of epistasis discussed in the main text, and for both a scenario in which the superior allele fixes first, T_{sup}, and a scenario in which the inferior allele fixes first, T_{inf}. Our goal is to assess just how "stuck" a population will be if it fixes the inferior allele first, especially in the case in which the inferior allele is locally optimal. The main body of the article has referred to this scenario as an evolutionary impasse but clearly there is a possibility of a population moving against selection to escape this local optimum. What is the expected time for this possibility? Also of particular interest is the case in which the inferior allele admits no beneficial mutations but does enable access to the superior allele via an initial neutral mutation: Is the expected time to reach the fittest allele in this case very different to the case with magnitude epistasis?

With regard to the advantage of sex addressed in this article, it should be noted that the probability of a sexual population fixing the superior allele first rather than the inferior allele is $P_{\text{sex}} \approx 1$, so the expected time to reach the fittest allele given that the inferior allele fixes first is all but irrelevant to sexuals. In contrast, this probability is appreciably nonzero for asexuals in a multilocus system (i.e., $P_{\text{asex}} < 1$). Accordingly, asexuals suffer the long waiting times to reach the fittest alleles given in this analysis but sexuals do not. (In principle, in scenarios in which the fittest allele differs from an inferior locally optimal allele by more than two mutations, this distinction is further increased, Watson and Jansen 2007).

When the superior allele fixes first
In case i, the fittest allele is XY. When the superior allele fixes first, the path to the fittest allele is $xy \rightarrow xY \rightarrow XY$, where both

transitions are supported by selection. In these calculations, we will let $s_{Xy} = s$ and $s_{xY} = 2s$. $T_{\text{sup}} = T_{xy \to xY} + T_{xY \to XY} = 1/4N\mu s + (1 + 2s)/2N\mu s$. In cases (2) to (4), the fittest allele is the superior allele xY, so when the superior allele fixes first $T_{\text{sup}} = T_{xy \to xY} = 1/4N\mu s$.

When the inferior allele fixes first
Case 1 (*no epistasis*)

$$T_{\text{inf}} = T_{xy \to xY} + T_{xY \to XY} = 1/2N\mu s$$
$$+ (1 + s)/4N\mu s \qquad \text{(two selected transitions)}$$

Case 2 (*sign epistasis but no local optima*)

$$T_{\text{inf}} = T_{xy \to xY} + T_{xY \to XY}$$
$$+ T_{XY \to xY} = 1/2N\mu s + (1 + s)/N\mu s$$
$$+ (1 + 3s/2)/N\mu s \qquad \text{(three selected transitions)}$$

Case 3 (*conditional neutrality*)

$$T_{\text{inf}} = T_{xy \to xY} + T_{xY \to XY} + T_{XY \to xY}$$
$$= 1/2N\mu s + 1/\mu + (1 + s)/2N\mu s$$
$$\text{(selected + neutral + selected transitions)}$$

Case 4 (*multiple optima*) (see eqs. 1–3, and the $N > N_{\text{critical}}$ condition, Weinreich and Chao 2005).

$$T_{\text{inf}} = T_{xy \to xY} + T_{xY \to XY \to xY}$$
$$= 1/2N\mu s + 1/16N\mu^2$$
$$\text{(selected + valley − crossing transitions)}$$

Summary

Case 1	Case 2	Case 3	Case 4
Given superior allele fixes first			
$1/4N\mu s +$ $(1+2s)/$ $2N\mu s$	$1/4N\mu s$	$1/4N\mu s$	$1/4N\mu s$
Given inferior allele fixes first			
$1/2N\mu s +$ $(1+s)/4N\mu s$	$1/2N\mu s +$ $(1+s)/$ $N\mu s + (1+$ $3s/2)/N\mu s$	$1/2N\mu s + 1/\mu +$ $(1+s)/2N\mu s$	$1/2N\mu s +$ $1/16N\mu^2$

If we assume $N \gg 1 \gg s \gg \mu$, then in cases where transitions that are not supported by selection are required we see that it is these transitions that dominate T_{inf}:

Thus in cases 1 and 2, where all transitions are supported by selection, T_{sup} and T_{inf} are both approximately proportional to $1/N\mu s$. So in these cases, the fact that sexuals are more likely to take T_{sup} rather than T_{inf} is of lesser consequence. But in case 3 and 4, T_{sup} remains proportional to $1/N\mu s$ but T_{inf} is proportional

Case 1	Case 2	Case 3	Case 4
Given superior allele fixes first			
$\approx 3/4N\mu s$	$1/4N\mu s$	$1/4N\mu s$	$1/4N\mu s$
Given inferior allele fixes first			
$\approx 3/4N\mu s$	$\approx 5/2N\mu s$	$\approx 1/\mu$	$\approx 1/16N\mu^2$

to $1/\mu$ in case 3 and $1/N\mu^2$ in case 4. Thus only in cases 3 and 4, where some transitions are required that are not supported by selection, can there be a significant difference between sexuals and asexuals (due to their having a different probability of taking the T_{sup} and T_{inf} paths).

To address the questions above that motivate this analysis: The possibility of a population moving against selection to escape the local optimum in case 4 has expected time approximately proportional to $1/N\mu^2$ whereas sexuals are likely to avoid this scenario and find the optimum in time proportional to $1/N\mu s$, a speedup proportional to s/μ. And the expected time to reach the fittest allele if the inferior allele fixes first is approximately proportional to $1/N\mu s$ in case 2 but approximately proportional to $1/\mu$ in case 3, so removing pathways that are supported by selection produces a difference in speed proportional to Ns even in the case in which neutral pathways are available. The former explains why the effect we model here is fundamentally different from the normal Fisher/Muller advantage of sex (see Fig. 2). The latter explains why true local optima are not required to see a significant advantage to sexuals (see Fig. 2, $s_{XY} = s_{Xy}$).

Appendix B
PARAMETRIC INVESTIGATIONS
Here, we investigate the sensitivity of the main result to changes in various parameters of the simulations. In each case, while varying one parameter, all other parameters are held at the default values: population size, $N = 10^5$, mutation rate, $\mu = 10^{-5}$, population mutation rate, $N\mu = 1$ (therefore), number of loci, $L = 50$, intergenic epistasis, $\varepsilon = 1$ (implicit, defined below), selection coefficient, $s = 0.02$ ($s_{Xy} = s$, $s_{XY} = 2s$, $s_{XY} = s/2$).

Varying population size, N
Figure B1. A shows no significant variation in P_{asex} until $N < 10{,}000$, and P_{sex} is still 1 even for $N = 1000$. $N = 200{,}000$ does not enable asexuals to increase P_{asex} noticeably.

Varying mutation rate, μ
Low mutation rates are sufficient for sexuals to succeed, although a mutation rate that is 10 times higher than our default begins to decrease the ability to converge on perfect genotypes. Higher
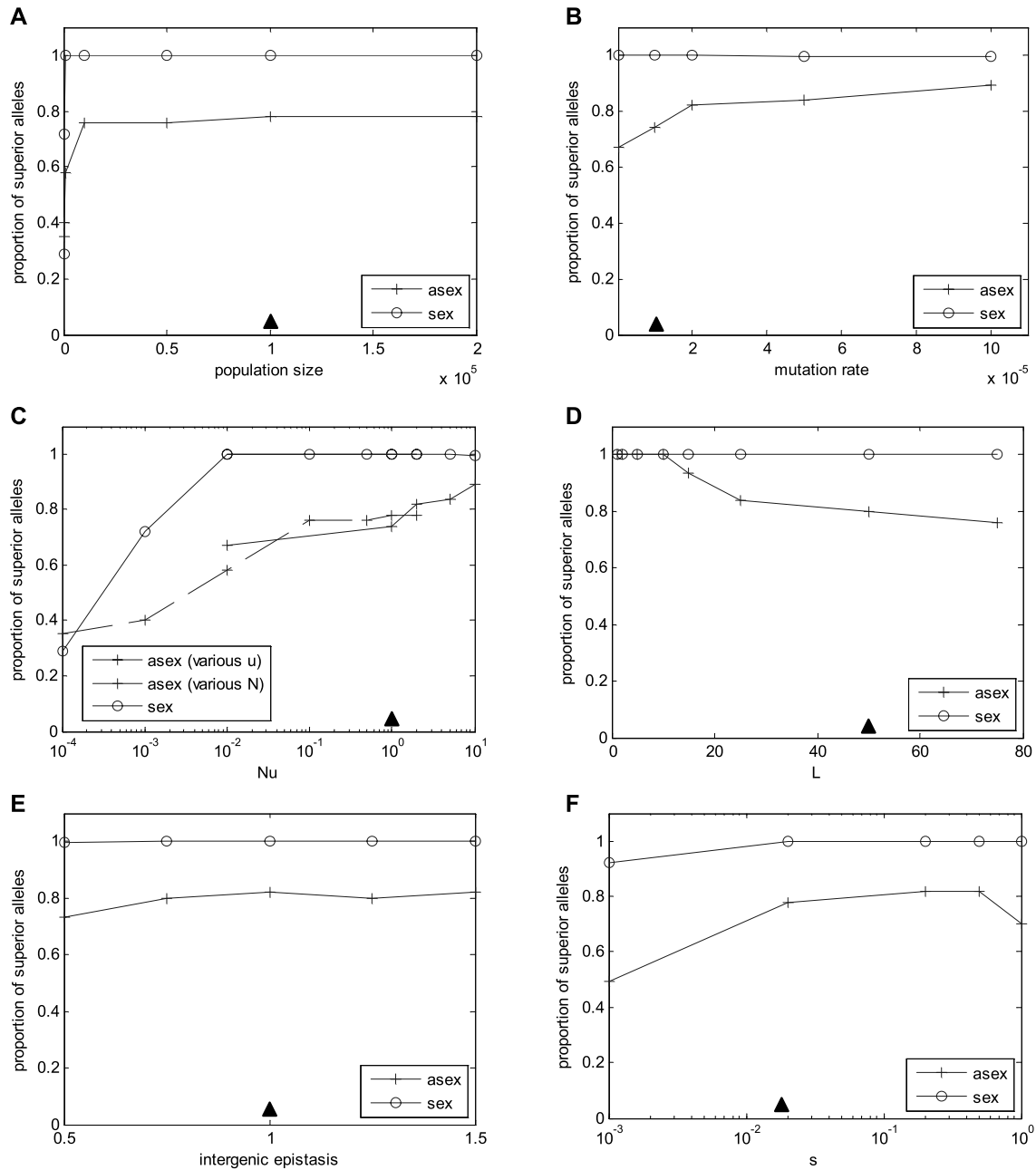
Figure B1. The proportion of superior alleles in sexual and asexual populations (i.e., P_{sex} and P_{asex}) after 3000 generations, with sensitivity to N, μ, $N\mu$, L, ε, s. (A) population size, $N = 10–2 \times 10^5$, (B) mutation rate, $\mu = 10^{-7}–10^{-4}$, (C) population mutation rate, $N\mu = 10^{-4}–10$, (D) number of loci, $L = 1–75$, (E) intergenic epistasis, $\varepsilon = 0.5$ to 1.5, (F) selection coefficient, $s = 10^{-3}–1$. In each case the default value of these parameters is indicated by the triangle marker on the horizontal axis. Only in the case of low number of loci (D) do we see asexuals succeeding in finding all superior alleles (see main text) in all other cases only sexuals exhibit $P = 1$. See discussion below.

mutation rates show a slightly higher proportion of superior alleles in asexuals but the probability of finding all 50 superior alleles with $P_{asex} = 0.9$ is still less than 1 in 100.

Varying population mutation rate, $N\mu$

$N\mu = 1$ (our default, given $N = 10^5$ and $\mu = 10^{-5}$) is in the region where a large difference is seen between sexuals and asexuals, but this region extends from $N\mu = 0.01$ to $N\mu = 0.5$.

(For $N\mu \leq 0.0001$ both sexuals and asexuals fail to find superior alleles reliably, for $N\mu = 10$ sexuals are beginning to fail to completely converge, hence P_{sex} has dropped very slightly from 1). The general consistency seen between results for $N\mu$ values created by varying μ (with $N = 10^5$) and by varying N (with $\mu = 10^{-5}$) suggests that it is the population mutation rate (allelic diversity) that determines the outcome rather than population size or mutation rate per se. (For the sexual data there

is no difference between data generated by varying μ and that generated by varying *N*).

Varying number of loci, L

Because the number of local optima is exponential in the number of loci, when the number of loci is sufficiently large (≥ 15), hitchhiking between loci prevents asexuals from fixing superior alleles reliably. Low *L* is the only case examined where asexuals succeed reliably, as predicted.

Varying intergenic epistasis, ε

Here the fitness of a genotype *G* with intergenic epistasis is $w'(G) = \exp(ln(w(G))^\varepsilon)$, where $w(G)$ is defined in the main text. Under this formulation, $\varepsilon = 1$ gives $w'(G) = w(G)$ that is, no epistasis. With $\varepsilon = 1$, the general structure of the log fitness landscape is linear, and creates positive or negative epistasis otherwise (Fig. B2). As predicted, intergenic epistasis shows no significant influence on the behavior of sexuals or asexuals; it is only the presence of local optima that is of consequence for our simulations. However, intergenic sign epistasis, where the genetic background at one locus may alter which allele was the superior allele at some other locus, would be another matter. (We note that the time for sexuals to reach $P_{sex} = 1$ is affected mildly by changes in ε, reducing from 1000 to 250 generations that is, greater positive epistasis increases the rate with which sexuals fix (all) beneficial alleles. Asexuals are similarly slower to get to the levels that they get to for lower ε, but in all but $\varepsilon = 0.5$ they have converged to suboptimal genotypes well before 3000 generations.)

Varying selection coefficient, s

The advantage of sexuals is robust to changes in *s*. We observe that at low *s*, neither sexual nor asexual populations are fully converged by 3000 generations. When *s* is high asexuals find less-fit genotypes (we speculate that this might be due to rapid loss of genotype diversity). But although the rate of accumulation of superior alleles varies, as one would expect, the advantage of sexuals is robust to changes in *s*.
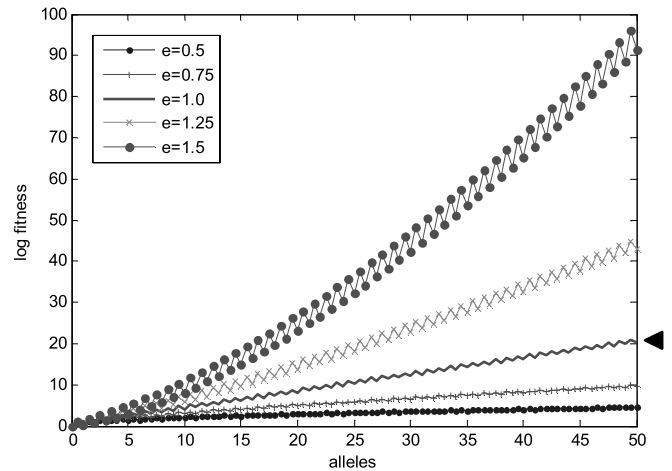

Figure B2. Log-fitness of sections through the fitness landscape for various values of intergenic epistasis. The centre line is linear in log fitness, that is, no epistasis, $\varepsilon = 1$, the default (triangle marker). Figure 1 of the main text shows a region of this $\varepsilon = 1$ curve.