DOI: 10.1534/genetics.108.089706

Hitchhiking Both Ways: Effect of Two Interfering Selective Sweeps on Linked Neutral Variation

Luis-Miguel Chevin,*,†,[1] Sylvain Billiard[‡] and Frédéric Hospital[§]

**UMR de Génétique Végétale, Ferme du Moulon, 91190 Gif sur Yvette, France, †Ecologie, Systématique et Evolution, UMR 8079, Université Paris-Sud, 91405 Orsay Cedex, France, §INRA, UMR1236 Génétique et Diversité Animales, 78352 Jouy-en-Josas, France and ‡Génétique et Evolution des Populations Végétales, Université des Sciences et Technologies de Lille 1, F-59655 Villeneuve d'Ascq Cedex, France*

Manuscript received March 28, 2008
Accepted for publication May 15, 2008

ABSTRACT

The neutral polymorphism pattern in the vicinity of a selective sweep can be altered by both stochastic and deterministic factors. Here, we focus on the impact of another selective sweep in the region of influence of a first one. We study the signature left on neutral polymorphism by positive selection at two closely linked loci, when both beneficial mutations reach fixation. We show that, depending on the timing of selective sweeps and on their selection coefficients, the two hitchhiking effects can interfere with each other, leading to less reduction in heterozygosity than a single selective sweep of the same magnitude and more importantly to an excess of intermediate-frequency variants relative to neutrality under some parameter values. This pattern can be sustained and potentially alter the detection of positive selection, including by provoking spurious detection of balancing selection. In situations where positive selection is suspected *a priori* at several closely linked loci, the polymorphism pattern in the region may also be informative about their selective histories.

THE search for molecular signatures of positive selection has been a matter of intense research and applications in the recent years, motivated by the hope to better understand the genetic bases of adaptation and the recent history of populations (Bamshad and Wooding 2003; Nielsen *et al.* 2007). The footprints of positive selection on neutral polymorphism are the consequence of the hitchhiking effect (Maynard Smith and Haigh 1974), and current methods to detect them encompass two main approaches. The first one is genome scans of neutral variation and is a top-down process. It consists of gathering polymorphism data widely distributed throughout the genome and summarizing them with a particular measure, be it the nucleotide diversity, the frequency spectrum of mutations (Nielsen *et al.* 2005), or the length and frequency of haplotypes [for ongoing selective sweeps (Sabeti *et al.* 2002; Voight *et al.* 2006)]. The loci exhibiting extreme values in the distribution of the measure are then considered as putative targets of positive selection (but see Teshima *et al.* 2006 for caveats of this method). The second approach, the candidate-gene approach, is a bottom-up process in which one wishes to test some evolutionary scenarios, for instance, for a gene (or QTL) of known function (see, for instance, Edelist *et al.* 2006). It consists of analyzing neutral polymorphism at a finer scale (of the order of the megabase or lower), to test if positive selection occurred, and to infer some parameters of the selective sweep such as the target and strength of selection. This fine-scale analysis can also be carried out in regions identified after a genome scan (a good example of this kind is Pool *et al.* 2006; for a more comprehensive review see Thornton *et al.* 2007). Here, we focus on this finer-scale analysis of polymorphism.

The most popular method for the fine-scale analysis of selective sweeps uses the information at several markers distributed in the small region of interest, to perform a composite likelihood-ratio test on the frequency spectrum (Kim and Stephan 2002), to jointly estimate the parameters of the selective sweep and the relative likelihood of selection *vs.* neutrality. This can be followed by a goodness-of-fit test to confirm the robustness of the estimated parameters against several demographic scenarios (Jensen *et al.* 2005). Though efficient, this method can be affected by ascertainment biases (Thornton and Jensen 2007). Moreover, some factors—*e.g.*, differences in recombination or mutation rates between the two sides of a selective sweep—can modify the fine-scale polymorphism pattern around the selective sweep in a systematic way (*i.e.*, nonstochastically). Here, we focus on one particular modifying factor, namely the presence of another locus under positive selection in the region of influence of a selective sweep. We wish to understand how the effect of a focal selective sweep is modified by the presence of another selective sweep in its vicinity.

Simultaneous positive selection at several linked loci was repeatedly reported for asexuals (Notley-McRobb and Ferenci 2000; Perfeito *et al.* 2007), where it was

[1]*Corresponding author:* UMR 8079, Bât. 360, Université Paris-Sud, 91405 Orsay Cedex, France. E-mail: luis-miguel.chevin@u-psud.fr

termed "clonal interference." The effect of such interference on probabilities of fixation in asexuals was described theoretically by Gerrish and Lenski (1998). In sexuals, positive selection at two closely linked loci not only decreases their probabilities of fixation (Barton 1995), but also builds up negative linkage disequilibrium between them (Hill and Robertson 1966; Felsenstein 1974) and slows down their dynamics, which altogether is called the Hill–Robertson effect. Notably, the overlap in time of positive selection at partially linked loci, with (Barton 1995) or without (Roze and Barton 2006) epistasis, is invoked in all population genetic models of the evolution of sex. This phenomenon is difficult to characterize empirically in natural populations. One of the reasons is that selection at two closely linked loci may be difficult to detect through its signature on neutral polymorphism without *a priori* information, since the signatures of both loci may be confounded. Moreover, the lack of knowledge about the effect of two interfering selective sweeps on neutral polymorphism makes it difficult to look for such signatures. Yet, in cases where one *a priori* suspects recent selection at two closely linked loci, signatures of selection can be found. This was done in two recent studies. The first one concerns two genes involved in sex-ratio distortion in *Drosophila simulans* (Derome *et al.* 2008). The second one deals with the domestication gene *Tb1* and the early-flowering gene *dwarf8* in maize (Camus-Kulandaivelu *et al.* 2008). These two studies at least suggest that successful selective sweeps at two tightly linked loci can occur in natural populations.

Some models where several selective forces interact on neutral polymorphism were published in the last decade. Kim and Stephan (2000) investigated the joint effects of positive and negative selection on neutral polymorphism and showed that the hitchhiking effect dominates in regions of low recombination, whereas background selection primarily explains the levels of neutral heterozygosity in regions with higher recombination. Kim and Stephan (2003) studied the hitchhiking effects of two selective sweeps that overlap in time, that is, the interplay of positive selection at two loci. Their aim was mainly to assess whether predictions made under the assumption that selective sweeps do not overlap still hold when there is at least partial overlap. They showed that because of the selective interference between the loci under selection, (i) their time to fixation increases, which leaves more time for recombination with the neutral locus to occur, and (ii) the probability of fixation is decreased for each beneficial mutation. The net effect is an overall decrease of the effect of selective sweeps relative to the case without interference.

Here, we further study the interplay of positive selection at two closely linked loci, with a different perspective. We focus on cases in which both beneficial mutations escape stochastic loss and get fixed and ask what the resulting pattern of neutral polymorphism is in the region. We want to know how a successful selective sweep at a linked locus alters the signature of a focal selective sweep on heterozygosity and on the site-frequency spectrum. We also investigate whether or not there is a particular signature of the action of two close selective sweeps and whether neutral polymorphism can carry information about the history of adaptive selection at two loci. We show that the interference of two selective sweeps can dramatically affect the signatures of positive selection, in particular by inducing an excess of intermediate-frequency variants in the frequency spectrum. This may paradoxically hinder our ability to detect adaptive selection in regions of the genome where it was most experienced.

DETERMINISTIC MODEL

Let us first use a deterministic argument to introduce the problem. We want to calculate the change in allelic frequencies at a neutral locus neu under the influence of hitchhiking effects from two loci under positive selection, sel_1 and sel_2, with selection coefficients s_1 and s_2, respectively. We assume that all loci are biallelic. The frequencies of the beneficial alleles at sel_1 and sel_2 are denoted p_{sel_1} and p_{sel_2}, respectively, and we denote p_{neu} the frequency of an arbitrarily chosen neutral allele at neu. The recombination rate between any pair of loci $\{l, m\}$ is denoted $r_{l,m}$. The fitness of an individual carrying X_{sel_1} copies of the beneficial allele at sel_1 ($X_{\mathrm{sel}_1} = 0, 1,$ or 2) and X_{sel_2} ($X_{\mathrm{sel}_2} = 0$, 1, or 2) copies of the beneficial allele at sel_2 is

$$W(X_{\mathrm{sel}_1}, X_{\mathrm{sel}_2}) = (1 + X_{\mathrm{sel}_1} s_1)(1 + X_{\mathrm{sel}_2} s_2); \qquad (1)$$

that is, we assume that fitness is additive within each locus and multiplicative between loci. The changes in frequencies at all loci due to selection at both sel_1 and sel_2, as well as other relevant parameters, were derived by exact recursions (see appendix a).

We denote C_U the linkage disequilibrium between a locus set U, defined as the covariance of their allelic states as in Barton and Turelli (1991) (see appendix a). The changes in frequencies at the selected loci can be written in a general form as

$$\Delta p_{\mathrm{sel}_i} = \frac{s_i p_{\mathrm{sel}_i} q_{\mathrm{sel}_i} + s_j C_{\mathrm{sel}_i,\mathrm{sel}_j} + s_i s_j (C_{\mathrm{sel}_i,\mathrm{sel}_j} + 2 p_{\mathrm{sel}_i} q_{\mathrm{sel}_i} p_{\mathrm{sel}_j})}{\overline{W}}, \qquad (2a)$$

where sel_i stands for the focal selected locus (sel_1 or sel_2) and sel_j stands for the other selected locus (sel_2 or sel_1, respectively), and $q_{\mathrm{sel}_i} = 1 - p_{\mathrm{sel}_i}$. The change in frequency at the neutral locus is

$$\Delta p_{\mathrm{neu}} = \frac{s_1 C_{\mathrm{sel}_1,\mathrm{neu}} + s_2 C_{\mathrm{sel}_2,\mathrm{neu}} + s_1 s_2 (2 p_{\mathrm{sel}_1} C_{\mathrm{neu},\mathrm{sel}_2} + 2 p_{\mathrm{sel}_2} C_{\mathrm{neu},\mathrm{sel}_1} + C_{\mathrm{neu},\mathrm{sel}_1,\mathrm{sel}_2})}{\overline{W}} \qquad (2b)$$

with

$$\overline{W} = 1 + 2s_1 p_{\mathrm{sel}_1} + 2s_2 p_{\mathrm{sel}_2} + 2s_1 s_2 (C_{\mathrm{sel}_1,\mathrm{sel}_2} + 2p_{\mathrm{sel}_1} p_{\mathrm{sel}_2}). \tag{2c}$$

Assuming $s_1 \ll 1$ and $s_2 \ll 1$, such that the terms involving products of the selection coefficients can be neglected, Equation 2 can be rewritten:

$$\Delta p_{\mathrm{sel}_i} \simeq s_i p_{\mathrm{sel}_i} q_{\mathrm{sel}_i} + s_j C_{\mathrm{sel}_i,\mathrm{sel}_j}, \quad \{i,j\} = \{1,2\} \text{ or } \{2,1\} \tag{3a}$$

$$\Delta p_{\mathrm{neu}} \simeq s_1 C_{\mathrm{sel}_1,\mathrm{neu}} + s_2 C_{\mathrm{sel}_2,\mathrm{neu}}. \tag{3b}$$

The approximate Equation 3a shows that when selection is weak, the frequency of the beneficial allele at a selected locus sel_i changes not only because of its own effect on fitness ($s_i p_{\mathrm{sel}_i} q_{\mathrm{sel}_i}$), as if it was alone, but also because of the hitchhiking with the other selected locus sel_j ($s_j C_{\mathrm{sel}_i,\mathrm{sel}_j}$), which depends on the other selection coefficient and the linkage disequilibrium, as if sel_i was neutral. This latter term captures the interference between the selected loci. At the neutral locus, when selection is weak the change in frequency is simply the sum of the effects of hitchhiking with both selected loci, each of which depends on the respective selection coefficient and on the linkage disequilibrium between the neutral locus and the locus under selection (3b).

At this stage, we can get a first feeling of the dynamics of the neutral allele frequency based on Equation 3b. If $C_{\mathrm{sel}_1,\mathrm{neu}}$ and $C_{\mathrm{sel}_2,\mathrm{neu}}$ have the same sign, then the two selective sweeps have cumulative effects on the change in frequency at the neutral locus: the hitchhiking effects are synergistic, and the neutral allele frequency will evolve faster than if it was hitchhiking with only one selected allele ("single selective sweep"). In contrast, if $C_{\mathrm{sel}_1,\mathrm{neu}}$ and $C_{\mathrm{sel}_2,\mathrm{neu}}$ have opposite signs, the hitchhiking effects are antagonistic and will tend to compensate each other. In such cases, the frequency of a neutral allele exposed to a "double" selective sweep may evolve *slower* than that under a "single" selective sweep.

The intensity of the interaction (synergy or antagonism) of hitchhiking effects depends on the selection coefficients as well as on the magnitudes of the linkage disequilibria $C_{\mathrm{sel}_1,\mathrm{neu}}$ and $C_{\mathrm{sel}_2,\mathrm{neu}}$, so in the long run, the change in frequency at the neutral locus also depends on the dynamics of $C_{\mathrm{sel}_1,\mathrm{neu}}$ and $C_{\mathrm{sel}_2,\mathrm{neu}}$. The initial values of those linkage disequilibria vary with the starting conditions. Their exact dynamics during the sweeps are too complicated to give here (see APPENDIX A). However, to a first-order approximation with small s_1 and s_2, the recursions for two- and three-locus linkage disequilibria read

$$C'_{\mathrm{sel}_1,\mathrm{sel}_2} \simeq (1 - r_{\mathrm{sel}_1,\mathrm{sel}_2}) C_{\mathrm{sel}_1,\mathrm{sel}_2} \big(1 + (1 - 2p_{\mathrm{sel}_1})s_1 + (1 - 2p_{\mathrm{sel}_2})s_2\big) \tag{4a}$$

$$C'_{\mathrm{sel}_i,\mathrm{neu}} \simeq (1 - r_{\mathrm{sel}_i,\mathrm{neu}}) \big(C_{\mathrm{sel}_i,\mathrm{neu}}(1 + (1 - 2p_{\mathrm{sel}_i})s_i) + s_j C_{\mathrm{sel}_1,\mathrm{sel}_2,\mathrm{neu}}\big) \quad (\{i,j\} = \{1,2\} \text{ or } \{2,1\}) \tag{4b}$$

$$C'_{\mathrm{sel}_1,\mathrm{sel}_2,\mathrm{neu}} \simeq (1 - \gamma)(C_{\mathrm{sel}_1,\mathrm{sel}_2,\mathrm{neu}}\big(1 + (1 - 2p_{\mathrm{sel}_1})s_1 + (1 - 2p_{\mathrm{sel}_2})s_2\big) - 2C_{\mathrm{sel}_1,\mathrm{sel}_2}(s_1 C_{\mathrm{sel}_1,\mathrm{neu}} + s_2 C_{\mathrm{sel}_2,\mathrm{neu}})) \tag{4c}$$

with $(1 - \gamma) = (1 - r_{A,B})(1 - r_{B,C})$, where loci A, B, and C represent the previously defined loci (neu, sel_1, sel_2) ordered along the chromosome (see APPENDIX A). In other words, $(1 - \gamma)$ is the probability that there is no recombination between the "first" and the "second" locus, and no recombination between the second and the "third" locus, where first, second, and third refer to the position of the loci on the chromosome, regardless of their status (neutral or selected). The linkage disequilibrium between the selected loci, $C_{\mathrm{sel}_1,\mathrm{sel}_2}$, either increases or decreases as a result of selection, depending on the sign of $(1 - 2p_{\mathrm{sel}_1})s_1 + (1 - 2p_{\mathrm{sel}_2})s_2$, and systematically decreases (in absolute value) with increasing recombination. Our main focus in our examination of the hitchhiking effect is the linkage disequilibria between the neutral locus and each of the selected loci, $C_{\mathrm{sel}_1,\mathrm{neu}}$ and $C_{\mathrm{sel}_2,\mathrm{neu}}$, as shown in Equation 3b. For each selected locus sel_i, $C_{\mathrm{sel}_i,\mathrm{neu}}$ is modified because of selection at sel_i and also because of selection at the other selected locus sel_j, by the means of the three-way linkage disequilibrium $C_{\mathrm{sel}_1,\mathrm{sel}_2,\mathrm{neu}}$ (Equation 4b).

Hence, the selective interference between sel_1 and sel_2 affects the variation of the neutral allele frequency through the higher-order interaction, *i.e.*, the three-locus linkage disequilibrium (Equation 4b). This makes sense and gives an illustration of the often difficult to understand meaning of three-locus linkage disequilibrium. When there is selective interference between sel_1 and sel_2, the linkage disequilibrium between these two selected loci does not directly affect the neutral allele frequency. However, regardless of the association between sel_1 and sel_2, a nonequilibrium repartition of neutral alleles among two-locus haplotypes at sel_1–sel_2 does affect the dynamics of linkage disequilibria $C_{\mathrm{sel}_1,\mathrm{neu}}$ and $C_{\mathrm{sel}_2,\mathrm{neu}}$ (Equation 4b), which in turn influences the hitchhiking effects; three-locus linkage disequilibrium is indeed a measure of this repartition. Note that the influence of selective interference through the three-locus linkage disequilibrium is directly apparent from the change in frequency at the neutral locus (Equation 2b) but should have an effect only when selection is strong. Again, recombination always decreases (in absolute value) the association between neu and sel_i; therefore the location of neu is critical to the outcome of the double selective sweeps. If neu is inside the interval delimited by sel_1 and sel_2, the difference between $r_{\mathrm{sel}_i,\mathrm{neu}}$ and $r_{\mathrm{sel}_j,\mathrm{neu}}$ is always smaller than if neu were outside of the interval. The difference between the rates of variation of $C_{\mathrm{sel}_1,\mathrm{neu}}$ and $C_{\mathrm{sel}_2,\mathrm{neu}}$ is then larger when the neutral locus is outside the interval. Moreover,

for a given distance between sel_1 and sel_2, $(1-\gamma)$ is always smaller when the neutral locus is outside of the interval than when it is inside. Hence, when neu is inside the interval, the interplay of hitchhikings from both selected loci (whether synergy or antagonism, depending on the starting conditions) is more likely to be sustained. In contrast, if neu is outside this interval, in the long run the hitchhiking effect by the closest selected locus dominates, and there is less opportunity for either antagonism or synergy of hitchhikings effects.

For the rest of this article, we define neu_m as the neutral locus located exactly in the middle of the interval delimited by sel_1 and sel_2 and neu_i as the neutral locus located at the same distance from sel_i as neu_m, but on the other side of sel_i, outside of the interval. The recombination rates with the loci under selection are thus, for neu_m, $r_{sel_1,neu_m} = r_{sel_2,neu_m} = r$ (by definition), and for neu_i ($i = 1$ or 2), $r_{sel_i,neu_i} = r$ (by definition) and $r_{sel_j,neu_i} = 3r(1-r)^2 + r^3 \simeq 3r + o(r)$. Hence, for neu_m, the linkage disequilibrium with the two loci under selection is similarly affected by recombination, whereas, for neu_1 and neu_2, the linkage disequilibrium with the farthest selected locus is three times more affected by recombination as the one with the closest selected locus. As a consequence, under similar strength of selection at both loci, we expect the neutral polymorphism at neu_m to reflect the interplay of selection at the two loci, whereas the polymorphism at neu_1 and neu_2 will carry mainly the signature of selection at the closest selected locus.

Apart from the dynamics of the selected loci and the changes in linkage disequilibria, the type of interaction between the selective sweeps (antagonism or synergy) strongly depends on the initial conditions. In the simplest case where both beneficial mutations enter the population at the same generation, they are most likely in negative linkage disequilibrium, and the probability that they are associated with different alleles at a neutral locus equals the heterozygosity at that generation (*i.e.*, the probability of drawing two different alleles at a locus). In a given fragment of sequence, there are several polymorphic sites, for which the frequency distribution of mutant alleles is well known (Ewens 2004). Thus, during one occurrence of a double selective sweep, several initial conditions regarding the initial linkage disequilibria $C_{sel_i,neu}$ are encountered among the various polymorphic sites. As a consequence, the double selective sweep affects not only the global neutral diversity (as measured by the heterozygosity or nucleotide diversity), but also the repartition of this diversity among sites (as measured by the frequency spectrum of mutations). Moreover, the initial frequencies of the neutral mutations dramatically influence their evolutionary dynamics and their final frequency at the end of the sweep. The frequency spectrum is then a key indicator here, and we wish to describe its evolution under the influence of two interfering selective sweeps, by following small sequence fragments located in the vicinity of the selected loci. Moreover, we wish to know if the general processes that we described using the deterministic model still matter when starting from a realistic initial distribution of neutral allelic frequencies.

Since the changes in frequencies at many tightly linked polymorphic sites are not analytically tractable, we used Monte Carlo simulations to address this question. This also allowed us to take into account the stochasticity inherent to every actual population, which may have important consequences in several aspects of the process. For instance, in a finite population, the two selected mutations need to end up on the same haplotype (by recombination) to both get fixed (Hill and Robertson 1966). This represents a qualitative shift that cannot appear in an analytical treatment and yet strongly influences the outcome of the selective sweeps. Moreover, forward Monte Carlo simulations allowed us to stochastically introduce new polymorphic sites through mutation during the selective sweep [infinite-site model (Ewens 2004)]. Finally, by simulating the sampling of gametes by the experimenter, we could include the sampling variance in our analysis. In the following, we present the approach and results of our forward simulations of interfering selective sweeps.

METHODS

Forward simulations: We used forward individual-based Monte Carlo simulations to investigate the effects of selection at two closely linked loci on neighbor neutral polymorphism. We simulated polymorphism at several sequence fragments along a chromosome region encompassing two sites under positive selection. Each fragment evolved under the infinite-site model of mutation. Recombination was allowed within and between fragments. The actual number of sites in each fragment was not explicitly defined; instead, a continuous model was used, in which the mutation parameter $\theta = 4N_e\mu$ and the recombination parameter $\rho = 4N_e r$ (where N_e is the effective population size) were defined at the level of the entire fragment.

At the beginning of each simulation, the initial conditions were settled for each fragment by generating the whole population by coalescence using the program "ms" (Hudson 2002). This provided realistic initial conditions regarding the distribution of polymorphism in each fragment, without having to simulate the complex genealogical relationships between the fragments, since we did not wish to measure the linkage disequilibrium between fragments. Although coalescence theory is generally used for samples that are small relative to the population size, Wakeley and Takahashi (2003) showed that when the sample size equals the effective population size, the error induced by using the coalescent is minute. Indeed, by neglecting multiple

coalescence events, the standard coalescent expectation underestimates by ~12% the expected number of alleles present in a single copy in the entire population, all other frequency classes remaining essentially unchanged. We checked that this approximation did not affect our results by artificially increasing the proportion of singletons by 12% in the initial population generated with ms and then running the forward simulations. The outcome was equivalent to that without increasing the number of singletons, thus validating the accuracy of our method (results not shown). We used $\theta = 5$ and $\rho = 10$ as parameters, such that the ratio ρ/θ was similar to that documented for Drosophila (Kliman *et al.* 2000; Przeworski *et al.* 2001).

We considered two selected loci: sel_1 and sel_2. The selective phase was simulated forward in time. It started with the introduction of the beneficial allele at sel_1 and ended when the beneficial alleles at sel_1 and sel_2 were both fixed. If any of the beneficial alleles was lost before fixation, the run was discarded and a new simulation was started again with the same initial conditions. For each locus under selection, the haplotype carrying the beneficial allele was introduced in five copies. This reduced computing time by lowering the risk that a beneficial allele was lost by drift in the early generations. This procedure is justified since our observations are conditioned on the final fixation of both mutations. Indeed, according to Barton (1998), conditional on its final fixation, a beneficial mutation rises quickly in frequency in early generations, and thus there is negligible opportunity for mutation or recombination to occur on the haplotype that carries it. In practice, for each selected locus, a single haplotype from ms was copied five times and the beneficial mutation was placed on it. Hence this approach was meant to model the rapid increase in frequency of the beneficial mutation in the early generations (conditional on fixation). It should not be confused with a selective sweep from the standing variation, where a mutation first drifts neutrally for several generations and then becomes selected when it is at a frequency $>1/(2N)$. In such a "soft" selective sweep, the beneficial mutation may initially be present on several distinct haplotypes. The neutral signature of such a soft sweep may be very different from that of a hard selective sweep, as Przeworski *et al.* (2005) showed, but this is not the topic of this article.

During the selective phase, mutation and recombination rates were defined at the individual rather than the population level, using the same μ and r as in the neutral phase. For each fragment, when mutation occurred in a gamete, it was simulated by randomly drawing a position inside the fragment out of a continuous uniform distribution and introducing a derived allele at this position. Recombination was simulated in the same manner between the neutral fragments and the sites under selection, as well as inside the fragments, using Haldane's mapping function assuming no interference.

Signatures of selection: At the end of the simulation, several measures were made. First, we computed the reduction of heterozygosity in the entire population. This was expressed as the ratio π/π_o of the observed nucleotide diversity per fragment over its value at the beginning of the selective phase. We also simulated the sampling of a small number of individuals ($2n = 50$ gametes) to assess the properties of the frequency spectrum and to perform some tests of selection. The samples were drawn conditionally on the presence of polymorphism in at least one fragment. The frequency spectrum was calculated as in Kim (2006). We computed the proportions of sites belonging to each frequency class (*i.e.*, from 1 to $2n - 1$) for each repeat and then averaged these proportions over all the repeats. For each simulation run, we also calculated several summary statistics for the frequency spectrum of mutations. The first one, Tajima's D (Tajima 1989), is the normalized difference between Watterson's (1975) estimator of θ based on the number of polymorphic sites and Tajima's (1983) estimator π based on the heterozygosity of sites. A negative value denotes an excess of low-frequency variants, indicative of positive selection or of population expansion, whereas a positive value denotes an excess of intermediate frequencies as can be caused by balancing selection. When necessary, we also calculated the H statistics defined by Fay and Wu (2000), in the standardized version proposed by Zeng *et al.* (2006). A negative value is indicative of an excess of very-high-frequency variants, which is a signature of positive selection. Finally, we used Zeng *et al.*'s (2006) E statistics, which contrast the abundance of low- *vs.* high-frequency variants. We report the mean values of these statistics over 500–1000 runs of simulations. We also assessed their respective powers to reject neutrality. This was calculated as the proportion of simulations for which the value of the statistics led to rejecting neutrality at the 5% significance level. Significance was assessed by running 10,000 coalescence simulations with ms (Hudson 2002) with the same sample size and the same number of polymorphic sites as the mean of selection simulations. We used the subset of ms samples in which at least one polymorphic site was present, as in Przeworski (2002). All statistics were used in one-sided tests, including Tajima's D, again as in Przeworski (2002). This means that we considered that positive and negative values of Tajima's D bear distinct information and lead to different evolutionary interpretations. As such they can be treated as different statistical tests instead of just being pooled as a global rejection of neutrality.

RESULTS

The symmetric case: We first consider the case in which mutations at sel_1 and sel_2 appear simultaneously in the population and have the same selection coefficients ($s_1 = s_2 = s = 0.1$). Though this is likely not the

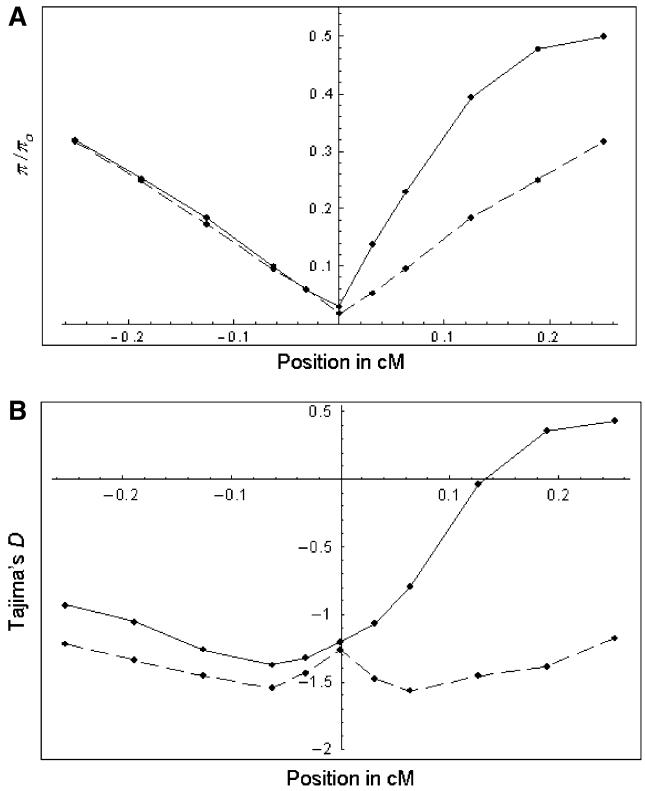

FIGURE 1.—Polymorphism patterns along the chromosome. (A) Reduction of heterozygosity (π/π_o) in the whole population; (B) Tajima's D in a sample of size $2n = 50$ chromosomes, as a function of the distance to sel_1. Solid line: selective sweeps at two close loci, with $s_1 = s_2 = s = 0.1$ and $r_{sel_1,sel_2}/s = 0.05$. Dashed line: single selective sweep with $s = 0.1$ and $r/s = 0.05$. Results are averaged over 500 simulations with population size $2N = 20{,}000$. sel_1 is at position 0 in both cases. In the case of two selective sweeps, sel_2 is located at 0.50 cM, so the rest of the pattern (not shown) would be symmetrical over $x = 0.25$.

most realistic situation, we use it as a case study to better understand and illustrate the various forces in action. As this situation is completely symmetric, we can consider only one-half of the chromosome segment, namely the sel_1 side from neu_1 to neu_m. In the following, neu_1 and neu_m do not refer to specific polymorphic sites, but rather to small chromosomal regions centered on the position described earlier for these loci.

Reduction of heterozygosity: Figure 1A shows the reduction of heterozygosity (quantified by π/π_o) at neutral markers along the chromosome. One selected locus (sel_1) is at position 0, and the other selected locus (sel_2) is on the right-hand side at +0.5 cM from sel_1 (not shown). The neutral locus at the right end of the graph is neu_m, located in the very middle of the interval [sel_1–sel_2]. Hence, the right side of the graph (positive abscissa values) gives the pattern for "inside" neutral markers located in the interval [sel_1–neu_m], and the left side of the graph (negative abscissa values) gives the pattern for "outside" neutral markers located in the interval [telomere–sel_1]. In the symmetrical case, the pattern on the right of neu_m is symmetrical (not shown). The solid line in Figure 1A gives the pattern of heterozygosity when selection acts at both loci sel_1 and sel_2 (hereafter we refer to this case as "double selective sweep"). As a comparison, the dashed line gives the pattern when there is selection at only one locus, here at sel_1 ("single selective sweep"). On the graph, the neutral loci on the left side (neu_1) and the right side (neu_m) are at the same distance from sel_1. This is used to compare inside and outside loci for the same recombination rate with the selected locus and also to compare single and double selective sweeps (see below).

The pattern observed in Figure 1A for a single sweep (dashed line) is the well-known classical picture (MAYNARD SMITH and HAIGH 1974; STEPHAN *et al.* 1992; KIM and STEPHAN 2002) except that here the diversity is not zero for a marker located at position zero (*i.e.*, on the selected locus sel_1), because mutation takes place during the course of the selective sweep in our forward simulation model. For a double selective sweep (solid line), the pattern on the left of sel_1, *i.e.*, outside the selected bracket, is very similar to that obtained for a single selective sweep at sel_1 of the same intensity. In contrast, between sel_1 and sel_2, *i.e.*, inside the selected bracket, the neutral polymorphism is substantially higher than outside the bracket or than the case of a single selective sweep.

This is the first main result of the simulations, consistent with the deterministic explanation above. In the case where there is interference between selective sweeps of similar intensities at two linked mutations, the pattern of polymorphism outside the selected bracket resembles that of a single selective sweep; *i.e.*, even when there is selection at both sel_1 and sel_2, neutral loci on the left of sel_1 are mostly affected by selection at sel_1, not at sel_2. In contrast, for neutral markers lying between the selected loci, the diversity at the end of the selective sweeps is the result of the combined effects of both hitchhikings. In particular, in the case of antagonistic selective sweeps that start at the same generation at sel_1 and sel_2, *more* polymorphism is maintained than in the case of a single selective sweep of the same intensity.

Frequency spectrum and Tajima's D: The frequency spectra at neu_1 and neu_m in a sample of size $2n = 50$ are shown in Figure 2. At neu_1, the spectrum is characterized by an excess of high-frequency derived variants, a lack of intermediate-frequency variants, and an excess of low-frequency variants relative to the neutral expectation. Taken together, these features are typical of a neutral locus partially linked to a locus under positive selection (TAJIMA 1989; FAY and WU 2000; PRZEWORSKI 2002). At neu_m, there is an excess of high-frequency variants, consistent with positive selection, but a lack of low-frequency variants. More importantly, there is an excess of variants at intermediate frequencies (from 15 to 35) relative to the standard neutral case. Taken alone, this latter feature is commonly interpreted as the outcome of selective forces maintaining diversity, *i.e.*,

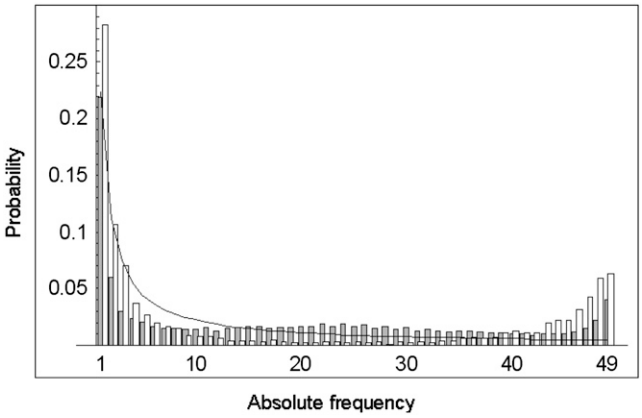

FIGURE 2.—Frequency spectrum of mutations in a sample of size $n = 25$ diploid individuals ($2n = 50$ chromosomes). Solid line, expectation under standard neutrality; open bars, neu_1 (outer locus); shaded bars, neu_m (inner locus). Results are averaged over 500 simulation runs. Parameters are as in Figure 1.

TABLE 1

Variation pattern of Tajima's *D*

	Standard deviation	$P(D < 0^a)$	$P(D > 0^a)$
Neutral equilibrium.	0.720	0.05	0.05
Single sweep	0.882	0.565	0.020
neu_1	1.079	0.45	0.044
neu_m	1.311	0.106	0.274

Standard deviation and proportion of significantly positive and negative values of Tajima's D in a sample of size $2n = 50$ chromosomes are shown. The neutral equilibrium values are from simulations of the standard coalescent. For the single selective sweep, the values are from forward simulations of a neutral locus located at a recombination distance r from a locus under positive selection such that $r/s = 0.05$, $s = 0.1$. For the case of two selective sweeps, neu_1 and neu_m are such that $r_{neu_1,sel_1}/s_1 = r_{sel_1,neu_2}/s_1 = 0.05$ and $s_1 = s_2 = 0.1$.

[a] Significance at the 5% level was assessed using standard coalescent simulations (10,000 runs).

balancing selection (CHARLESWORTH 2006). However, taken together the features of the frequency spectrum at neu_m—excess of high-frequency derived variants, excess of intermediate-frequency variants, and lack of low-frequency variants relative to the neutral expectation—appear as a distinctive pattern of a double selective sweep.

To understand how the neutral-frequency spectrum changes along the chromosome in the case of a double selective sweep, we used Tajima's D as a summary statistic because it is the most sensitive to perturbations of intermediate-frequency classes (TAJIMA 1989; PRZEWORSKI 2002). Figure 1B shows the pattern of Tajima's D along the chromosome, with the same formalism as in Figure 1A. For the single selective sweep (dashed line), Tajima's D is negative in the entire region considered, as a consequence of the lack of intermediate-frequency variants generated by the strong positive selection. Note that the values are higher for the marker that includes the target of selection than for markers in the close flanking regions, as a consequence of the smaller number of polymorphic sites in the former than in the latter. In the case of a double sweep (solid line), Tajima's D is globally higher than for a single selective sweep, even though the selection coefficient at sel_1 is the same. Moreover, the pattern is strongly asymmetric: Tajima's D inside the selected interval is higher than outside the interval. In the case of Figure 1B, Tajima's D is even *positive* in the middle of the selected interval. This is the second important result of this article, which was not directly predictable from the deterministic model above: the interference of two selective sweeps has more impact on the frequency spectrum than on the reduction of heterozygosity. In our illustrative example, neu_m exhibits a reduced heterozygosity, which is consistent with positive selection in the region (Figure 1A, right edge of the solid line), whereas Tajima's D, which summarizes the frequency spectrum, does not carry any signature of positive selection and is even positive (Figure 1B, right edge of the solid line).

The frequency spectrum is obviously affected by the stochasticity inherent to the finite population size and to the sampling, so there is variation in Tajima's D between repeats, which we report in Table 1. For the sake of simplicity we show only results for neu_m and neu_1, as well as for a neutral locus equivalent to neu_1 in the case of a single selective sweep. The standard deviation of Tajima's D in case of a double selective sweep is larger than that for a single sweep of the same intensity. Moreover, the standard deviation at neu_m is much larger than that at neu_1. Indeed, neu_m is more directly under the combined effects of two hitchhikings than neu_1, so slight changes in the starting conditions can lead to more variation in the final state at neu_m than at neu_1.

We also report in Table 1 (columns 3 and 4) the power to reject neutrality in our simulations, using Tajima's D in one-sided tests (see METHODS). Interestingly, at neu_m it is substantially more probable to reject neutrality through a significantly positive value (>27% of the simulations) than through a significantly negative value. This is not true for neu_1, for which the powers of the tests on the right side and on the left side are comparable to those for a single selective sweep of the same intensity.

Area of influence: We explored the range of recombination values between the selected loci for which the selective sweep at sel_2 had an influence on the polymorphism pattern generated by the selective sweep at sel_1. This was done by increasing the distance between sel_1 and sel_2, while keeping the selection coefficients constant, and relocating the neutral markers such that neu_m remained in the middle of the interval and neu_1 lay outside the interval at the same distance from sel_1. The results are shown in Figure 3, where the reduction in heterozygosity (π/π_o) at neu_m and neu_1 is plotted as a

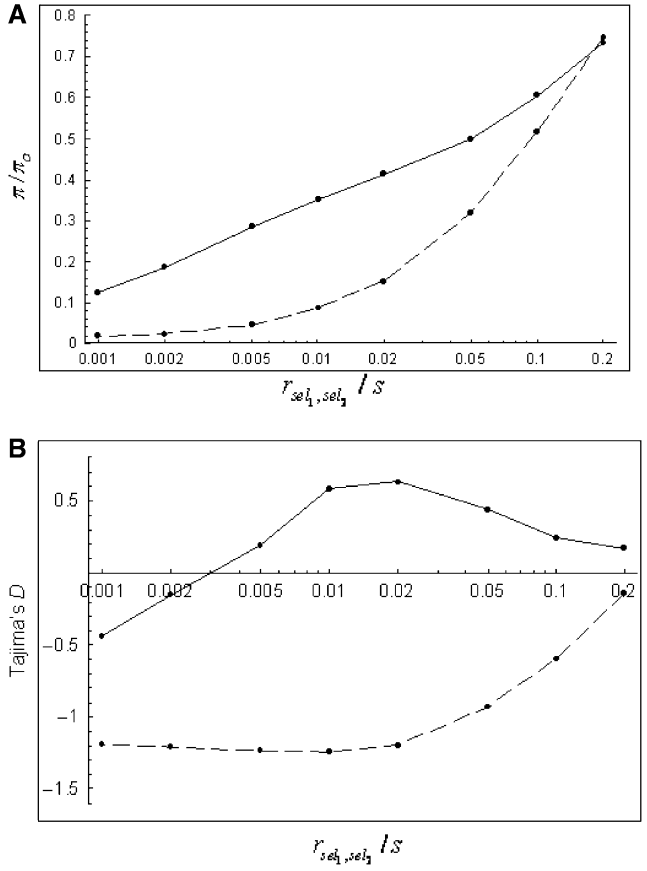

FIGURE 3.—Influence of the distance between selected loci. (A) Reduction of heterozygosity in the whole population; (B) Tajima's D at neu_m (solid line) or neu_1 (dashed line), in a sample of size $2n = 50$, plotted against the distance between the selected loci expressed as a ratio $r_{\text{sel}_1,\text{sel}_2}/s$ between recombination and the selection coefficient. Mean results are over 1000 repeats, with all the parameters as in Figure 1 except for $r_{\text{sel}_1,\text{sel}_1}$.

function of the ratio $r_{\text{sel}_1,\text{sel}_2}/s$ between recombination and the selection coefficient. Interestingly, neu_m remains more polymorphic than neu_1 for values of $r_{\text{sel}_1,\text{sel}_2}/s$ spanning more than two orders of magnitude (Figure 3A). This range corresponds to that usually documented for the region of influence of a selective sweep (FAY and WU 2005). Again, the frequency spectrum is more sensitive than the polymorphism level to the interference between selective sweeps (Figure 3B). For instance, at $r_{\text{sel}_1,\text{sel}_2}/s = 0.2$, neu_m and neu_1 have similar reduction in heterozygosity but different Tajima's D values, with a slightly positive value at neu_m. The strongest asymmetry in Tajima's D between neu_1 and neu_m occurs at $r_{\text{sel}_1,\text{sel}_2}/s = 0.02$ (*i.e.*, when the recombination rate between the neutral locus and the locus under selection is such that $r/s = 0.01$). At this point, the difference between Tajima's D for neu_m and neu_1 is 1.83. Note that as $r_{\text{sel}_1,\text{sel}_2}/s$ increases, Tajima's D at the middle of the interval has reduced positive values, but those extend over a larger chromosomal region.

Duration of the signature: The footprints left by selective sweeps on neutral variation are obviously transient, since mutation and drift eventually restore the heterozygosity and the frequency spectrum to their neutral equilibria. The duration of such a signature is a key issue in the detection of selection in natural populations and has been recently a subject of much interest (PRZEWORSKI 2002, 2003; JENSEN *et al.* 2005). In our case, we wished to know how the particular pattern of polymorphism induced by a double selective sweep evolved after the end of the sweeps. Figure 4 shows the power to reject neutrality after the fixation of both beneficial mutations, using three summary statistics for the frequency spectrum: Tajima's D, Fay and Wu's H, and Zeng's E (see METHODS). After the end of the selective sweeps, the power of Fay and Wu's H decreases much faster than that of Tajima's D (Figure 4, A and B), as was already discussed in PRZEWORSKI (2002). For all three summary statistics, the power to reject neutrality outside the sel_1–sel_2 interval remains higher than inside the interval, over the entire period considered. ZENG *et al.* (2006) emphasized that the power of their new statistic E increased *after* the end of the selective sweep (or bottleneck), coinciding with the decrease of H, so that these two statistics had somehow antagonistic behaviors because they both depended on high-frequency variants. Our simulation results show that after two simultaneous selective sweeps, the power of E increases more slowly for neu_m than for neu_1. Thus E relays the information contained in H, including the differences between the powers to reject neutrality at loci located inside or outside the selected interval. Altogether, our results indicate that the signature left by antagonistic hitchhiking effects may persist for a long time after fixation of the beneficial mutations.

Relaxing the symmetry: Until now, we have focused on an illustrative, completely symmetric case, in which selective sweeps at sel_1 and sel_2 occurred simultaneously and where both mutations had the same selection coefficient. In practice, it is unlikely that two beneficial mutations arise at the same generation at two closely linked loci. Also, selection coefficients may vary importantly between beneficial mutations. The interaction between selective sweeps is expected to depend on the synchronicity of the beneficial mutation events at sel_1 and sel_2, as well as on their relative selection coefficients, which determine how long in time the sweeps will overlap. To assess the influence of these parameters on the final pattern of polymorphism, we ran simulations where the beneficial allele at sel_1 appeared first and then was allowed to reach a threshold frequency p_t before the beneficial allele at sel_2 was introduced. The threshold frequency p_t was transformed into a scaled time τ to account for the fact that the trajectory of a beneficial allele is not linear in time (see APPENDIX B). The selection coefficient s_1 was kept constant, while s_2 was varied such that $s_2/s_1 = \frac{1}{2}$, 1, or 2.

Figure 5 shows the resulting Tajima's D at neu_m, neu_1, and neu_2. At outside loci, neu_1 and neu_2 (Figure 5, A and

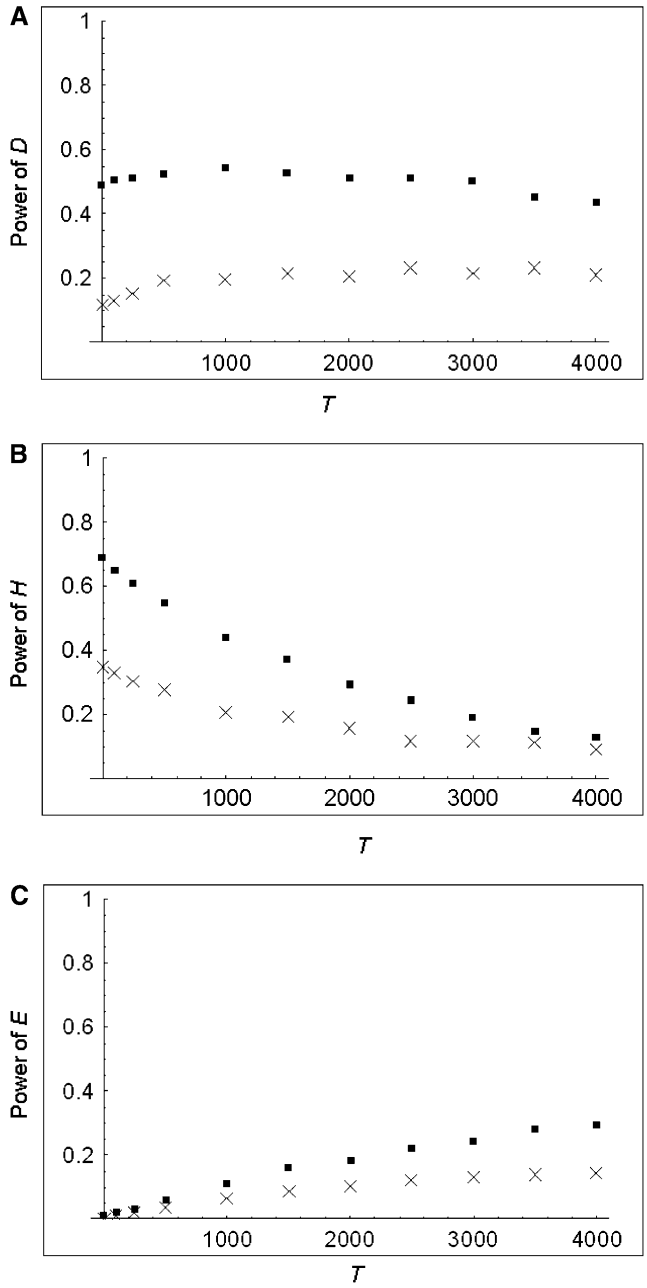

FIGURE 4.—Power to reject neutrality after the sweeps: proportion of simulations (of 500–1000 repeats) that reject neutrality at the 5%-significance level using (A) D, (B) H, or (C) E as summary statistics, plotted against the time T (in generations) since fixation of the last beneficial mutation. Significance was assessed with standard coalescent simulations (10,000 runs), and all statistics were used in a one-sided test for negative values. Cross, neu_m; box, neu_1.

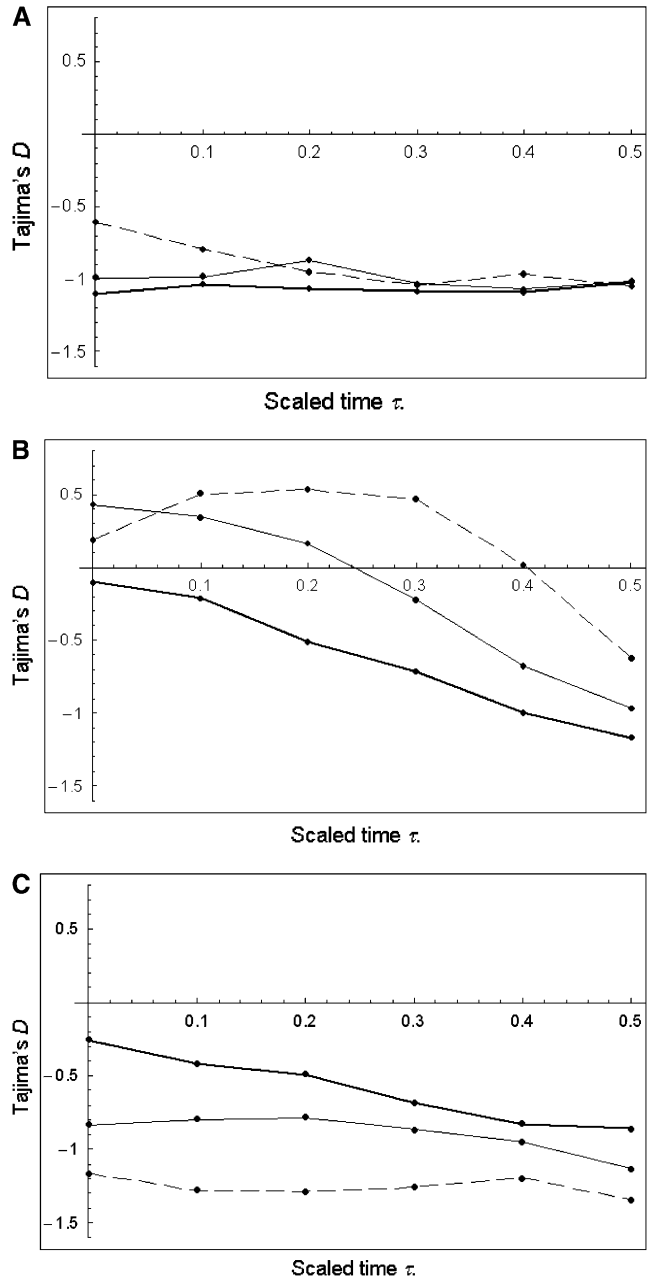

FIGURE 5.—Delayed selective sweeps with varying selection coefficients: final Tajima's D at (A) neu_1, (B) neu_m, and (C) neu_2 as a function of the scaled delay τ between the introduction of the beneficial mutations at sel_1 and at sel_2 (see APPENDIX B). Thin line, $s_2 = s_1$; thick line, $s_2 = s_1/2$; dashed line, $s_2 = 2s_1$. All other parameters are as in Figure 1.

C), the delay between selective sweeps has little influence on Tajima's D at fixation. At neu_1, the impact of the hitchhiking by sel_2 was substantial only when $s_2 > s_1$ and τ was close to 0, *i.e.*, when the selective sweeps had little delay (Figure 5A, dashed line). At neu_2 (Figure 5C), the final Tajima's D obviously depends on s_2 as it is the selection coefficient of the closest selected locus, but the influence of selection at sel_1 is weak except when $s_2 < s_1$. In contrast, at neu_m (Figure 5B), the frequency spectrum is under strong influence of both hitchhiking effects, and the outcome is highly dependent on the timing of the sweeps and on the ratio of the selection coefficients. When $s_2 \leq s_1$ (Figure 5B, thin and thick lines), Tajima's D is maximal at $\tau = 0$, *i.e.*, when the sweeps tend to be simultaneous, and decreases rapidly with increasing τ. When $s_2 > s_1$, there is a nonzero value

of τ that maximizes Tajima's D (Figure 5B, dashed line). This is because the dynamics of the selected allele frequencies are different from each other; hence a delay can enhance the antagonism of hitchhiking effects by allowing the beneficial alleles at sel_1 and sel_2 (i) to enter synchronously in the critical phase of a hitchhiking effect, in which the dynamics of the beneficial allele is quasi-deterministic, while this allele remains at a low frequency (Barton 1998), and (ii) to reach nonnegligible frequencies at similar times, so that several recombination events can produce haplotypes hosting the two favorable alleles in coupling. Note that for this to happen, the weaker mutation must start increasing in frequency earlier, so it has more chances to escape loss by drift due to the interference with the stronger mutation (Barton 1995). Therefore, this scenario is also the most likely to be encountered in real data exhibiting fixation at both selected loci. As expected, in all three situations there is a value of τ for which Tajima's D becomes lower at neu_m than at outer loci, indicating that hitchhiking effects in the middle of the interval switch from antagonism to synergy. This occurs all the earlier (in time) when the second selective sweep has a lower selection coefficient, because selective interference between the beneficial mutations is then reduced.

Note that in nonsymmetric cases, we always kept neu_m at the middle of the selected interval, although antagonism between hitchhiking effects is not necessarily maximal at this point in the case of selective sweeps of different intensities. Hence, our simulation results at neu_m are conservative regarding the antagonism of hitchhiking effects.

DISCUSSION AND CONCLUSIONS

Two selective sweeps of comparable intensities that arise close to each other tend to interfere in their effects on neutral variation. They can maintain globally higher levels of polymorphism than a single selective sweep of the same intensity, because selective interferences can slow down the dynamics of each mutation, allowing for more recombination (Kim and Stephan 2003). Here, we further show that, regardless of the trajectories of beneficial mutations in time, two interfering selective sweeps can compete in their hitchhiking effects simply by dragging along different neutral alleles. The resulting polymorphism pattern (as quantified by the nucleotide diversity π) is highly asymmetric, the region between the selected loci being the most subject to the interplay of the two hitchhiking effects. More importantly, we show that the interference of selective sweeps can distort the frequency spectrum in the direction of an excess of intermediate frequencies at neutral sites between the two selected loci, which is often interpreted as a signature of balancing selection (Charlesworth 2006).

Here, we conditioned the simulations on the fixation of both beneficial mutations. The actual fixation probability of beneficial mutations cannot be calculated directly from our simulations, since these mutations were introduced in several copies to decrease simulation time. The decrease in the probability of fixation as a consequence of selective interference was studied in detail in Barton (1995) and can be substantial. The conclusions of this study are thus more accurately applicable to cases where selection coefficients are large and of the same order of magnitude, for which the probability of joint fixation of both mutations is not negligible. Note that generally theoretical studies of selective sweeps based on coalescent simulations assume fixation of the beneficial mutation. Most of these studies also rely on the assumption that the product Ns is of the order of 500–1000 while the population size is very large (of order 10^6), such that the selection coefficient and the fixation probability are of order 10^{-3} (see, *e.g.*, Fay and Wu 2000, 2005; Przeworski 2002). Also, interference of selective sweeps could well occur between beneficial mutations already present in the population and initially neutral, for instance, following a rapid environmental shift, which greatly decreases the risk of stochastic loss (Hermisson and Pennings 2005). Such selective sweeps from the standing variation are expected to leave a footprint different from that of a hard sweep (Innan and Kim 2004; Przeworski *et al.* 2005). Yet, since most neutral mutations are expected to be in low frequency in a natural population (Ewens 2004), it is quite possible that very few copies (if not a single one) of the beneficial mutation actually sweep through the population, hence turning the soft sweep into a quasi-hard sweep.

It may be argued that asymmetry in the polymorphism pattern may well arise by chance in a single selective sweep. Indeed, Figure 1 shows the mean of several simulations corresponding to an expected pattern, while obviously there is variation between repeats. Hence, some single-sweep simulations could exhibit a pattern similar to the one expected under interfering selective sweeps. Nevertheless, in the context of a candidate region where selection is searched for, the current practice is to use several markers distributed throughout the region. As the number of markers increases, it is less and less likely that an asymmetric pattern will be observed by chance for all the markers. For instance, in Figure 1, there are five markers on each side of sel_1. Using $|\log(\pi(\text{left side})/\pi(\text{right side}))| > 0.5$ (where $|\ |$ denotes absolute value) as a criterion for asymmetry for each couple of markers equally distant from sel_1, the probability that asymmetry is in the same direction for all markers is 2.5 times higher under interfering selective sweeps than in the case of a single selective sweep. Palaisa *et al.* (2004) observed marked asymmetry at multiple markers in a genetic region. In such cases, a deterministic explanation might be involved rather than

just chance variation, and the occurrence of a second interfering selective sweep should be considered together with other possible causes of asymmetry, such as differences in recombination or mutation rates between both sides of the selective sweep.

Santiago and Caballero (2005) showed that in a highly subdivided population, a selective sweep can induce an increase of heterozygosity and an excess of intermediate-frequency variants in demes other than that where the beneficial mutation originated. This is because the selective sweep can force the introduction of neutral alleles that were previously absent or in negligible frequency in those demes because of low migration. Our study of interfering selective sweeps can be understood in the light of their results by using an analogy in which the alleles at one selected locus define demes for the other selected locus, and recombination is viewed as "migration" from one genetic background to the other. This analogy is the rationale for the so-called "structured coalescent" approach of selective sweeps (Kaplan *et al.* 1989). In our context, the case where both beneficial mutations are initially in strong negative linkage disequilibrium and carry different neutral alleles is similar to that in Santiago and Caballero (2005), where a selective sweep starting in one deme hitchhikes an allele absent in another deme. Indeed, the focal selective sweep introduces neutral polymorphism in the other selected background, thus reducing the effect of the other selective sweep, and reciprocally. This illustrative analogy is not a mere equivalence, though, since in the case of interfering selective sweeps, the sizes of the "demes" change with selection. There is also selective interference between the selected loci, which alters the process by slowing down the dynamics at each locus, so our results are not redundant with those of Santiago and Caballero (2005).

We focused on interference between sweeps at reasonably distant selected loci, which results in the asymmetric pattern described in Figure 1, when the initial linkage disequilibrium is negative. In contrast, in cases where the beneficial mutations are too closely linked to recombine in a reasonable time (for instance, when they are inside the same gene), and yet have similar enough selection coefficients to be maintained at high frequencies for a long time, positive selection can contribute to maintaining high levels of nucleotide diversity very close to the target of selection. This situation is what was termed trafficking by Kirby and Stephan (1996). At the most extreme, several beneficial mutations could arise at the same site and several copies of the same allele, identical in state but not by descent, could provoke interfering selective sweeps. This was studied as a particular case of "soft sweeps" by Pennings and Hermisson (2006), who focused on the signatures left by selection at a site where a beneficial mutation was introduced recurrently by mutation during the course of the sweep. We believe that this study could contribute to generalizing the somehow extreme (though very enlightening) cases of soft sweeps with recurrent mutation and of "trafficking" to arbitrarily distant interfering sweeps, including by attempting to assess the physical scale of the interaction between two selective sweeps (Figure 3). Though limited by the selective interference that decreases the fixation probabilities at each locus, sweep interference may be more likely to happen than soft sweeps with recurrent mutations or trafficking, because it involves larger chromosomal regions, which increases the probability of occurrence of two beneficial mutations.

Beneficial substitutions may not be evenly distributed over time, but rather concentrated in short time periods following environmental changes, when a previously well-adapted population needs to climb a new adaptive peak (as, for instance, in Orr 1998). If so, the simultaneous occurrence of several beneficial mutations may not be unlikely, and interference of selective sweeps may alter to some extent our ability to detect positive selection in genome scans, adding a new confounding factor to demography (Jensen *et al.* 2005) or variable genomic features (mutation, recombination). Perhaps more readily, the search for interfering selective sweeps could be helpful in specific studies focusing on smaller candidate regions, in which several putative targets of selection have already been identified. In such cases, the analysis of the polymorphism pattern could provide information not only about the presence of selection, but also about the synchronicity of selective sweeps or the origin (migration *vs.* mutation) of beneficial alleles This could yield valuable insights into the adaptive history of a species (Camus-Kulandaivelu *et al.* 2008).

We assumed here that selective sweeps had independent (multiplicative) effects on fitness. Epistasis between loci contributing to adaptive traits has already been shown to generate linkage disequilibrium between those loci (Caicedo *et al.* 2004). Epistasis between selected loci may also influence the neutral polymorphism pattern of interfering sweeps in a specific manner, so that it could be possible to identify selective interactions *a posteriori*. For instance, a recent article revealed a double selective sweep at two closely linked chromosomal regions involved in the sex-ratio distortion of *D. simulans* (Derome *et al.* 2008). Though both regions are compulsory for meiotic drive to occur in the lab (Montchamp-Moreau *et al.* 2006), the functional relationship between these two regions is still questioned in natural populations (C. Montchamp-Moreau, personal communication). It may be possible to use the polymorphism pattern in this region to try to elucidate how those loci interact in natural populations. More work is needed to investigate if there can actually be a molecular signature of the interaction between loci.

We thank Frantz Depaulis as well as two anonymous reviewers for helpful comments on an earlier version of this manuscript. This work was supported by a Bourse de Docteur Ingénieur grant from the

Centre National de la Recherche Scientifique (CNRS) to L.-M.C. F.H. and L.-M.C. are supported by grant ANR-06-BLAN-0128 from the CNRS.

LITERATURE CITED

Bamshad, M., and S. P. Wooding, 2003 Signatures of natural selection in the human genome. Nat. Rev. Genet. **4:** 99–111.

Barton, N. H., 1995 Linkage and the limits to natural selection. Genetics **140:** 821–841.

Barton, N. H., 1998 The effect of hitch-hiking on neutral genealogies. Genet. Res. Camb. **72:** 123–133.

Barton, N. H., and M. Turelli, 1991 Natural and sexual selection on many loci. Genetics **127:** 229–255.

Caicedo, A. L., J. R. Stinchcombe, K. M. Olsen, J. Schmitt and M. D. Purugganan, 2004 Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. Proc. Natl. Acad. Sci. USA **101:** 15670–15675.

Camus-Kulandaivelu, L., L.-M. Chevin, C. Tollon-Cordet, A. Charcosset, D. Manicacci *et al.*, 2008 Patterns of molecular evolution associated with two selective sweeps in the Tb1-Dwarf8 region in maize. Genetics (in press).

Charlesworth, D., 2006 Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. **2:** e64.

Derome, N., E. Baudry, D. Ogereau, M. Veuille and C. Montchamp-Moreau, 2008 Selective sweeps in a 2-locus model for sex-ratio meiotic drive in Drosophila simulans. Mol. Biol. Evol. **25:** 409–416.

Edelist, C., C. Lexer, C. Dillmann, D. Sicard and L. H. Rieseberg, 2006 Microsatellite signature of ecological selection for salt tolerance in a wild sunflower hybrid species, Helianthus paradoxus. Mol. Ecol. **15:** 4623–4634.

Ewens, W. J., 2004 *Mathematical Population Genetics—I. Theoretical Introduction.* Springer-Verlag, New York.

Fay, J. C., and C. I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Fay, J. C., and C. I. Wu, 2005 Detecting hitchhiking from patterns of DNA polymorphism, pp. 65–77 in *Selective Sweeps*, edited by D. Nurminsky. Landes Bioscience, Georgetown, TX.

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

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

Hermisson, J., and P. S. Pennings, 2005 Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics **169:** 2335–2352.

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

Hospital, F., C. Dillmann and A. E. Melchinger, 1996 A general algorithm to compute multilocus genotype frequencies under various mating systems. Comput. Appl. Biosci. **12:** 455–462.

Hudson, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18:** 337–338.

Innan, H., and Y. Kim, 2004 Pattern of polymorphism after strong artificial selection in a domestication event. Proc. Natl. Acad. Sci. USA **101:** 10667–10672.

Jensen, J. D., Y. Kim, V. B. DuMont, C. F. Aquadro and C. D. Bustamante, 2005 Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics **170:** 1401–1410.

Kaplan, N. L., R. Hudson and C. H. Langley, 1989 The "hitchiking effect" revisited. Genetics **123:** 887–899.

Kim, Y., 2006 Allele frequency distribution under recurrent selective sweeps. Genetics **172:** 1967–1978.

Kim, Y., and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics **155:** 1415–1427.

Kim, Y., and W. Stephan, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics **160:** 765–777.

Kim, Y., and W. Stephan, 2003 Selective sweeps in the presence of interference among partially linked loci. Genetics **163:** 389–398.

Kirby, D. A., and W. Stephan, 1996 Multi-locus selection and the structure of variation at the white gene of *Drosophila melanogaster.* Genetics **144:** 635–645.

Kirkpatrick, M., T. Johnson and N. Barton, 2002 General models of multilocus evolution. Genetics **161:** 1727–1750.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman *et al.*, 2000 The population genetics of the origin and divergence of the Drosophila simulans complex species. Genetics **156:** 1913–1931.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. Camb. **23:** 23–35.

Montchamp-Moreau, C., D. Ogereau, N. Chaminade, A. Colard and S. Aulard, 2006 Organization of the sex-ratio meiotic drive region in *Drosophila simulans.* Genetics **174:** 1365–1371.

Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark *et al.*, 2005 Genomic scans for selective sweeps using SNP data. Genome Res. **15:** 1566–1575.

Nielsen, R., I. Hellmann, M. Hubisz, C. Bustamante and A. G. Clark, 2007 Recent and ongoing selection in the human genome. Nat. Rev. Genet. **8:** 857–868.

Notley-McRobb, L., and T. Ferenci, 2000 Experimental analysis of molecular events during mutational periodic selections in bacterial evolution. Genetics **156:** 1493–1501.

Orr, H. A., 1998 The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution **52:** 935–949.

Palaisa, K., M. Morgante, S. Tingey and A. Rafalski, 2004 Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. Proc. Natl. Acad. Sci. USA **101:** 9885–9890.

Pennings, P. S., and J. Hermisson, 2006 Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genet. **2:** e186.

Perfeito, L., L. Fernandes, C. Mota and I. Gordo, 2007 Adaptive mutations in bacteria: high rate and small effects. Science **317:** 813–815.

Pool, J. E., V. Bauer DuMont, J. L. Mueller and C. F. Aquadro, 2006 A scan of molecular variation leads to the narrow localization of a selective sweep affecting both Afrotropical and cosmopolitan populations of *Drosophila melanogaster.* Genetics **172:** 1093–1105.

Przeworski, M., 2002 The signature of positive selection at randomly chosen loci. Genetics **160:** 1179–1189.

Przeworski, M., 2003 Estimating the time since the fixation of a beneficial allele. Genetics **164:** 1667–1676.

Przeworski, M., J. D. Wall and P. Andolfatto, 2001 Recombination and the frequency spectrum in Drosophila melanogaster and Drosophila simulans. Mol. Biol. Evol. **18:** 291–298.

Przeworski, M., G. Coop and J. D. Wall, 2005 The signature of positive selection on standing genetic variation. Evol. Int. J. Org. Evol. **59:** 2312–2323.

Roze, D., and N. H. Barton, 2006 The Hill-Robertson effect and the evolution of recombination. Genetics **173:** 1793–1811.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature **419:** 832–837.

Santiago, E., and A. Caballero, 2005 Variation after a selective sweep in a subdivided population. Genetics **169:** 475–483.

Stephan, W., T. H. E. Wiehe and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. **41:** 237–254.

Stephan, W., Y. S. Song and C. H. Langley, 2006 The hitchhiking effect on linkage disequilibrium between linked neutral loci. Genetics **172:** 2647–2663.

Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105:** 437–460.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Teshima, K. M., G. Coop and M. Przeworski, 2006 How reliable are empirical genomic scans for selective sweeps? Genome Res. **16:** 702–712.

Thornton, K. R., and J. D. Jensen, 2007 Controlling the false-positive rate in multilocus genome scans for selection. Genetics **175:** 737–750.

Thornton, K. R., J. D. Jensen, C. Becquet and P. Andolfatto, 2007 Progress and prospects in mapping recent selection in the genome. Heredity **98:** 340–348.

Voight, B. F., S. Kudaravalli, X. Wen and J. K. Pritchard, 2006 A map of recent positive selection in the human genome. PLoS Biol. **4:** e72.
Wakeley, J., and T. Takahashi, 2003 Gene genealogies when the sample size exceeds the effective size of the population. Mol. Biol. Evol. **20:** 208–213.
Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.
Zeng, K., S. Shi, Y. X. Fu and C. I. Wu, 2006 Statistical tests for detecting positive selection by utilizing high frequency SNPs. Genetics **174:** 1431–1439.

Communicating editor: J. Wakeley

APPENDIX A

Here, we want to describe the interactions between two selected loci and their neutral background. This can be tackled using the methodology of Barton and Turelli (1991) and Kirkpatrick *et al.* (2002), as done by Stephan *et al.* (2006). However, this is not fully necessary here, as the problem can be studied with a simple three-locus model: the two selected loci, plus a neutral locus. Different situations are investigated by changing the location of the neutral locus relative to the selected loci. In this appendix, we derive recursion for the gene frequencies at the three loci, the two- and three-locus linkage disequilibrium, and mean fitness.

For the sake of generality, let us consider three loci A, B, and C, located in that order on a chromosome, with r_1 (resp. r_2) the recombination rate between loci A and B (resp. B and C). The recombination rate between extreme loci A and C is then

$$R = r_1 + r_2 - 2r_1r_2. \tag{A1}$$

Assume that each locus has two alleles denoted by upper- and lowercase letters (A, a, B, b, C, and c). There are $K = 8$ possible gametic haplotypes:

$$\begin{array}{cccccccccc} \text{Gamete} & abc & abC & aBc & aBC & Abc & AbC & ABc & ABC \\ k & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}. \tag{A2}$$

Let x_k be the frequency of gamete haplotype k. We have

$$\sum_{k=1}^{K} x_k = 1 \tag{A3}$$

and from basic genetic definitions, we can write expressions for the frequencies of the uppercase genotype at one, two, and three loci:

$$\begin{array}{lll} p_A = x_5 + x_6 + x_7 + x_8 & p_{AB} = x_7 + x_8 & \\ p_B = x_3 + x_4 + x_7 + x_8 & p_{AC} = x_6 + x_8 & p_{\mathrm{ABC}} = x_8. \\ p_C = x_2 + x_4 + x_6 + x_8 & p_{BC} = x_4 + x_8 & \end{array} \tag{A4}$$

The two-locus linkage disequilibrium between loci A and B writes

$$C_{AB} = p_{AB} - p_Ap_B. \tag{A5}$$

The linkage disequilibria C_{AC} and C_{BC} for the two other pairs of loci are obtained seemingly by replacement, and finally the three-locus linkage disequilibrium is

$$C_{ABC} = p_{ABC} - C_{BC}p_A - C_{AC}p_B - C_{AB}p_C - p_Ap_Cp_B. \tag{A6}$$

There are 36 possible diploid genotypes $\{(i, j);\ i \leq j\}$ to consider. The probabilities $P(i, j, k)$ that a parent formed by the gametes i and j produces the gamete k after meiosis are given in Table A1. This table was derived using the Mathematica notebooks defined in Hospital *et al.* (1996).

Each locus may be selected or neutral depending on the case considered. This does not change the probabilities in Table A1, but simply the selection coefficient attributed to each locus.

The fitness of a diploid genotype composed of two gametic haplotypes i and j is

$$w(i, j) = (1 + X_{\mathrm{sel}_1}(i, j)s_1)(1 + X_{\mathrm{sel}_2}(i, j)s_2), \tag{A7}$$

where $X_{\mathrm{sel}_1}(i, j)$ is the number of copies of the favorable allele at selected locus sel_1, and similarly for sel_2, and where s_1 and s_2 are the corresponding selection coefficients. Note that in the text, we rather define the fitness (written with uppercase W) directly from X_{sel_1} and X_{sel_2} without reference to the haplotypes for the sake of clarity, so

TABLE A1

Probabilities of recombination at three loci during meiosis ($q = 1 - r$)

Parental gametes		Offspring gametes							
i	j	abc	abC	aBc	aBC	Abc	AbC	ABc	ABC
abc	abc	1	0	0	0	0	0	0	0
abc	abC	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	0	0
abc	aBc	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	0	0	0
abc	aBC	$\frac{1}{2}q_2$	$\frac{1}{2}r_2$	$\frac{1}{2}r_2$	$\frac{1}{2}q_2$	0	0	0	0
abc	Abc	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0	0	0
abc	AbC	$\frac{1}{2}Q$	$\frac{1}{2}R$	0	0	$\frac{1}{2}R$	$\frac{1}{2}Q$	0	0
abc	ABc	$\frac{1}{2}q_1$	0	$\frac{1}{2}r_1$	0	$\frac{1}{2}r_1$	0	$\frac{1}{2}q_1$	0
abc	ABC	$\frac{1}{2}q_1q_2$	$\frac{1}{2}q_1r_2$	$\frac{1}{2}r_1r_2$	$\frac{1}{2}q_2r_1$	$\frac{1}{2}q_2r_1$	$\frac{1}{2}r_1r_2$	$\frac{1}{2}q_1r_2$	$\frac{1}{2}q_1q_2$
abC	abC	0	1	0	0	0	0	0	0
abC	aBc	$\frac{1}{2}r_2$	$\frac{1}{2}q_2$	$\frac{1}{2}q_2$	$\frac{1}{2}r_2$	0	0	0	0
abC	aBC	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	0	0
abC	Abc	$\frac{1}{2}R$	$\frac{1}{2}Q$	0	0	$\frac{1}{2}Q$	$\frac{1}{2}R$	0	0
abC	AbC	0	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0	0
abC	ABc	$\frac{1}{2}q_1r_2$	$\frac{1}{2}q_1q_2$	$\frac{1}{2}q_2r_1$	$\frac{1}{2}r_1r_2$	$\frac{1}{2}r_1r_2$	$\frac{1}{2}q_2r_1$	$\frac{1}{2}q_1q_2$	$\frac{1}{2}q_1r_2$
abC	ABC	0	$\frac{1}{2}q_1$	0	$\frac{1}{2}r_1$	0	$\frac{1}{2}r_1$	0	$\frac{1}{2}q_1$
aBc	aBc	0	0	1	0	0	0	0	0
aBc	aBC	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0
aBc	Abc	$\frac{1}{2}r_1$	0	$\frac{1}{2}q_1$	0	$\frac{1}{2}q_1$	0	$\frac{1}{2}r_1$	0
aBc	AbC	$\frac{1}{2}r_1r_2$	$\frac{1}{2}q_2r_1$	$\frac{1}{2}q_1q_2$	$\frac{1}{2}q_1r_2$	$\frac{1}{2}q_1r_2$	$\frac{1}{2}q_1q_2$	$\frac{1}{2}q_2r_1$	$\frac{1}{2}r_1r_2$
aBc	ABc	0	0	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0
aBc	ABC	0	0	$\frac{1}{2}Q$	$\frac{1}{2}R$	0	0	$\frac{1}{2}R$	$\frac{1}{2}Q$
aBC	aBC	0	0	0	1	0	0	0	0
aBC	Abc	$\frac{1}{2}q_2r_1$	$\frac{1}{2}r_1r_2$	$\frac{1}{2}q_1r_2$	$\frac{1}{2}q_1q_2$	$\frac{1}{2}q_1q_2$	$\frac{1}{2}q_1r_2$	$\frac{1}{2}r_1r_2$	$\frac{1}{2}q_2r_1$
aBC	AbC	0	$\frac{1}{2}r_1$	0	$\frac{1}{2}q_1$	0	$\frac{1}{2}q_1$	0	$\frac{1}{2}r_1$
aBC	ABc	0	0	$\frac{1}{2}R$	$\frac{1}{2}Q$	0	0	$\frac{1}{2}Q$	$\frac{1}{2}R$
aBC	ABC	0	0	0	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$
Abc	Abc	0	0	0	0	1	0	0	0
Abc	AbC	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
Abc	ABc	0	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0
Abc	ABC	0	0	0	0	$\frac{1}{2}q_2$	$\frac{1}{2}r_2$	$\frac{1}{2}r_2$	$\frac{1}{2}q_2$
AbC	AbC	0	0	0	0	0	1	0	0
AbC	ABc	0	0	0	0	$\frac{1}{2}r_2$	$\frac{1}{2}q_2$	$\frac{1}{2}q_2$	$\frac{1}{2}r_2$
AbC	ABC	0	0	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
ABc	ABc	0	0	0	0	0	0	1	0
ABc	ABC	0	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
ABC	ABC	0	0	0	0	0	0	0	1

$$W(X_{\mathrm{sel}_1}, X_{\mathrm{sel}_2}) = w(i,j). \tag{A8}$$

The frequencies x' of the gametes at the next generation are obtained by

$$x'_k = \sum_{i=1}^{K}\sum_{j=i}^{K}\frac{1}{\bar{w}}\delta_{i,j}x_ix_jP(i,j,k)w(i,j) \tag{A9}$$

with

$$\bar{w} = \sum_{i=1}^{K}\sum_{j=i}^{K}\delta_{i,j}x_ix_jw(i,j) \tag{A10}$$

and

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 2 & \text{if } i \neq j, \end{cases} \tag{A11}$$

where $P(i, j, k)$ is taken from Table A1, $w(i, j)$ is the fitness of diploid genotype (i, j) computed as in (A7), and $\bar{w}$ is the mean fitness in the population.

Now, we can use (A9) to compute the various quantities defined in (A4)–(A6) from one generation to the next. We use a prime ($'$) to denote the quantity at the next generation. It turns out that for all these quantities, with the notable exception of the three-locus linkage disequilibrium (A6), the expression does not depend on the respective positions of the loci (*i.e.*, whether the neutral locus is "between" or "outside" the selected loci). Hence, without loss of generality we can write these quantities in terms of loci neu, sel_1, and sel_2 without taking account of their order on the chromosome. We get for the variation of gene frequency at any of the selected loci

$$\begin{aligned} \Delta p_{\text{sel}_i} &= p'_{\text{sel}_i} - p_{\text{sel}_i} \\ &= \frac{s_i p_{\text{sel}_i}(1 - p_{\text{sel}_i}) + s_j C_{\text{sel}_i,\text{sel}_j} + s_i s_j (C_{\text{sel}_i,\text{sel}_j} + 2p_{\text{sel}_i}(1 - p_{\text{sel}_i})p_{\text{sel}_j})}{\bar{w}}, \end{aligned} \tag{A12}$$

where i indicates the selected locus considered and j the other selected locus.

For the variation of gene frequency at the neutral locus we get

$$\begin{aligned} \Delta p_{\text{neu}} &= p'_{\text{neu}} - p_{\text{neu}} \\ &= \frac{s_1 C_{\text{sel}_1,\text{neu}} + s_2 C_{\text{sel}_2,\text{neu}} + s_1 s_2 (2p_{\text{sel}_1} C_{\text{neu},\text{sel}_2} + 2p_{\text{sel}_2} C_{\text{neu},\text{sel}_1} + C_{\text{neu},\text{sel}_1,\text{sel}_2})}{\bar{w}}. \end{aligned} \tag{A13}$$

And for the mean fitness,

$$\bar{w} = 1 + 2s_1 p_{\text{sel}_1} + 2s_2 p_{\text{sel}_2} + 2s_1 s_2 (C_{\text{sel}_1,\text{sel}_2} + 2p_{\text{sel}_1} p_{\text{sel}_2}). \tag{A14}$$

For the linkage disequilibrium at the next generation between the neutral locus and one selected locus, we get

$$\begin{aligned} C'_{\text{neu},\text{sel}_i} = {} & \frac{(C_{\text{neu},\text{sel}_i} C_{\text{sel}_1,\text{sel}_2} + C_{\text{neu},\text{sel}_j}(1 - p_{\text{sel}_i})p_{\text{sel}_i}) s_1 s_2}{\bar{w}} \\ & + \frac{(1 - r_{\text{neu},\text{sel}_i})\big(C_{\text{neu},\text{sel}_1,\text{sel}_2} s_j + C_{\text{neu},\text{sel}_i}(2p_{\text{sel}_j} s_j + 1)\big)(s_i + 1)}{\bar{w}} \\ & - \Delta p_{\text{neu}} \Delta p_{\text{sel}_i}. \end{aligned} \tag{A15}$$

And for the linkage disequilibrium at the next generation between the two selected loci,

$$\begin{aligned} C'_{\text{sel}_1,\text{sel}_2} = {} & \frac{(C_{\text{sel}_1,\text{sel}_2}(s_1 + 1)(s_2 + 1))(1 - r_{\text{sel}_1,\text{sel}_2})}{\bar{w}} \\ & + \frac{\Big(C^2_{\text{sel}_1,\text{sel}_2} + (p_{\text{sel}_1} - 1)p_{\text{sel}_1}(p_{\text{sel}_2} - 1)p_{\text{sel}_2}\Big) s_1 s_2}{\bar{w}} - \Delta p_{\text{sel}_1} \Delta p_{\text{sel}_2}. \end{aligned} \tag{A16}$$

Finally, the expression for the linkage disequilibrium at the next generation between three loci is too long to display. Instead, we give the frequency of the three-locus haplotype at the next generation,

$$\begin{aligned} p'_{\text{neu},\text{sel}_1,\text{sel}_2} = \frac{1}{\bar{w}} (& r_{\text{neu},\text{sel}_1}(1 + s_1) s_2 (C_{\text{neu},\text{sel}_2} C_{\text{sel}_1,\text{sel}_2} - C_{\text{neu},\text{sel}_1,\text{sel}_2} p_{\text{sel}_2}) \\ & + r_{\text{neu},\text{sel}_2} s_1 (1 + s_2)(C_{\text{neu},\text{sel}_1} C_{\text{sel}_1,\text{sel}_2} - C_{\text{neu},\text{sel}_1,\text{sel}_2} p_{\text{sel}_1}) \\ & - r_{\text{sel}_1,\text{sel}_2} C_{\text{sel}_1,\text{sel}_2} p_{\text{neu}} (s_1 + 1)(s_2 + 1) \\ & - (1 - \gamma) C_{\text{neu},\text{sel}_1,\text{sel}_2} (s_1 + 1)(s_2 + 1) \\ & - r_{\text{neu},\text{sel}_2} C_{\text{neu},\text{sel}_2} p_{\text{sel}_1} \big((p_{\text{sel}_1} + 1)s_1 + 1\big)(s_2 + 1) \\ & - r_{\text{neu},\text{sel}_1} C_{\text{neu},\text{sel}_1} p_{\text{sel}_2} (s_1 + 1)\big((p_{\text{sel}_2} + 1)s_2 + 1\big) \\ & + \frac{1}{2}(p_{\text{neu},\text{sel}_1,\text{sel}_2}(1 + \bar{w} + s_1 + s_2 + (1 + p_{\text{sel}_1} + p_{\text{sel}_2} - p_{\text{sel}_2} p_{\text{sel}_1}) s_1 s_2))), \end{aligned} \tag{A17}$$

where

$$(1-\gamma) = (1-r_1)(1-r_2); \tag{A18}$$

i.e., $(1-\gamma)$ is the probability that there is no recombination either between the first and second locus or between the second and third locus on the chromosome. Note that $(1-\gamma)$ is the only term that depends on the positions of the loci on the chromosome, whereas all other terms depend only on the status of the loci (neutral or selected).

One can then plug (A17), as well as the recursion for two-locus linkage disequilibrium, and gene frequencies into (A6).

APPENDIX B

This appendix explains the calculation of the scaled delay time between the beneficial mutations at sel_1 and sel_2. This scaled time is used to account for the dynamics of the beneficial allele at sel_1, which spends more time at low and high frequencies than at intermediate frequencies. Taking advantage of the fact that

$$p_{sel_1}/(1-p_{sel_1}) \simeq \varepsilon(1-\varepsilon)\exp(s_1 t),$$

where ε is the beginning of the deterministic phase, the expected trajectory of the first beneficial mutation is

$$p_{sel_1} = \frac{\varepsilon}{\varepsilon + (1-\varepsilon)\exp(-s_1 t)},$$

which coincides with Stephan *et al.*'s (1992) Equation 3a. By reversing this equation, the expected time before reaching a frequency p_{sel_1} is

$$t(p_{sel_1}) = \frac{\log(p_{sel_1}/(1-p_{sel_1})) - \log(\varepsilon/(1-\varepsilon))}{s_1}.$$

The threshold frequency of the beneficial mutation at sel_1 when the beneficial mutation at sel_2 is introduced is p_t. We define the scaled delay time between the selected mutations as

$$\tau = \frac{t(p_t)}{t(1-\varepsilon)} = \frac{\log(p_t/(1-p_t)) - \log(\varepsilon/(1-\varepsilon))}{2\log(\varepsilon/(1-\varepsilon))}.$$

It is the proportion of the expected total duration of the deterministic phase reached by the selective sweep at sel_1, when the beneficial mutation occurs at sel_2. τ has meaning only for frequencies at which the dynamics of the selected locus are nearly deterministic; therefore $\tau = 0$ corresponds to $p_t = \varepsilon$ and $\tau = 1$ to $p_t = 1-\varepsilon$ (ε was set to 40/20,000 on the basis of the observed course of the simulations). Note that the actual "simultaneous" case treated above ($p_t = 1/(2N)$) is not strictly equivalent to $\tau = 0$, as it includes the frequencies at which the dynamics are governed by the stochastic process.