Epistasis and the selective advantage of sex and recombination

Viviane M. de Oliveira* and Juliana K. da Silva
Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Dois Irmãos 52171-900, Recife, Pernambuco, Brazil

Paulo R. A. Campos†
Departamento de Física, Universidade Federal Rural de Pernambuco, Dois Irmãos 52171-900, Recife, Pernambuco, Brazil

(Received 1 April 2008; revised manuscript received 20 July 2008; published 5 September 2008)

The understanding of the central mechanisms favoring sex and recombination in real populations is one of the fundamental issues in evolutionary biology. Based on a previous stochastic formulation for the study of sex, here we aim to investigate the conditions under which epistasis favors the fixation of the sexual mode of reproduction in a given population. In addition, we try to identify the evolutionary forces which contribute to this process. One considers a finite population model which assumes the existence of a recombination modifier allele that can activate the recombination mechanism. We have found that sex is very little favored in a scenario of antagonistic epistasis, and this advantage only occurs in a narrow range of values of the selection coefficient s_d. On the other hand, synergistic epistasis favors recombination in a very broad domain. However, the major mechanism contributing to the spreading of the modifier allele depends on the range of values of s_d. At large s_d, background selection favors recombination since it increases the efficacy of selection, while at low s_d Muller's ratchet is the leading mechanism.

DOI: 10.1103/PhysRevE.78.031905 PACS number(s): 87.23.Kg, 87.10.Mn, 02.50.Ey

I. INTRODUCTION

Several issues remain unsolved in evolutionary biology. Of the greatest open questions in evolutionary biology, we could mention the maintainance of genetic diversity [1–3], the origin of life [4–8], and the evolution of sex, recombination, and epistasis [9–15]. In the current work we address the last point. Several theories have been formulated to understand the evolution of sex and recombination [16]. One of the intriguing challenges in this issue is to understand how sexual replication became the predominant mode of reproduction, despite the current theories not ensuring selective advantage of sexual reproduction in all evolutionary scenarios.

One of the theories invoked to explain the evolution of sex and recombination is based on Muller's ratchet hypothesis [17,18]. Muller's ratchet is an evolutionary process that has been implicated in the extinction of asexual species, the evolution of mitochondria, the degeneration of the Y chromosome, and the evolution of microbes. The accumulation of deleterious mutations by the loss of the fittest individuals in an asexual population is known as Muller's ratchet [18]. In an infinitely large population that is subjected to natural selection and deleterious mutations, the fraction of individuals which are mutation-free is given by $\exp(-U_d/s_d)$ [19], when a multiplicative fitness landscape is assumed, where U_d is the deleterious mutation rate and s_d is the selective parameter. In a finite population, stochastic events become relevant in determining population evolution, and their intensities are proportional to $1/N$, where N is the population size. In a scenario where back mutations are negligible, which is expected to hold on large genomes, and for small population sizes, a continuous influx of deleterious mutations will first lead to the loss of the mutation-free individuals, and subsequently to a continuous extinction of the most adapted individuals in the population. In this way, recombination is expected to have a crucial role in adaptation since it enables that individuals carrying segregating deleterious mutations can recombine to form a better adapted one. This latter feature manifests even more intensely when advantageous mutations occur, because recombination can bring together those beneficial mutations arising in distinct lineages, thus reducing clonal interference [18,20–22].

Recently, some models have been proposed to investigate the circumstances that enable the sexual mode of replication to bring an evolutionary profit. Interesting stochastic approaches have been formulated [11,12]. These formulations propose finite population models to identify the mechanisms which favor the fixation of a modifier recombination allele in a given population. Starting from a completely asexual population, these models assume that a single recombination modifier allele, which increases the rate of recombination or activates recombination, invades the asexual population, and then the number of events at which the modifier allele has become successful is counted. By "successful" one means that the modifier allele has spread through the whole population. Gordo and Campos have demonstrated that Muller's ratchet is the major evolutionary force dictating the advantage of recombination [11], whereas Keightley and Otto [12] argue that it is Hill-Robertson interference [23] that favors sex. The Hill-Robertson effect describes that the linkage between selected sites reduces the efficacy of selection in finite populations. Regardless of the framework, both works have found an appreciable advantage for recombination relative to a neutral allele, and this gain becomes larger as one considers larger population sizes.

*viviane@deinfo.ufrpe.br
†paulo.campos@df.ufrpe.br

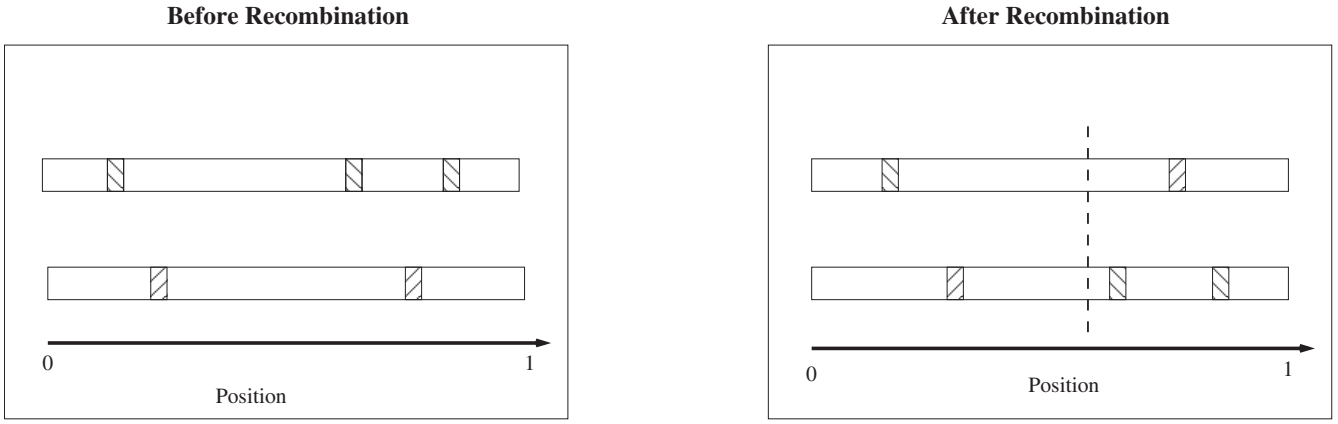

FIG. 1. Snapshot of how recombination works. In this instance, the individual 1 (upper) has three deleterious mutations and individual 2 (lower) has two deleterious mutations. The positions of these five mutations are completely distinct. Remember that the position of each mutation is a real number in the interval (0,1]. The position of the genetic exchange is indicated by the vertical dashed line. Those mutations apart from that line will be swapped between the two individuals. Hence, after recombination individual 1 receives one deleterious mutation from individual 2 and transfers two mutations to individual 2.

Based on the model proposed by Gordo and Campos, we will study how epistatic interactions between genes can affect the scenarios established by these previous investigations. The assumption that genes have independent effects on adaptation is very simplifying [24,25]. Actually, many mutations interact with one another in complex ways. For instance, the combined effects of two or more mutations on fitness can be much greater or much smaller than predicted from their individual effects. These deviations from the expectation based on the assumption of independent effects are called epistasis.

Our paper is organized as follows. Section II describes the model we use in our simulations. Section III presents our simulation results, and finally in Sec. IV we present our conclusions.

II. THE MODEL

Our computer simulations follow the standard Wright-Fisher model, i.e., most adapted individuals have a higher chance of producing offspring. The model assumes a population of constant size N that evolves according to the following life cycle: mutation, selection, and recombination. Each organism is represented by an infinitely large genome $S=(s_1,s_2,\ldots,s_\infty)$, where each site s_α can take two possible values $s_\alpha=0$ (original state) or 1 (a mutation has occurred). This situation corresponds to the classical infinite-site model introduced by Watterson [26]. The assumption of the infinite-site model is very appropriate when the genome size is large. For instance, the HIV genome has 9749 nucleotides—about the same size as other retroviruses. For *Escherichia Coli*, the genome size is about 5 Mega base pairs, and it contains about 4000 genes, and so this approximation seems quite reasonable.

The adaptation value of any individual is solely a function of the number of deleterious mutations, k, and it has the simple form

$$\omega_k = (1 - s_d)^{k^\alpha}, \tag{1}$$

where α is the epistasis parameter, and s_d is the selection coefficient. The case $\alpha=1$ corresponds to a multiplicative fitness landscape, where each mutation has an independent effect on the adaptation of the organism. When $\alpha>1$ (synergistic epistasis) or $\alpha<1$ (antagonistic epistasis), the effect of each new mutation on adaptation depends on the previous number of mutations. For $\alpha>1$ each newly added mutation has an even stronger effect, while for $\alpha<1$, each newly added mutation has its effect attenuated. Of course, Eq. (1) is a rather simplified way to estimate fitness when taking into account interactions among genes. Actually, quantitative effects of epistasis have been difficult to discern because they are difficult to estimate [27]. In some environments there is a trade-off between s_d and α, so that one can only be optimized at the expense of the other [28]. So, when one estimates fitness as in Eq. (1), one basically means that there is a bias toward antagonistic or synergistic epistasis.

During reproduction, an offspring inherits all the mutations from its parent plus an additional amount n which is obtained from a Poisson distribution of the mean U_d, i.e.,

$$P(n) = \frac{e^{-U_d} U_d^n}{n!}. \tag{2}$$

U_d is the rate of deleterious mutations. When a new mutation is generated its position in the genome is taken from a uniform distribution in the interval (0, 1].

After reproduction and mutation, recombination takes place. Recombination proceeds as follows. We form $N/2$ pairs of individuals and then check for each pair whether they have the modifier allele of recombination. If only one individual carries the modifier allele then they recombine with probability $r/2$. In the case that both organisms share the modifier allele then they recombine with probability r. Once recombination occurs, we randomly determine the position for the genetic exchange from a uniform distribution in the interval (0, 1]. In Fig. 1 we show a snapshot of the recombination scheme.

III. RESULTS

In the following we will present our simulation results. The main quantity of interest in our study is the relative

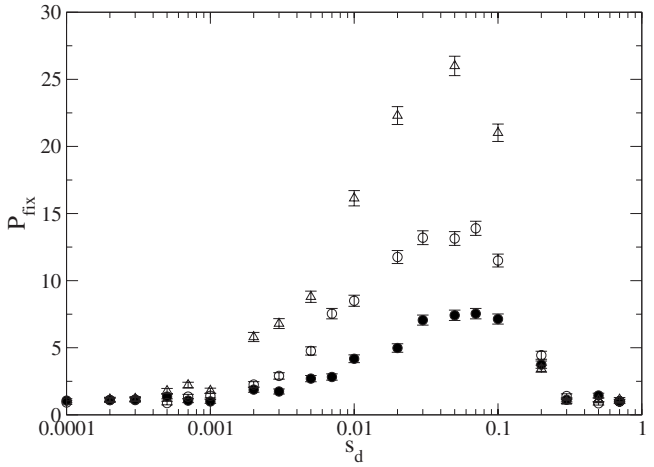

FIG. 2. Relative probability of fixation of the recombination modifier allele as a function of the selective parameter s_d. The distinct curves correspond to different population sizes N. The parameter values are $N=500$ (solid circles), 1000 (empty circles), and 2000 (triangles), $U_d=1.0$, $r=0.5$, and $\alpha=1.0$ (multiplicative fitness landscape). The data points are the results for $50N$ simulations.

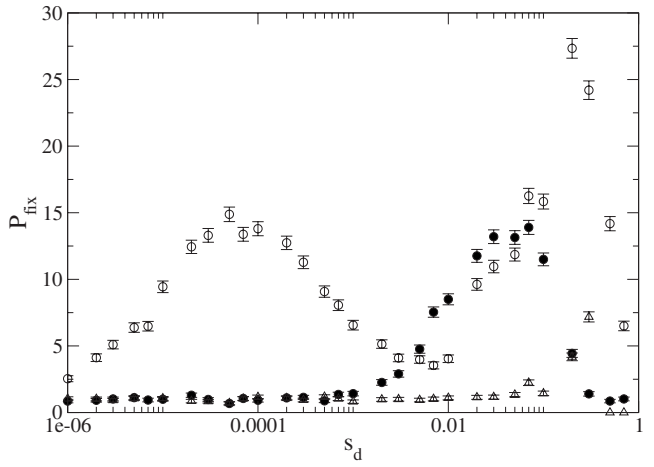

FIG. 3. Relative probability of fixation of the recombination modifier allele as a function of the selective parameter s_d. The distinct curves correspond to different values of epistasis α: $\alpha=0.5$ (empty triangles), 1.0 (filled circles), and 2.0 (empty circles). The other parameters are $N=1000$, $U_d=1.0$, and $r=0.5$. The data points are the results for $50N$ simulations.

probability of fixation of the recombination modifier allele. The modifier allele reaches fixation when it is inherited by every organism in the population. We find the relative probability of fixation of the modifier allele, P_{fix}, by calculating the fraction of runs in which the modifier allele has become fixed divided by the probability of fixation of a neutral allele, which for a population of constant size N is $1/N$.

Before introducing the single modifier allele, we let the population evolve for 1000 generations. At generation $t=1000$ a randomly chosen individual receives the modifier allele, and then the population evolves up to fixation or complete loss of the modifier allele.

Figure 2 shows the relative probability of fixation of the modifier allele as a function of the selective parameter s_d. In this figure, we present the results for the multiplicative fitness landscape, $\alpha=1$, and for fixed value of mutation rate $U_d=1.0$ (this value of mutation rate is compatible with estimated genomic mutation rates for *Drosophila melanogaster* [29] and RNA viruses [30]). The different curves denote distinct population sizes N. Because the probability of fixation of a neutral mutation is $1/N$, a larger population size requires a more extensive statistical analysis and consequently a larger computational cost. For each population of size N, we have simulated $50N$ independent runs. As we can observe, the advantage of recombination is prominent at intermediate values of the selection coefficient. At large values of s_d (strong selection regime), the loss of the most adapted individuals is very unlikely. In this regime Muller's ratchet does not act. For instance, when $s_d>0.3$ no advantage for recombination is obtained. At intermediate and very small values of s_d, Muller's ratchet is effective, resulting in a continuous accumulation of deleterious mutations. Although the advantage of the recombination modifier allele vanishes in the very weak selection regime. Gordo and Campos [11] have shown that the benefit of recombination disappears around $s_d=1/N$, where selection ceases to be effective. Furthermore, they have demonstrated that the range in which recombination brings an evolutionary profit coincides with the interval in which recombination can stop the ratchet, and the position of the maximum value of P_{fix} occurs at the point at which $d\omega/dt$ is maximized in the corresponding asexual population.

From now on, we are going to investigate how epistasis alters the previous scenario. Figure 3 shows the relative probability of fixation, P_{fix}, for some values of epistasis α. At a first glance, the results suggest that epistasis strongly changes the previous scenario for the multiplicative fitness landscape, both qualitatively and quantitatively. One may distinguish two situations: synergistic epistasis ($\alpha>1$) and antagonistic epistasis ($\alpha<1$). For antagonistic epistasis, we see that the modifier allele is favored in a very narrow interval of s_d, where selection is very strong. In addition, the maximum advantage reached by the allele is also reduced compared to the multiplicative case ($\alpha=1$). In contrast, synergistic epistasis displays a different scenario, where the modifier recombination allele is favored for almost all the range of s_d considered in the figure. The relative probability P_{fix} is now a double-peaked function, being more pronounced at very large values and also at small values of the selective parameter. The size of the peak placed at high s_d is bigger than in the multiplicative case, whereas the size of the peak located at small values of s_d has about the same amplitude as for $\alpha=1$.

In order to better comprehend how epistasis alters the fate of the modifier alleles, we next compare how mean population fitness evolves. The mean population fitness is estimated as

$$\bar{\omega}=\frac{1}{N}\sum_{i=1}^{N}\omega_i. \tag{3}$$

The mean population fitness as a function of time is shown in Figs. 4 and 5. In these plots, we compare the trajectory of the

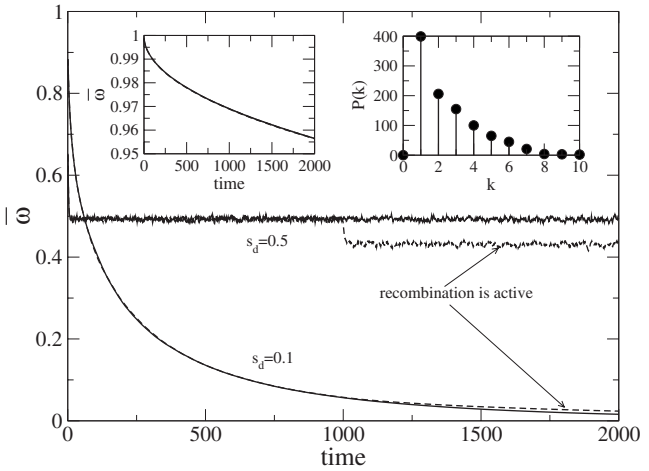

FIG. 4. Mean population fitness as a function of time. The solid line is the mean fitness for an asexual population and the dashed line corresponds to a population subjected to recombination. The parameters are $N=1000$, $U_d=1.0$, $r=0.5$, and $s_d=0.1$ (lower curves), 0.5 (upper curves), and 1×10^{-3} (left inset). The inset in the right of the figure shows the distribution of deleterious mutations in the population at generation $t=1000$.

mean fitness of an asexual population with one in which the sexual mode of reproduction is activated at generation $t=1000$, i.e., every individual receives the modifier allele. The aim is to check the change in fitness produced by the change of the mode of reproduction. In Fig. 4 we have considered $\alpha=0.5$, which corresponds to antagonistic epistasis. In the inset of the figure the parameters are $N=1000$, $U_d=1.0$, $s_d=10^{-3}$, and $r=0.5$, and in this case we cannot distinguish the fitness trajectories, which means that the gain in fitness brought about by recombination is negligible. In the main plot, we show simulation results for $s_d=0.1$ and 0.5, keeping all other parameters constant. One observes that, when $s_d=0.1$, the increase of fitness due to recombination is small, and the ratchet still continues to click, sustaining the accumulation of deleterious mutations. In a small region of s_d ($s_d=0.1-0.4$) there is a slight increase of fitness due to recombination, which promotes the fixation of the modifier allele more frequently than expected for a neutral allele. Strikingly, when the population is subjected to very strong selection, for instance $s_d=0.5$ (as shown in the figure), recombination has a deleterious effect, since it produces less adapted individuals. This can be understood by looking at the distribution of the number of mutations, $P(k)$, at generation $t=1000$, which is the moment at which recombination is activated (see the inset on the right of Fig. 4). Although most individuals have one deleterious mutation, the distribution is very broad, and so individuals carrying two, three, or even more mutations can coexist in the population at high frequencies. When those individuals in the less adapted classes recombine with those in the one-mutation class, from this exchange results individuals with an intermediate number of mutations. For antagonistic epistasis this is wasteful, since the deleterious effect of each new mutation is larger when the individual has fewer mutations.

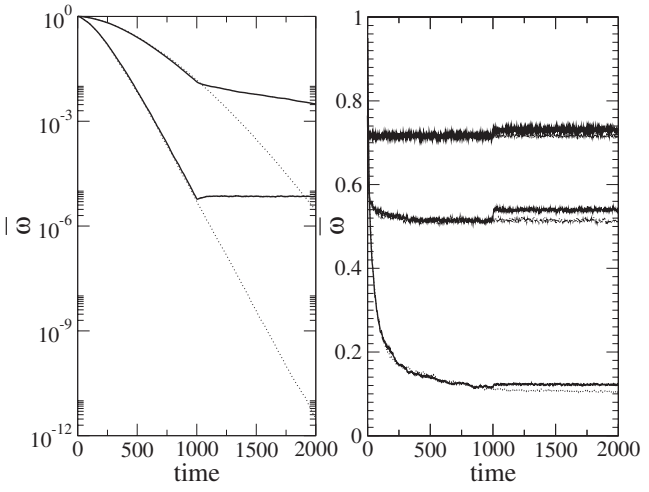

FIG. 5. Mean population fitness as a function of time. The dashed line is the mean fitness for an asexual population and the solid line corresponds to a population subjected to recombination. The parameters are $N=1000$, $U_d=1.0$, and $r=0.5$. Left panel shows $\bar{\omega}$ for the weak selection regime, where $s_d=1\times10^{-5}$ (upper curves) and 1×10^{-4} (lower curves). In the right panel, we have considered $s_d=0.01$ (bottom curves), 0.1 (intermediate curves), and 0.5 (top curves).

A more complex scenario emerges for synergistic epistasis. In order to perceive the double-peaked shape for P_{fix}, we investigate how $\bar{\omega}$ evolves at the two distinct regions of the selective parameter s_d where a noticeable advantage for recombination was observed. Figure 5 shows $\bar{\omega}$ as a function of time for five values of s_d: $s_d=1\times10^{-5}$ and 1×10^{-4} (left panel), and 1×10^{-2}, 0.1, and 0.5 (right panel). Once more, we compare fitness evolution for the situations in which recombination is acting with the ones where recombination is inactive. Viewing the left panel, one clearly notices the continuous decline of fitness in the absence of recombination (dotted lines), the drop in fitness being faster for $s_d=10^{-4}$. From the same plot, we also observe that recombination prevents Muller's ratchet when $s_d=1\times10^{-4}$, while for $s_d=10^{-5}$ it has the effect of slowing down the ratchet but not ultimately stopping it. The point $s_d=1\times10^{-4}$ is around the position of the maximum advantage enjoyed by the modifier allele, as seen from Fig. 3, and coincides with the fastest fitness decline (data not shown), which supports the argument given by Gordo and Campos [11] that Muller's ratchet is the major mechanism favoring recombination in this regime of selection. On the other hand, the success of recombination in reaching fixation at large values of s_d cannot be attributed to Muller's ratchet, since in this regime the ratchet does not click even in the absence of recombination (right panel of Fig.5). Nevertheless, an evolutionary advantage is gathered by the recombination allele in this second regime since it increases the efficacy of background selection, i.e., the rate at which deleterious mutations are eliminated from the population. This is clear when we plot together the mean population fitness when recombination is active and when it is not. In these examples, recombination increases the fitness immediately after its activation.

For the sake of completeness, now we will assume that beneficial mutations can also take place at a constant rate U_b during reproduction. The selective effect s_b of each beneficial

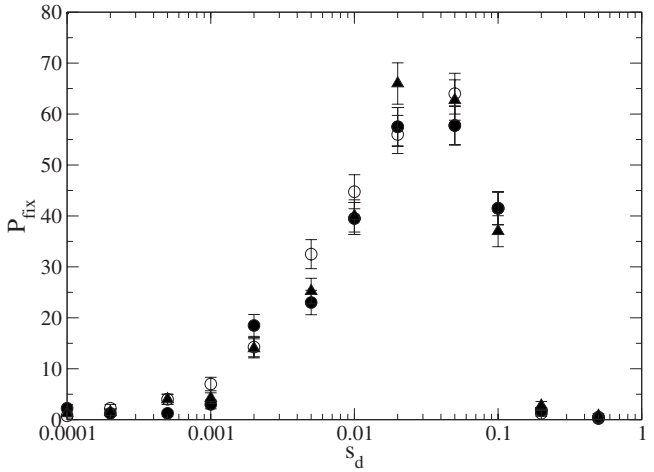

FIG. 6. Effect of beneficial mutations on P_{fix}. The parameter values are $N=5000$, $r=0.5$, $U_d=1.0$, $s_b=0.01$, and $U_b=1\times10^{-6}$ (filled circles), 1×10^{-5} (empty circles), and 1×10^{-4} (filled diamonds).

mutation is kept constant. As before, Fig. 6 displays the probability of fixation of the recombination modifier allele as a function of s_d, but now considering three distinct values of mutation rate U_b. One can perceive that the advantage brought about by the influx of advantageous mutations is not comparable to those promoted by Muller's ratchet and/or the Hill-Robertson effect.

IV. CONCLUSIONS

In this work we have surveyed the role of epistatic interactions in the evolution of sex and recombination. We have performed extensive simulations of a recently proposed model [11] to investigate the circumstances under which recombination brings an evolutionary advantage, assuming independent effects of mutations. Conversely, our study assumes that mutations interact and so their effects are no longer independent. The statistical analysis relies on determining how the nature of these interactions, antagonistic or synergistic, changes the scenario presented for the multiplicative fitness landscape.

By studying the relative probability of fixation of the modifier allele of recombination, we have seen that the selective advantage of recombination for antagonistic epistasis is very small and restricted to a narrow region of large values of the selective parameter s_d, where recombination confers a slight increase of fitness. On the other hand, synergistic epistasis extends the evolutionary advantage of recombination to a very broad domain, when compared to the non-epistasis situation. Depending on the value of s_d, the advantage of recombination is brought about by different evolutionary mechanisms. For small values of s_d, Muller's ratchet is the leading mechanism providing fixation of the modifier allele, since in this regime the recombination is very effective in stopping the ratchet, or even slowing down the fitness decline. This phase corresponds to the one observed favoring recombination in the multiplicative fitness landscape. The great difference is that under synergistic epistasis this phase is shifted toward lower values of the selection coefficient. However, another phase emerges in which a broad and large peak appears at intermediate and large values of s_d. In this region it is not Muller's ratchet that contributes to promote fixation of the recombination allele, but background selection, which becomes more efficient due to recombination by breaking the linkage between deleterious mutations and reducing the strength of the Hill-Robertson interference. The maximum advantage reached by the recombination allele is even higher than that seen in the Muller's ratchet phase. So far, in synergistic epistasis we can conclude that actually both Muller's ratchet and Hill-Robertson interference are mechanisms responsible for the advantage of sex.

The main message of our work is to show that epistasis affects the success of recombination in becoming fixed in populations, by the action of different evolutionary mechanisms which determine its advantage: Muller's ratchet and background selection. This finding gives a contribution in showing directions in which recombination could have emerged. Of course, we still need to identify whether other evolutionary mechanisms can display similar roles in promoting recombination, and also the conditions under which all these evolutionary mechanisms can be manifest. In this context, investigations on more complex models which incorporate population structure and variable effects of mutations are welcome. Recent studies show that the ratchet clicks faster in subdivided populations [31]. Another important feature of natural populations is that the environments they inhabit are not homogeneous but fluctuate in time and space. Empirical evidence shows that several organisms engage in a sexual model of reproduction when the environmental conditions are stressful [32,33].

ACKNOWLEDGMENTS

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the program PRONEX/MCT-CNPq-FACEPE. J.K.S. was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). The authors acknowledge Isabel Gordo for a critical reading of this manuscript.

[1] F. Tajima, Genetics **123**, 585 (1989).
[2] B. Charlesworth, Mol. Biol. Evol. **15**, 538 (1998).
[3] I. Gordo and P. R. A. Campos, BMC Evol. Biol. **7**, 116 (2007).
[4] M. Eigen, Naturwiss. **58**, 465 (1970).
[5] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[6] M. R. Michod, Am. Zool. **23**, 15 (1983).
[7] P. R. A. Campos, J. F. Fontanari, and P. F. Stadler, Phys. Rev.

E **61**, 2996 (2000).
[8] D. Alves, P. R. A. Campos, A. T. C. Silva, and J. F. Fontanari, Phys. Rev. E **63**, 011911 (2000).
[9] J. S. Maynard, *The Evolution of Sex* (Cambridge University Press, Cambridge, U.K., 1978).
[10] A. S. Kondrashov, J. Hered **84**, 372 (1993).
[11] I. Gordo and P. R. A. Campos, Genetics **179**, 621 (2008).
[12] P. Keightley and S. P. Otto, Nature (London) **443**, 89 (2006).
[13] E. Tannenbaum, Phys. Rev. E **77**, 011915 (2008).
[14] B. Lee and E. Tannenbaum, Phys. Rev. E **76**, 021909 (2007).
[15] E. Tannenbaum, Phys. Rev. E **73**, 061925 (2006).
[16] S. A. West, C. M. Lively, and A. F. Read, J. Evol. Biol. **12**, 1003 (1999).
[17] H. J. Muller, Mutat Res. **1**, 2 (1964).
[18] J. Felsenstein, Genetics **78**, 737 (1974).
[19] J. Haigh, Theor Popul. Biol. **14**, 251 (1978).
[20] T. F. Cooper, PLoS Biol. **5**, 1899 (2007).
[21] R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930).
[22] H. J. Muller, Am. Nat. **66**, 118 (1932).
[23] W. G. Hill and A. Robertson, Genet. Res. **8**, 269 (1966).
[24] S. F. Elena and R. E. Lenski, Nature (London) **390**, 395 (1997).
[25] N. Beerenwinkel, L. Pachter, B. Sturmfels, S. F. Elena, and R. E. Lenski, BMC Evol. Biol. **7**, 60 (2007).
[26] G. A. Watterson, Theor Popul. Biol. **7**, 256 (1975).
[27] D. S. Falconer, *Introduction to Quantitative Genetics* (Longman, New York, 1989).
[28] C. O. Wilke and C. Adami, Proc. R. Soc. London, Ser. B **268**, 1469 (2001).
[29] P. D. Keightley, Genetics **138**, 1315 (1994).
[30] J. W. Drake and J. J. Holland, Proc. Natl. Acad. Sci. U.S.A. **96**, 13910 (1999).
[31] P. R. A. Campos, J. Combadão, F. Dionisio, and I. Gordo, Phys. Rev. E **74**, 042901 (2006).
[32] M. Abdullah and R. H. Borts, Proc. Natl. Acad. Sci. U.S.A. **98**, 14524 (2001).
[33] I. Grishkan, A. Koroi, E. Nevo, and S. P. Wasser, Proc. R. Soc. London, Ser. B **270**, 13 (2003).