

Joint Effects of Genetic Hitchhiking and Background Selection on Neutral Variation

Yuseob Kim and Wolfgang Stephan

Department of Biology, University of Rochester, Rochester, New York 14627

Manuscript received December 7, 1999
Accepted for publication March 20, 2000

ABSTRACT

Due to relatively high rates of strongly selected deleterious mutations, directional selection on favorable alleles (causing hitchhiking effects on linked neutral polymorphisms) is expected to occur while a deleterious mutation-selection balance is present in a population. We analyze this interaction of directional selection and background selection and study their combined effects on neutral variation, using a three-locus model in which each locus is subjected to either deleterious, favorable, or neutral mutations. Average heterozygosity is measured by simulations (1) at the stationary state under the assumption of recurrent hitchhiking events and (2) as a transient level after a single hitchhiking event. The simulation results are compared to theoretical predictions. It is shown that known analytical solutions describing the hitchhiking effect without background selection can be modified such that they accurately predict the joint effects of hitchhiking and background on linked, neutral variation. Generalization of these results to a more appropriate multilocus model (such that background selection can occur at multiple sites) suggests that, in regions of very low recombination rates, stationary levels of nucleotide diversity are primarily determined by hitchhiking, whereas in regions of high recombination, background selection is the dominant force. The implications of these results on the identification and estimation of the relevant parameters of the model are discussed.

IT has been suggested that the “hitchhiking effect” of a strongly selected allele on the frequencies of neutral DNA polymorphisms at linked loci may play an important role in determining the patterns of genetic variation across eukaryotic genomes (Maynard Smith and Haigh 1974; Ohta and Kimura 1975; Kaplan *et al.* 1989; Stephan *et al.* 1992; Gillespie 1994; Barton 1998). Furthermore, it has been shown that, on the basis of the observed patterns of variation, the relevant parameters of the underlying selective process can be estimated (Wiehe and Stephan 1993; Stephan 1995). As predicted by the theory of hitchhiking (Birky and Walsh 1988, and references above), the level of genetic variation is usually positively correlated with the rate of recombination, but divergence between closely related species is nearly unaffected by recombination (Begun and Aquadro 1992). On the other hand, the theory of “background selection” (leading to a reduction of effective population size by recurrent deleterious mutations) proposed by Charlesworth *et al.* (1993) makes qualitatively similar predictions (Hudson and Kaplan 1995; Nordborg *et al.* 1996). Hudson and Kaplan (1995) and Charlesworth (1996) showed that background selection can explain genome-wide patterns in *Drosophila melanogaster* polymorphism data. However, a

similarly good fit has been obtained when the hitchhiking model alone is applied to the same data set (Stephan 1995). Other studies supported hitchhiking over background selection for explaining patterns of genetic variation at (mostly) individual loci (Schlötterer *et al.* 1997; Nurminsky *et al.* 1998; Stephan *et al.* 1998; Benassi *et al.* 1999), whereas in some cases background selection was thought to be sufficient in explaining the results. Thus, it appears that the relative importance of hitchhiking and background selection in determining the level of genetic variation remains essentially unknown.

Previous studies of the hitchhiking effect used a model in which recurrent deleterious mutations at linked loci are not included. However, since the rate of deleterious mutations is believed to be high (Keightley and Eyre-Walker 1999), hitchhiking events are likely to occur in chromosomal regions where the standing level of variation is already reduced by background selection (Charlesworth 1996). Therefore, the effect of background selection on the process of hitchhiking should be investigated to assess the relative importance of these two forces.

Peck (1994) and Barton (1995) studied the reduction of the fixation probability of strongly selected alleles due to background selection. Stephan *et al.* (1999) derived results for the effect of background selection on the nucleotide diversity and fixation probability at a partially linked, weakly selected locus. But genetic variation at a neutral locus that is partially linked to

Corresponding author: Yuseob Kim, Department of Biology, University of Rochester, Rochester, NY 14627.
E-mail: yuse@troi.cc.rochester.edu

both a locus under strong positive selection and a locus under background selection has not been investigated. In the latter case, the dynamics of favorable and deleterious alleles may interfere with each other, making it difficult to predict the outcome of this interaction by analyzing these processes individually. In this article, we investigate this problem using simulations based on three-locus, two-allele models. We measure heterozygosity at a neutral locus using two different models of genetic hitchhiking. In the first model, we analyze the stationary level of heterozygosity caused by background selection and recurrent hitchhiking events; in the second one, we study the effect of background selection and a single hitchhiking event on heterozygosity. The results of this analysis have significant implications for our understanding of selective processes in natural populations and for the identification and estimation of relevant parameters of these processes.

STATIONARY LEVEL OF HETEROZYGOSITY CAUSED BY BACKGROUND SELECTION AND RECURRENT HITCHHIKING EVENTS (MODEL 1)

In this section, we investigate the stationary level of heterozygosity determined by recurrent substitutions of favorable alleles and continuous removal of deleterious alleles by background selection. We use a simple discrete-generation model of a diploid population of size N . (Note that a list of parameters is provided in Table 1.) Since we assume that the fitness effects of alleles within a locus combine multiplicatively, this model is equivalent to that of a haploid population of size $2N$ that undergoes random conjugation and recombination. We consider a three-locus model such that the three loci are located on a chromosome in the following order: The first locus (*Del*) experiences recurrent deleterious mutations. The mutation of the wild-type allele (A) to a deleterious allele (a) with selective disadvantage t occurs at a rate u (per gene per generation). That is, as in Stephan *et al.* (1999), background selection acts at a single locus. The second locus (*Fav*) is under positive selection such that a favorable allele (B) with selection coefficient s is introduced as described below. Mutation from an ancestral (m) to a derived neutral allele (M) occurs at the third locus (*Neu*). The genetic background (haplotype) on which the mutations to B and to M occur is chosen randomly in proportion to its frequency. The recombination fractions between *Del* and *Fav* and between *Fav* and *Neu* are r_1 and r_2 , respectively. The two other possibilities of gene order, *i.e.*, *Del-Neu-Fav* and *Fav-Del-Neu*, were also studied. For all three gene orders, the simulation and theoretical methods are similar. So we focus on the case of *Del-Fav-Neu*.

Simulation methods: For this model, there are eight possible haplotypes (ABM , ABm , AbM , Abm , aBM , aBm , abM , and abm) in the population. Therefore, the dynamics of the system can be completely described by the

TABLE 1

Definitions of parameters

N	Diploid population size
N_1, N_2	Effective population size at the <i>Fav</i> and the <i>Neu</i> locus, respectively
s, t	Selection coefficients of favorable and deleterious alleles, respectively
r_1, r_2	Recombination fraction between <i>Del</i> and <i>Fav</i> and between <i>Fav</i> and <i>Neu</i> , respectively
α	Strength of directional selection = $2Ns$
u	Rate of deleterious mutations per generation per locus
u_f	Rate of favorable mutations per generation per population
μ	Rate of neutral mutations per generation per locus (nucleotide)
Π	Reduction factor of the fixation probability of favorable mutations
Φ	Fixation probability of favorable mutations
f_H, f_B	Reduction factors of the stationary level of heterozygosity due to recurrent hitchhiking events and background selection, respectively
h	Reduction of heterozygosity due to a single hitchhiking event
π	Heterozygosity per nucleotide
θ	Expected heterozygosity without hitchhiking (= $4N_e\mu$)
H	Cumulative heterozygosity of one trajectory of a neutral allele
p_a, p_B	Allele frequencies of a and B , respectively
ρ	Recombination rate per nucleotide per generation
τ	Number of generations until the last hitchhiking event
ν	Number of advantageous substitutions per generation per nucleotide

changes of eight haplotype frequencies. It is straightforward to derive a set of equations describing the deterministic change of haplotype frequencies by selection, recombination, and deleterious mutations (appendix a). When one copy of B or M is introduced in the population, haplotype frequencies are changed accordingly before they are subjected to selection. To incorporate the effects of finite population size, multinomial sampling of different haplotypes was simulated after their frequencies were changed according to the deterministic equations. We used the random binomial number generator of Press *et al.* (1992) with some modification.

Each simulation starts with a population of Abm and abm haplotypes. The initial frequency of a is given as u/t . Then, one copy of the favorable allele (B) is introduced in the population at rate u_f if the population is fixed for b (note that u_f is a mutation rate per population per generation, whereas u is per gene per generation). If B is fixed, all B 's are converted to b . Therefore, with $u, u_f > 0$, a mutation-selection balance at the *Del* locus

and occasional directional selection at the *Fav* locus occur simultaneously. To measure the standing level of genetic variation at *Neu*, we used the method suggested by Charlesworth *et al.* (1993). An allele *M* is introduced at the beginning of each simulation run and introduced again whenever it is lost in the population by drift. If *M* is fixed in the population, at the next generation all *M*s are converted to *m* and then another *M* is introduced. The frequency of *M*, *y*, is monitored until *M* is lost or fixed. During this period, heterozygosity, $2y(1 - y)$, is summed over generations. The expected value of this sum, *H*, in the neutral model is 2.0 (Kimura 1969, 1971). Heterozygosity per site per generation, π , for a given hypothetical process of mutation, with rate μ , is then obtained by “spreading out” trajectories of *M* over time according to the mutation process; *i.e.*, $\pi = 2N\mu H$. To observe the change of the level of genetic variation, we only need to measure *H* without modeling a specific mutation process (Charlesworth *et al.* 1993).

This procedure is based on the principle of ergodicity, which says that averaging a random variable of a stationary stochastic process over time leads to the same result as averaging this quantity at any given time point over different realizations of the process. Assuming that the selective phase during hitchhiking is very short, we may consider each selective sweep instantaneous. Since these hitchhiking events are modeled as a time-homogeneous Poisson process, their effect is a shortening of the trajectories of the neutral allele *M*, independent of time. As a consequence, the expectation of *H* can be found by evaluating this random variable at arbitrary time points during a particular realization of the process and by averaging over these values, or by evaluating an ensemble of realizations of the process at a particular time. A total of 10^8 introductions of *M* were made consecutively in each simulation run and the mean value of *H* was obtained. The mean numbers of generations until loss and until fixation were also recorded.

Theoretical predictions: We combined the previous theoretical results of hitchhiking and background selection using the following assumptions. Deleterious mutations occur very frequently, leading to the establishment of a mutation-selection balance. We assume that this mutation-selection balance at *Del* affects the fixation probability and the effective population size at the *Fav* locus. On the other hand, we assume that directional selection at *Fav* has no influence on the mutation-selection balance at *Del*. Therefore, the equilibrium frequency of the deleterious allele (*a*) is maintained during the selective phase. The combined effects of background selection and hitchhiking on heterozygosity may therefore be approximated by well-known formulas (Kaplan *et al.* 1989; Stephan *et al.* 1992; Wiehe and Stephan 1993) that describe the effects of hitchhiking on neutral variation as a function of linkage, effective population size, and strength of selection, except that

these latter quantities have to be corrected to allow for the occurrence of background selection.

The effect of background selection at a single locus on a linked locus is given by

$$f_B(r) = 1 - \frac{u}{\lambda(1 + r(1 - \delta)/\lambda)^2} \quad (1)$$

(Hudson and Kaplan 1995; Nordborg *et al.* 1996), where $f_B(r)$ is the relative reduction of heterozygosity or effective population size as a function of the recombination fraction, *r*, between the two loci. Then, without hitchhiking, heterozygosity at *Neu* is predicted to be $\pi_0 f_B(r_1 + r_2)$, where π_0 is the average nucleotide heterozygosity in the neutral equilibrium model. To incorporate the effect of hitchhiking, it follows from the above considerations that the strength of directional selection at *Fav* is characterized by $\alpha_1 = 2N_1s$, where $N_1 = Nf_B(r_1)$, and that the fixation probability of *B* is

$$\Phi = (1 - e^{-2s})\Pi, \quad (2)$$

where $\Pi = \Pi(u, t, s, r_1)$ is the reduction of the fixation probability due to background selection (may be obtained by solving Equations 15a and 15b of Barton 1995 numerically). To obtain the effect of repeated substitutions at *Fav* on average heterozygosity at *Neu*, we use a theoretical argument by Wiehe and Stephan (1993). The reduction of expected heterozygosity by a single hitchhiking event is given by

$$h = h(r_1, r_2) = \frac{2r_2}{s} \alpha_1^{-2r_2/s} \Gamma\left(-\frac{2r_2}{s}, \frac{1}{\alpha_1}\right), \quad (3)$$

(Stephan *et al.* 1992), where $\Gamma(\dots)$ is the incomplete gamma function. Using coalescent arguments (Kaplan *et al.* 1989; Wiehe and Stephan 1993), *h* can be related to a sample quantity. For a sample of size 2, consider a coalescent process, with time running backward. Then it can be shown that *h* is equal to the probability of entering the selective phase with two ancestral genes and exiting it with two ancestral genes. In other words, *h* is equal to the probability that the neutral locus escapes hitchhiking by recombining away from the favored allele, while the latter is on its way to fixation. The rate of hitchhiking events per generation experienced by the *Neu* locus is therefore given by $1 - h$ times the probability of selective fixations at the *Fav* locus, which is $u_f\Phi$. As the history of the sample is traced back, regular coalescent events may occur at the rate of $1/2N_2$, parallel to hitchhiking events, where $N_2 = Nf_B(r_1 + r_2)$. Without hitchhiking, the expected time to the most recent common ancestor of a sample of size 2 is $2N_2$. Therefore, on the time scale of $2N_2$ generations, the rate of occurrence of a coalescent or a hitchhiking event is $1 + 2N_2 u_f\Phi(1 - h)$, and the expected time back until one of these events occurs is $1/[1 + 2N_2 u_f\Phi(1 - h)]$. Thus, due to hitchhiking the expected total time of the genealogy of the *Neu* locus is reduced by the factor

$$f_H(r_1, r_2) = \frac{1}{1 + 2N_2 u_f \Phi(1 - h)}. \quad (4)$$

Thus, combining the effects of background selection and hitchhiking, the value of H is predicted to be

$$E[H] = 2f_B(r_1 + r_2) f_H(r_1, r_2). \quad (5)$$

In summary, we predicted the combined effect of hitchhiking and background selection by using a previously known analytic solution of the hitchhiking effect and by modifying the effective population sizes at the linked loci and the fixation probability of the favorable allele such that background selection is taken into account.

Simulation results: Table 2 shows the results of simulations for the gene arrangement *Del-Fav-Neu* in model 1. In most cases, we used $t = 0.02$, which is close to the mean heterozygote effect of deleterious mutations estimated from *D. melanogaster* (Crow and Simmons 1983). The choices of s , r_1 , r_2 , u , and u_f are rather arbitrary. Since simulation time increases with population size, we used $2N = 10^5$, which is smaller than typical *Drosophila* population sizes. We obtained accurate results for H , the fixation probability of B , and the mean time to loss (T_0) and fixation (T_1) of M expected under standard theories (simulations 1-1, 1-4, and 1-8). The fixation probability of M was close to $1/2N$ in all the simulation runs for model 1 (data not shown), which agrees with the fact that selection at linked loci does not affect the fixation probability of neutral mutants (Birky and Walsh 1988). The H values obtained for all parameter values agree well with our simple theoretical predictions (Equation 5). This result indicates that the effects of background selection and hitchhiking can be combined in a predictable way. However, the combined effects are not simply multiplicative. For example, comparing simulations 1-3, 1-4, and 1-6, the H values were reduced by the factors of 0.67 and 0.84 by background selection and hitchhiking, respectively, when each process took place without the other. However, when both processes occurred at the same time, the reduction factor of H was 0.61, larger than the product $0.67 \times 0.84 = 0.57$. This discrepancy can be fully explained by the expected change of f_H from 0.836 (for 1-4) to 0.894 (for 1-6). This suggests that the effect of hitchhiking diminishes with background selection. This nonmultiplicative combination of hitchhiking and background selection produces even more striking results when the effect of hitchhiking is very strong: In simulations 1-9 and 1-10, it is shown that the reduction of heterozygosity is smaller when hitchhiking is combined with background selection than when hitchhiking occurs in the absence of background selection. Simulations for the other arrangements of genes, *i.e.*, *Del-Neu-Fav* and *Fav-Del-Neu*, gave qualitatively the same results (data not shown).

In simulations 1-13, 1-14, and 1-15, where $s > t$, the equilibrium frequency of the deleterious allele at *Del* is likely to be perturbed by the substitution of B , which

violates the assumptions of our theoretical predictions. However, for these parameter values theoretical predictions given by (5) still agree well with simulation results (Table 2). We further address this problem for model 2 below.

Generalizations and implications: The results obtained from our three-locus analysis suggest the following generalizations. Consider a chromosome of a finite physical length throughout which the recombination rate per nucleotide per generation, ρ , is constant. A position on the chromosome is described by the number of nucleotides, l , away from the reference locus *Neu*, where a positive (negative) value of l defines a locus to the right (left) of *Neu*. *Neu* is located l_L and l_R nucleotides away from the left and right ends of the chromosome, respectively. Deleterious mutations can occur at any position along the chromosome at a rate u (per nucleotide per generation). Single hitchhiking events may occur according to a time-homogeneous Poisson process caused by advantageous substitutions at randomly chosen loci. The question we ask is, What is the joint effect of these forces on neutral polymorphism at *Neu*? Equation 5 suggests that nucleotide diversity, π , can be approximated as

$$\pi \approx \pi_0 f_B(\rho) \frac{\rho}{\rho + k\alpha f_B(\rho) \nu(\rho)}, \quad (6)$$

where π_0 is the neutral equilibrium value of nucleotide diversity, $\alpha = 2Ns$, k is a constant (see below), $\nu(\rho)$ is the expected number of selected substitutions per nucleotide per generation at *Neu*, and $f_B(\rho)$ is the reduction factor of effective population size at *Neu* due to background selection.

Equation 6 can be derived in a similar way as the corresponding equation without background selection (Wiehe and Stephan 1993). Let $\nu(l, \rho)$ and $\alpha(l, \rho)$ be the expected number of advantageous substitutions and the strength of selection at a position l , respectively. The rate at which a neutral polymorphism at *Neu* undergoes hitchhiking caused by selected substitutions between l and $l + dl$ nucleotides away is then given by $\nu(l, \rho) [1 - h(l, \rho)] dl$, where

$$h(l, \rho) = \frac{2\rho|l|}{s} \alpha(l, \rho)^{-2\rho|l|/s} \Gamma\left(-\frac{2\rho|l|}{s}, \frac{1}{\alpha(l, \rho)}\right). \quad (7)$$

The equation above assumes that the recombination fraction scales linearly with physical distance, which can be justified as the effect of hitchhiking is limited to a small genetic distance. The expected number of selected substitutions causing a hitchhiking effect at *Neu* is then obtained as

$$k_h = \int_{-M^*/\rho}^{M^*/\rho} \nu(l, \rho) [1 - h(l, \rho)] dl \quad (8a)$$

$$\approx \nu(\rho) \int_{-M^*/\rho}^{M^*/\rho} [1 - h(l, \rho)] dl, \quad (8b)$$

TABLE 2
Results for model 1 (gene order: *Del-Fav-Neu*)

	s	t	u/t	u_f	r_1	r_2	Φ^a	H ^b	T_0^c	T_1^c
1-1	0.02	0.02	0	0	—	—	—	1.9718 ± 0.0295 (2)	22.41	197,890
1-2	0.02	0.02	0.2	0	10 ⁻³	10 ⁻³	—	1.6835 ± 0.0308 (1.6682)	20.10	168,797
1-3	0.02	0.02	0.4	0	10 ⁻³	10 ⁻³	—	1.3410 ± 0.0245 (1.3364)	17.88	135,652
1-4	0.02	0.02	0	10 ⁻⁴	—	10 ⁻³	0.0394 (0.0392)	1.6866 ± 0.0300 (1.6724)	21.70	166,386
1-5	0.02	0.02	0.2	10 ⁻⁴	10 ⁻³	10 ⁻³	0.0353 (0.0353)	1.4599 ± 0.0263 (1.4506)	19.71	144,445
1-6	0.02	0.02	0.4	10 ⁻⁴	10 ⁻³	10 ⁻³	0.0337 (0.0339)	1.2258 ± 0.0222 (1.1949)	17.64	121,186
1-7	0.02	0.02	0.2	5 × 10 ⁻⁴	10 ⁻³	10 ⁻³	0.0350 (0.0353)	0.9623 ± 0.0164 (0.9532)	18.47	94,472
1-8	0.02	0.02	0	1.86 × 10 ^{-3d}	—	10 ⁻³	0.0388 (0.0392)	0.4212 ± 0.0061 (0.4307)	17.73	43,165
1-9	0.02	0.02	0	4.16 × 10 ^{-3d}	—	10 ⁻⁴	0.0391 (0.0392)	0.1178 ± 0.00104 (0.1214)	15.30	10,014
1-10	0.02	0.02	0.2	4.20 × 10 ^{-3d}	10 ⁻⁴	10 ⁻⁴	0.0340 (0.0343)	0.1319 ± 0.00567 (0.1360)	14.49	11,068
1-11	0.002	0.02	0.2	0.97 × 10 ^{-3d}	2 × 10 ⁻³	10 ⁻⁴	0.00339 (0.00335)	1.4698 ± 0.0264 (1.4243)	19.70	144,497
1-12	0.0002	0.02	0.2	7.84 × 10 ^{-3d}	2 × 10 ⁻³	10 ⁻⁵	0.000329 (0.000334)	1.4649 ± 0.0260 (1.4201)	19.76	149,882
1-13	0.02	0.01	0.2	2 × 10 ⁻⁴	10 ⁻³	10 ⁻³	0.0381 (0.0380)	1.3013 ± 0.0189 (1.2914)	19.69	126,672
1-14	0.02	0.01	0.4	2 × 10 ⁻⁴	10 ⁻³	10 ⁻³	0.0373 (0.0376)	1.1417 ± 0.0170 (1.1251)	18.42	112,261
1-15	0.02	0.005	0.4	2 × 10 ⁻⁴	10 ⁻³	10 ⁻³	0.0385 (0.0388)	1.1946 ± 0.0176 (1.2066)	19.33	122,904

Each simulation result is obtained by 10⁸ consecutive introductions of M; 2N = 10⁵.

^a Observed (predicted) probability of fixation of allele B.

^b Observed mean ± standard error of H; the predicted value of H (Equation 5) is in parentheses.

^c Observed means of T_0 and T_1 , respectively.

^d u_f is the actual number of beneficial mutations divided by the total number of generations in a simulation.

where M^* is the maximal recombination distance allowed (see Wiehe and Stephan 1993). The approximation leading to (8b) works because the hitchhiking term $[1 - h(l, \rho)]$ has its maximum at $l = 0$ and declines rapidly to zero for $l \neq 0$ relative to the function $v(l, \rho)$, which varies slowly with l . The integral in (8b) can be evaluated for constant $\alpha(\dots)$, namely $\alpha(\dots) = \alpha_0$, because the integrand is insensitive to $\alpha(\dots)$. Therefore, we find approximately

$$k_h \approx \frac{ksv(\rho)}{\rho}, \quad (9a)$$

where k denotes the integral

$$k = \int_{-2M/\alpha_0}^0 \left[1 + u\alpha_0^u \Gamma\left(u, \frac{1}{\alpha_0}\right) \right] du, \quad (9b)$$

and $M = 2NM^*$. As shown in Wiehe and Stephan (1993), k depends only weakly on α_0 .

Assuming deleterious mutations have a uniform selective disadvantage, t , and effects of background selection at many loci combine multiplicatively, it can be shown that

$$f_b(\rho) \approx \exp\left[-\frac{u}{t} \int_{-l}^R \frac{1}{\{1 + \rho|x|(1 - \hat{t})/\hat{t}^2\}} dx\right] \quad (10)$$

(Nordborg *et al.* 1996) and

$$v(\rho) \approx v_0 \exp\left[-\frac{u}{t} \int_{-l}^R \left\{ \frac{2t}{\sqrt{(t+s+\rho|x|)^2 - 4st + t+s+\rho|x|}} \right\}^2 dx\right] \quad (11)$$

(Barton 1995, Equation 17a), where v_0 is the expected number of selected substitutions per nucleotide on a free-recombining chromosome. Incorporating (10) and (11) into (6), Figure 1 shows the relationship of nucleotide diversity π/π_0 and recombination rate ρ for background selection, hitchhiking, and for the joint action of both processes. It is noteworthy that for low recombination rates the function describing the joint effects of background selection and hitchhiking approaches that of hitchhiking alone, whereas for higher recombination rates background selection is the dominant force determining the level of diversity. This is because in regions of low recombination the background selection terms $f_b(\rho)$ cancel, unless background selection is extremely strong and/or the rate of favorable substitutions is extremely low, such that $f_b(\rho)v(\rho)$ converges to zero. The expression resulting from (6) for small ρ suggests that background selection affects only the fixation probability of advantageous mutations, not their strength. As a consequence, background selection may reduce the effect of hitchhiking; that is, lower levels of nucleotide diversity are expected in regions of very low recombination when hitchhiking operates alone than in the presence of background selection. A similar result was found for our three-locus model (see above).

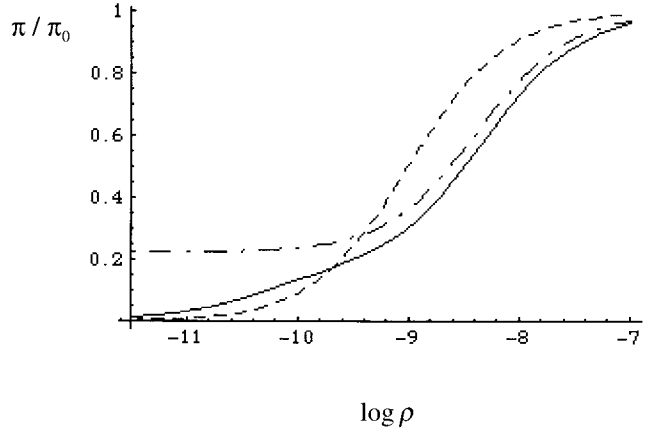


Figure 1.—Relative nucleotide diversity, π/π_0 , against per-nucleotide recombination rate, ρ . The model described in the text is used with $t = s = 0.02$, $u = 1.5 \times 10^{-9}$, $l_L = l_R = 10^7$, $k\alpha v_0 = 10^{-9}$. The continuous line (—), produced by Equations 6, 10, and 11, represents the joint effect of hitchhiking and background selection. The graphs of relative diversity determined by background selection alone (- · -) and hitchhiking alone (- -) are produced by $\pi = \pi_0 f_b(\rho)$ (Equation 10) and $\pi = \pi_0 v(\rho) / (\rho + k\alpha v_0)$ (Wiehe and Stephan 1993), respectively.

TRANSIENT PATTERNS OF HETEROZYGOSITY AFTER A SINGLE HITCHHIKING EVENT UNDER THE INFLUENCE OF BACKGROUND SELECTION (MODEL 2)

We use the same three-locus model as above, but assume different mutational processes for the *Fav* and *Neu* loci. We are interested in transient patterns of heterozygosity at *Neu* at given time points after a hitchhiking event, rather than the stationary level of heterozygosity measured in model 1. We define time, T , as the number of generations before present ($T = 0$). One selected substitution, from b to B , at *Fav* occurs at a fixed time in the past ($T = \tau$). It is assumed that the previous substitution at *Fav* took place very long ago such that the level of genetic variation at *Neu* has recovered to its equilibrium level before the selected substitution occurs at $T = \tau$; *i.e.*, we are analyzing the effect of a single hitchhiking event. At the *Neu* locus, mutant alleles, M , are introduced in the past at $T = \lambda$ ($0 < \lambda < \infty$), such that they may occur at any generation with an equal rate, μ (per gene). Each mutant has a certain probability of still segregating at $T = 0$, thus contributing to heterozygosity at $T = 0$. Therefore, heterozygosity can be determined by adding up all the contributions made by the mutations in the past (Kimura 1969, Equation 5). This suggests a simulation procedure in which heterozygosity at $T = 0$ is measured by summing over all trajectories of M that occurred in the past. That is, in each simulation run we follow the trajectory of allele M and record its frequency at $T = 0$. By repeating this procedure, we obtain the distribution of the frequency of M and thus the average heterozygosity at $T = 0$. As in model 1, recurrent deleterious mutations occur at the *Del* locus.

Simulation methods: The recurrence equations and multinomial sampling method that were described in model 1 are used to simulate the dynamics of allele frequency changes. In the simplest way, model 2 can be simulated by introducing a neutral mutant, M , at $T = \lambda$, where λ is uniformly distributed between 0 and L ($\gg N$), and by recording its frequency at $T = 0$. If M is lost or fixed before $T = 0$, its frequency is recorded as zero. Thus, the frequency distribution of M at $T = 0$ can be obtained by repeating the above procedure many times. However, this straightforward simulation scheme is too time-consuming because the length of the time window, L , has to be set to a large value. To circumvent this problem, we use the following procedures, which allow us to keep L reasonably small.

For simplicity, consider first the model without background selection. We assume that, at the beginning of the time window ($T = L$), the population is in mutation-drift equilibrium such that the number and the frequency distribution of a segregating allele at Neu can be described by the standard neutral theory (Kimura 1983). Then, we introduce mutants at $T = L$ with initial frequency $i/2N$ ($i = 1, \dots, 2N - 1$), where the probability of frequency $i/2N$ is proportional to $1/i$. The rest of the mutants are introduced at $T = \lambda$ with initial frequency $1/2N$, where λ is uniformly distributed between 0 and $L - 1$. The ratio of mutants appearing at $T = L$ and $T = \lambda$ is adjusted such that the expected number of segregating sites is at equilibrium. For a given mutation rate (per generation per locus), μ , the probability that allele M is segregating at the Neu locus at $T = L$ is given by θa_{2N} , where $\theta = 4N\mu$ and $a_n = \sum_{i=1}^n 1/i$. The expected number of new mutants occurring between 0 and L is $2N\mu L$. Therefore, the proportion of mutants, δ , introduced at $T = L$ is $2a_{2N}/(2a_{2N} + L)$. Since one mutant occurs in each simulation run, the value of μ is given by $(1 - \delta)/(2NL)$. The expected value of heterozygosity at $T = 0$ at equilibrium (without hitchhiking) contributed by each M is θ (see appendix b). The favorable mutation at the Fav locus occurs at a fixed generation ($\tau < L$) before present. The mutation to the favorable allele (B) occurs if a neutral mutant that appeared at $T = \lambda$ ($> \tau$) is still segregating in the population at $T = \tau$. If B is lost, haplotype frequencies at $T = \tau$ (immediately before the favorable mutation occurred) are restored and the mutation to B occurs again. This procedure is repeated until B is fixed. While B is segregating, the frequency of M changes but the time counter T is arrested. It resumes decreasing after B is fixed. Therefore, this procedure simulates a situation where neutral allele frequencies change instantaneously in one generation by a hitchhiking effect. For example, if M is lost or fixed during the substitution of B , its frequency is recorded as 0 or 1, respectively, at $T = \tau - 1$.

To incorporate background selection on Del , the same procedure is used but with the following changes. We assume that, when neutral mutants are introduced, the

population is in a deleterious mutation-selection equilibrium. Therefore, the frequency of the deleterious allele (a), q , is set to u/t when each replicate of the simulation starts with a new introduction of M . One copy of M randomly associates with A or a at $T = \lambda$. The expected number of segregating sites at $T = L$ is now $\sim 4N_2\mu a_{2K}$, where K is the closest integer to $N_2 = N_{\text{eff}}^*(r_1 + r_2)$ (see Equation 1); the latter is the effective population size at Neu when background selection has been taken into account. δ is now given by $2N_2a_{2K}/(2N_2a_{2K} + NL)$. The initial frequency of M at $T = L$ is $i/2K$ ($i = 1, \dots, 2K - 1$), where the probability of frequency $i/2K$ is proportional to $1/i$. We assume no linkage disequilibrium between Del and Neu at $T = L$; thus the frequency of the AM haplotype, for example, is given by $(1 - u/t)(i/2K)$. However, we used $L > \tau + 10^3$, so that haplotype frequencies immediately before the hitchhiking event depend little on the initial frequencies.

Theoretical predictions: To predict average heterozygosity (π) at $T = 0$ for model 2, we used the same approach as for model 1 by modifying the effective population sizes at the linked loci due to background selection. A simple solution can be obtained as

$$E[\pi] = \int_0^1 2y(1 - y)\varphi(y)he^{-\tau/2N_2}dy + 2N\mu \int_0^\tau 2\frac{1}{2N}\left(1 - \frac{1}{2N}\right)e^{-t/2N_2}dt, \quad (12)$$

where $h = h(r_1, r_2)$ and $N_2 = N_{\text{eff}}^*(r_1 + r_2)$, as defined above, and $\varphi(y)$ is the distribution of mutant allele frequencies immediately before the hitchhiking event. The first and second terms represent the heterozygosities contributed by neutral mutants that appeared before and after the hitchhiking event, respectively. Assuming that the equilibrium level of genetic variation has been attained before the hitchhiking event and background selection does not change the shape of the distribution of mutant frequencies from that of neutrality, $\varphi(y)$ can be replaced by θ/y , where $\theta = 4N_2\mu$. Mean heterozygosity determined by $\varphi(y)$ then becomes reduced by h and decays further by $e^{-\tau/2N_2}$. The second term is derived similarly. Then, (12) reduces to

$$E[\pi] = \theta\{1 - (1 - h)e^{-\tau/2N_2}\}. \quad (13)$$

This equation describes the recovery of genetic variation as a function of time. It generalizes results obtained by previous studies of genetic hitchhiking (Wiehe and Stephan 1993; Perltitz and Stephan 1997) and population bottlenecks (Tajima 1989). In contrast to those studies that were derived under the assumption of neutrality, (13) takes background selection into account.

Simulation results: We introduced 2×10^7 M alleles independently in a diploid population of a size $2N = 10^5$, as described above, and observed its frequency at $T = 0$ (Figure 2 and Table 3). The frequency distribu-

Number of segregating alleles

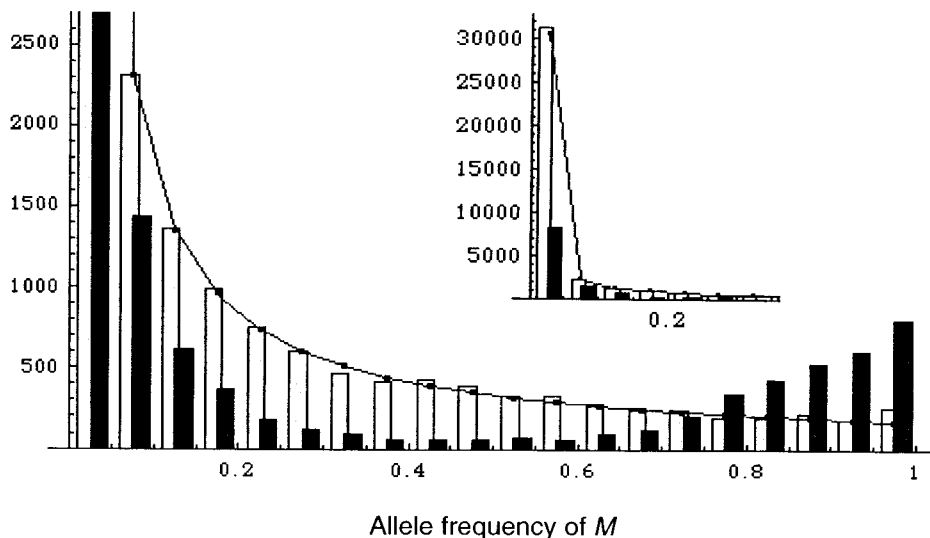


Figure 2.—Allele frequency distribution of M before and after a hitchhiking event. Shaded bars represent the frequency data obtained at $T = 0$ in simulation 3-2 of Table 3 (before hitchhiking). Small squares connected by lines show the expected number of M 's segregating in each frequency interval. The expected number of M 's segregating in the frequency interval $(y, y + dy)$ is assumed to be $\theta/ y dy$, where $\theta = 4N_2\mu$. μ is determined by simulation, as explained in the text. Then, the expected number of M 's in a frequency interval $(i/20, (i + 1)/20]$, $i = 0, \dots, 19$, is given by $\int_{0.05i/20}^{0.05(i+1)/20} \theta/ y dy$. Solid bars represent the data obtained in simulation 3-3 (immediately after hitchhiking).

tions of allele M at $T = 0$, before and after the hitchhiking event, are shown in Figure 2. Comparing the observed and expected distributions before hitchhiking, it is clearly seen that background selection does not change the shape of the frequency distribution of a linked neutral locus. Significant excess of low-frequency alleles was not observed. Immediately after the hitchhiking event, however, the number of intermediate-frequency alleles was greatly reduced, and that of high-frequency alleles increased significantly. This effect was previously observed by J. Fay and C.-I Wu (personal communication). We further discuss this observation below.

Table 3 summarizes the simulation results for hetero-

zygosity (π), observed immediately after the hitchhiking event (except 2-2). Predicted values (using Equation 13) agree well with the simulation results. Figure 3A also shows that (13) accurately predicts heterozygosities at various time points after the hitchhiking event. For $\tau = 1$, the average reduction of heterozygosity by a hitchhiking event, predicted by h (Equation 3), simply corresponds to π/θ . π/θ does not change significantly from 2-1 (0.48) to 2-3 (0.48) or from 2-4 (0.065) to 2-5 (0.067), which implies that the reduction of the effective population size at Fav by background selection does not weaken the effect of a single hitchhiking event, at least when $s = t$. This result differs from model 1, where the effect of hitchhiking decreased as the effect of back-

TABLE 3
Results for model 2 (gene order: *Del-Fav-Neu*)

	s	t	u/t	r_1	r_2	τ	Seg. M^a	$\pi(\times 10^{-5})^b$	$\theta(\times 10^{-5})$	p_a^c
2-1	0.02	—	0	—	10^{-3}	1	17,535	9.654 (9.982)	19.95	—
2-2	—	0.02	0.2	10^{-3}	10^{-3}	-1	41,476	16.94 (16.65)	16.65	—
2-3	0.02	0.02	0.2	10^{-3}	10^{-3}	1	14,455	7.939 (8.162)	16.65	0.200 ± 0.008
2-4	0.02	0.02	0.2	10^{-3}	10^{-4}	1	7,323	1.061 (1.078)	16.37	0.200 ± 0.008
2-5	0.02	0.02	0.4	10^{-3}	10^{-4}	1	5,535	0.858 (0.812)	12.78	0.404 ± 0.016
2-6	0.005	0.02	0.2	10^{-3}	10^{-4}	1	7,462	3.061 (3.183)	16.37	0.200 ± 0.007
2-7	0.02	0.005	0.2	10^{-3}	10^{-4}	1	8,211	1.167 (1.086)	17.27	0.228 ± 0.092
2-7a ^d	0.02	0.005	0.2	10^{-3}	10^{-4}	1	8,224	1.116 —	17.27	0.198 ± 0.055
2-7b ^d	0.02	0.005	0.2	10^{-3}	10^{-4}	1	8,137	1.217 —	17.27	0.379 ± 0.095
2-8	0.02	0.005	0.2	10^{-4}	10^{-4}	1	7,690	1.073 (1.070)	16.27	0.259 ± 0.156
2-8a ^d	0.02	0.005	0.2	10^{-4}	10^{-4}	1	7,655	1.060 —	16.27	0.202 ± 0.083
2-8b ^d	0.02	0.005	0.2	10^{-4}	10^{-4}	1	7,648	1.244 —	16.27	0.566 ± 0.064

Simulation results based on 2×10^7 introductions of M ; $2N = 10^5$, $L = 10^4$.

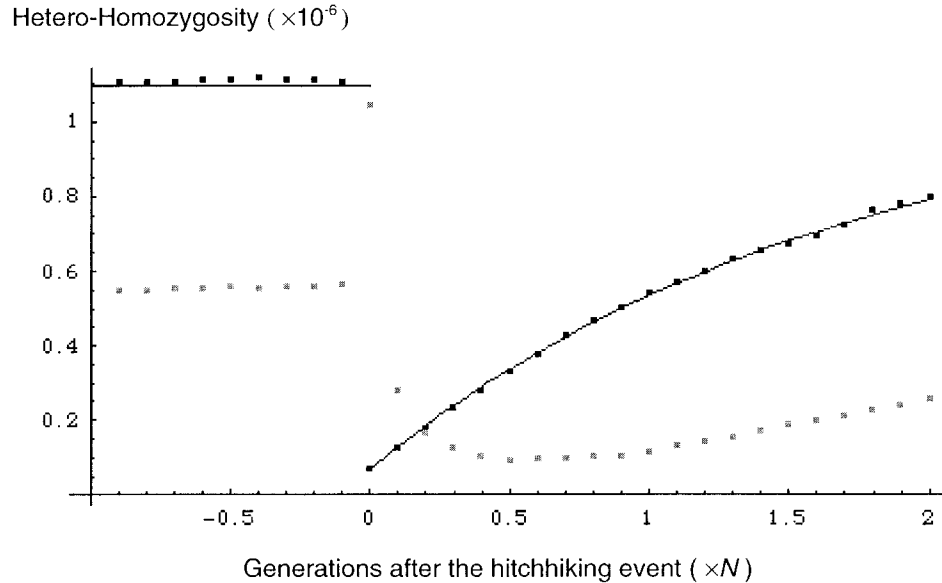
^a Number of segregating alleles recorded at $T = 0$.

^b Mean values of π at $T = 0$ and its predicted value (using Equation 13) in parentheses.

^c Mean \pm standard deviation of the frequency of allele a , recorded at the first generation when the frequency of B is >0.5 in each substitution of B .

^d The mutation to B occurs in initial linkage with either A or a , respectively.

A



B

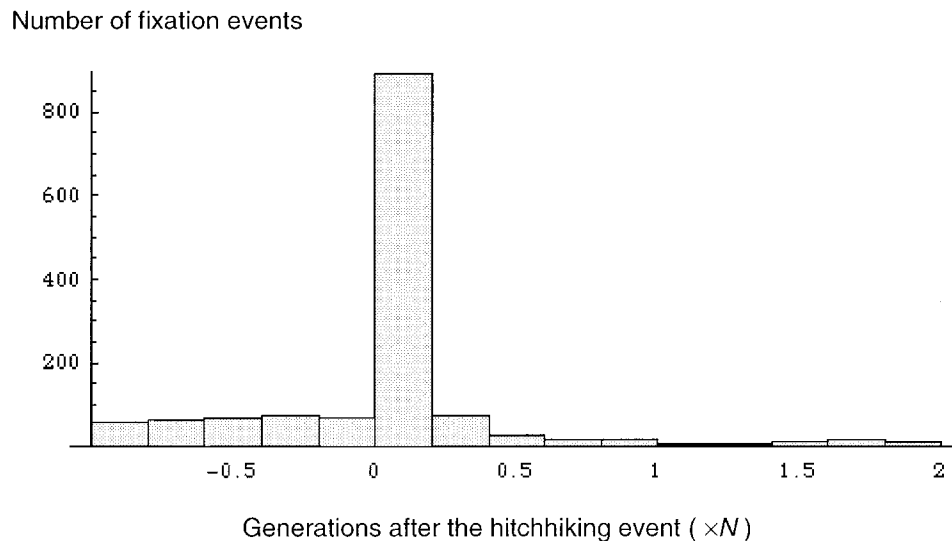


Figure 3.—Changes of heterozygosity, homozygosity, and fixation rates at *Neu* over time after a hitchhiking event. The parameters are $2N = 10^5$, $L = 15,000$, $\tau = 10,000$, $s = t = 0.02$, $u/t = 0.2$, $r_1 = 10^{-3}$, $r_2 = 10^{-4}$, and number of *M*'s introduced = 10^8 . (A) In addition to $T = 0$, the frequency of *M* was recorded every 500 generations. Mean heterozygosity and homozygosity were calculated at each generation. Solid squares represent observed heterozygosities. Lines were drawn for the expected values of heterozygosity using Equation 13. Shaded squares represent homozygosity. (B) Whenever the *M* allele is fixed, the time of this event was recorded. The histogram shows the number of fixation events at each time interval. The interval between 0 and $0.2N$ generations after the hitchhiking includes the fixation events occurring during the substitution of *B*.

ground selection increased. This discrepancy is caused by the fact that N_2 and Φ are significantly reduced by background selection but h is relatively insensitive to changes of $\alpha_1 = 2N_1s$ (discussed above).

We also investigated the perturbation of the mutation-selection balance at *Del* during the substitution of *B* at *Fav* and its effect on heterozygosity after the fixation of *B*. It was previously observed that the frequency of the deleterious allele, p_a , deviates most from its equilibrium value, u/t , when the increase of the frequency of *B*, p_B , is greatest. p_a returns toward u/t after p_B exceeds 0.5 (data not shown). We thus recorded p_a in the first generation after p_B became >0.5 during the substitution process of *B*. When $s \leq t$, p_a remained very close to u/t . When $s = 0.02$ and $t = 0.005$ (2-7 and 2-8), the mean

and standard deviation of p_a increased significantly, as expected. However, the observed and expected value of π still agreed very well. To further investigate the change of p_a , we conducted additional simulations in which *B* was introduced in initial linkage either with *A* or with *a* (2-7a and 2-7b and 2-8a and 2-8b). The initial linkage with *A* did not change p_a significantly. However, the linkage with *a* greatly elevated p_a . Surprisingly, mean heterozygosities after the hitchhiking event were relatively close to each other despite a large difference in p_a during the selective phase, although small increases in heterozygosity for the case of initial linkage with *a* were observed (2-7b and 2-8b). If the increase of the deleterious allele frequency caused the reduction of effective population size at linked loci, it would have

resulted in a lower heterozygosity after the fixation of *B*. However, the result is in the opposite direction. One might argue that the reduction of effective population size will weaken the strength of directional selection and thus result in higher values of π . However, it was shown above that a decrease of α_1 does not change h significantly (2-4 and 2-5). Therefore, the agreement of the observed value of heterozygosity with its prediction needs to be further explored, as the underlying assumption—constant p_a during the selective phase—is violated (see discussion).

Finally, we investigated the increase of homozygosity of derived neutral alleles after a hitchhiking event. We observed the level of homozygosity of *M* at various time points before and after the hitchhiking event (Figure 3A). Homozygosity increased sharply immediately after hitchhiking. However, it dropped quickly and, after a short time ($<0.5N$ generations), the homozygosity/heterozygosity ratio decreased below its standing level before the hitchhiking event. When we decreased the strength of the hitchhiking effect by increasing r_2 from 10^{-4} to 10^{-3} , the immediate increase of homozygosity was smaller than shown in Figure 3A, but the decrease of homozygosity over time was slower than in Figure 3A (data not shown). The rapid change of homozygosity over time implies that the high-frequency alleles produced by hitchhiking are quickly fixed in the population. We confirmed this by recording fixation events of allele *M* over time (Figure 3B). There was a great increase of fixation events during and shortly after the substitution of *B*. As hitchhiking events cannot change the average substitution rate of neutral alleles (Birky and Walsh 1988), the transient increase of the fixation rate should be followed by a period of a low fixation rate. Indeed, we observed a reduction of fixation events following the period of a high fixation rate (Figure 3B). The same pattern was observed when we replaced the hitchhiking event with a population bottleneck in the simulation (data not shown).

DISCUSSION

We demonstrated by simulations that formulas for background selection and hitchhiking can be combined to predict genetic variation at a linked neutral locus, despite the fact that these processes may interfere with one other. Analytic solutions previously known for the hitchhiking effect agreed well with our simulation results when effective population size and the fixation probability of the selected allele were modified by background selection. Two different simulation procedures were used in models 1 and 2. In these models, background selection occurs at one locus, as if deleterious mutations distributed over an entire chromosome were collapsed into a single locus. Therefore, the results obtained in this study might not be directly applicable to the realistic situation where background selection

results from deleterious mutations at many loci. However, since it was shown that the effects of deleterious mutations at two loci combine multiplicatively to reduce genetic variation (Hudson and Kaplan 1995; Nordborg *et al.* 1996) and the fixation probability of a favorable allele (Barton 1995), background selection at many loci is not likely to change the overall results of this study.

Our simulation results indicated that (5) and (13) are approximately correct even if the frequency, p_a , of the deleterious allele deviates from its equilibrium value due to strong directional selection at linked loci. In simulations 2-7b and 2-8b, we expected further reduction of heterozygosity at $T = 0$ because p_a increased significantly during the selective phase, which might mean a further reduction of effective population size. However, the following argument shows that an increase of p_a does not necessarily imply a decrease of effective population size. Hudson and Kaplan (1995) explained how background selection can reduce effective population size and thus the size of gene genealogies. In a population in mutation-selection balance, two ancestral genes can have a common ancestor in the previous generation only if two genes have the same number of deleterious alleles at linked loci. As time runs backward, ancestral genes are preferentially found in chromosomes with no deleterious alleles at linked loci because chromosomes carrying deleterious alleles have a small probability of having descendants. Therefore, the rate at which gene lineages coalesce increases as the number of chromosomes with no deleterious alleles decreases. However, if a favorable allele *B* that was initially linked with deleterious alleles goes to fixation (simulations 2-7b and 2-8b), some ancestral genes must be found on chromosomes with deleterious allele *a* during the selective phase. This is because all the descendants at *Neu* after fixation should be in linkage with allele *B*, which was initially in linkage with *a*. The association between *B* and *a* decays by recombination as time goes backward to the early stage of the selective phase. Therefore, the increase of p_a in the middle of the selective phases in 2-7b and 2-8b may not affect effective population size as the ancestral genes are found on chromosomes with *a* as well as those without *a*. Slight increases of heterozygosities in 2-7b and 2-8b indicate that this effect slightly increases, rather than decreases, effective population size.

The generalization of model 1 leading to (6) describes the overall relationship between recombination rate and genetic variation. Equation 6 can be used to estimate the parameters of background selection and/or hitchhiking in natural populations. The intensity of a selective sweep, αv ($\alpha = 2Ns$ and v is the number of strongly selected substitutions per nucleotide per generation), in *D. melanogaster* populations has previously been estimated without incorporating background selection (Wiehe and Stephan 1993; Stephan 1995).

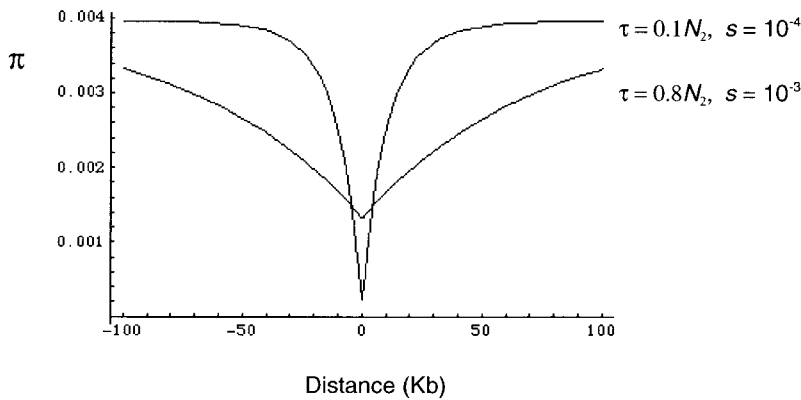


Figure 4.—Heterozygosities over a physical distance. Graphs were produced from Equation 13. Distance is defined to be zero at the location of *Fav*. Recombination rates are assumed to follow Haldane's map function with $\rho = 10^{-9}$. The effect of background selection is uniform over this region, with $N_2 = 10^6$ and $\mu = 10^{-9}$.

Therefore, this method was thought to have overestimated the effect of hitchhiking. However, (6) suggests that, in regions of very low recombination, the reduction of heterozygosity is mainly determined by hitchhiking unless the effect of background selection is extremely strong. This condition is likely to be met in *D. melanogaster* populations for the following reason. Charlesworth (1996) predicted a pattern of genetic variation across the *D. melanogaster* genome using a per-haploid genome mutation rate 0.48, which was obtained from the mutation accumulation studies by Mukai *et al.* (1974) and Ohnishi (1977). As a result, the expected level of heterozygosity was very close to the observed level of nucleotide diversity. However, recent surveys of the rates and effects of deleterious mutations in *D. melanogaster* suggest about fivefold lower values of the mutation rate (Keightley and Eyre-Walker 1999). If true, the expected level of heterozygosity explained by background selection should be significantly higher than that obtained by Charlesworth (1996) (see Equation 10), and the remaining reduction of heterozygosity should be explained by hitchhiking. As the value of $\alpha\nu$ was mainly determined by loci in regions of low recombination in which the relationship between nucleotide diversity and recombination is approximately linear, the estimation of Wiehe and Stephan (1993) and Stephan (1995) appears to be valid, but the interpretation of their results has to take into account that ν now depends on background selection (see Equation 6).

Our results can also be used to interpret recent observations of genetic variation on the *Y* chromosomes of *D. melanogaster* and *D. simulans*. Zurovcova and Eanes (1999) reported strongly reduced nucleotide diversity in the dynein gene *Dhc-Yh3*. Background selection is unlikely to explain this result since the *Y* chromosome encodes only six known genes and, therefore, background selection is probably not very strong. However, according to Equation 6, hitchhiking—even if it is very rare—is consistent with the observed extreme reduction of diversity on the nonrecombining *Y* chromosome.

Separation of the product $\alpha\nu$ into its components, *i.e.*, the rate and the strength of directional selection,

requires the measurement of the time between consecutive hitchhiking events or the selection coefficients of selected alleles. Equation 13 suggests that τ and s cannot be estimated separately from levels of nucleotide diversity, even if θ for the region is known. A possible solution is to find a local reduction of heterozygosity in a chromosomal region as a result of a single hitchhiking event. Figure 4 shows that combinations of τ and s produce unique patterns of expected heterozygosity over a physical distance. Therefore, a joint estimation of τ and s could be made by fitting (13) to multilocus polymorphism data in a chromosomal region. This approach will be useful in regions of high recombination where local reduction spans over a relatively short distance and the estimate of θ can be obtained from data from adjacent regions that are assumed to be close to the equilibrium level of heterozygosity.

We thank two reviewers and Bruce Walsh for valuable comments on the manuscript. This work was supported in part by National Science Foundation grant DEB-9896179 and by funds from the University of Rochester to W.S., and by an Ernst Caspari fellowship to Y.K.

LITERATURE CITED

- Barton, N. H., 1995 Linkage and the limits to natural selection. *Genetics* **140**: 821–884.
- Barton, N. H., 1998 The effect of hitch-hiking on neutral genealogies. *Genet. Res.* **72**: 123–133.
- Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.
- Benassi, V., F. DePaulis, G. K. Meghlaoui and M. Veuille, 1999 Partial sweeping of variation at the *Fbp2* locus in a west African population of *Drosophila melanogaster*. *Mol. Biol. Evol.* **16**: 347–353.
- Birky, C. W., and J. B. Walsh, 1988 Effects of linkage on rates of molecular evolution. *Proc. Natl. Acad. Sci. USA* **85**: 6414–6418.
- Charlesworth, B., 1996 Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet. Res.* **68**: 131–149.
- Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- Crow, J. F., and M. J. Simmons, 1983 The mutation load in *Drosophila*, pp. 1–35 in *The Genetics and Biology of Drosophila*, edited by M. Ashburner, H. L. Carson and J. N. Thompson. Academic Press, London.
- Gillespie, J. H., 1994 Alternatives to the neutral theory, pp. 1–17

- in *Non-neutral Evolution: Theories and Molecular Data*, edited by B. Golding, Chapman and Hall, London.
- Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. *Genetics* **141**: 1605–1617.
- Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The 'hitchhiking effect' revisited. *Genetics* **123**: 887–899.
- Keightley, P. D., and A. Eyre-Walker, 1999 Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* **153**: 515–523.
- Kimura, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**: 893–903.
- Kimura, M., 1971 Theoretical foundation of population genetics at the molecular level. *Theor. Popul. Biol.* **2**: 174–208.
- Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, United Kingdom.
- Maynard Smith, J., and J. Haigh, 1974 The hitchhiking effect of a favourable gene. *Genet. Res.* **23**: 23–35.
- Mukai, T., R. K. Cardellino, T. K. Watanabe and J. F. Crow, 1974 The genetic variance for viability and its components in a population of *Drosophila melanogaster*. *Genetics* **78**: 1195–1208.
- Nordborg, M., B. Charlesworth and D. Charlesworth, 1996 The effect of recombination on background selection. *Genet. Res.* **67**: 159–174.
- Nurminsky, D. I., M. V. Nurminskaya, D. De Aguiar and D. L. Hartl, 1998 Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* **396**: 572–575.
- Ohnishi, O., 1977 Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster* II. Homozygous effects of polygenic mutations. *Genetics* **87**: 529–545.
- Ohta, T., and M. Kimura, 1975 The effect of a selected linked locus on heterozygosity of neutral alleles (the hitchhiking effect). *Genet. Res.* **25**: 313–326.
- Peck, J., 1994 A ruby in the rubbish: beneficial mutations, deleterious mutations, and the evolution of sex. *Genetics* **137**: 597–606.
- Perlitz, M., and W. Stephan, 1997 The mean and variance of the number of segregating sites since the last hitchhiking event. *J. Math. Biol.* **36**: 1–23.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, 1992 *Numerical Recipes in C*. Cambridge University Press, Cambridge, United Kingdom.
- Schlötterer, C., C. Vogl and D. Tautz, 1997 Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics* **146**: 309–320.
- Stephan, W., 1995 An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorphism data. *Mol. Biol. Evol.* **12**: 959–962.
- 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., L. Xing, D. A. Kirby and J. M. Braverman, 1998 A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proc. Natl. Acad. Sci. USA* **95**: 5649–5654.
- Stephan, W., B. Charlesworth and G. McVean, 1999 The effect of background selection at a single locus on weakly selected, partially linked variants. *Genet. Res.* **73**: 133–146.
- Tajima, F., 1989 The effect of change in population size on DNA polymorphism. *Genetics* **123**: 597–601.
- Wiehe, T. H. E., and W. Stephan, 1993 Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. *Mol. Biol. Evol.* **10**: 842–854.
- Zurovcova, M., and W. F. Eanes, 1999 Lack of nucleotide polymorphism in the Y-linked sperm flagellar dynein gene *Dhc-Yh3* of *Drosophila melanogaster* and *D. simulans*. *Genetics* **153**: 1709–1715.

Communicating editor: J. B. Walsh

APPENDIX A

Eight haplotype frequencies (x_1, x_2, \dots, x_8) representing *ABM, ABm, AbM, Abm, aBM, aBm, abM, and abm*,

respectively, change deterministically by selection, recombination, and deleterious mutations to (x'_1, x'_2, \dots, x'_8). Equations for recombination are easily derived from a table of random matings where double recombination events are ignored.

1. Selection:

$$\begin{aligned}\bar{w} &= \sum_i \sum_j w_{ij} x_i x_j, \\ \bar{w}_i &= \sum_j w_{ij} x_j, \\ y_i &= x_i \bar{w}_i / \bar{w} \quad (i, j = 1, \dots, 8),\end{aligned}$$

where w_{ij} is the fitness of an individual with haplotypes i and j . For example, $w_{15} = (1 - \delta)(1 + s)^2$.

2. Recombination:

$$\begin{aligned}z_1 &= y_1 - r_1 D_1 - r_2 D_5, & z_5 &= y_5 + r_1 D_1 - r_2 D_7, \\ z_2 &= y_2 - r_1 D_2 + r_2 D_5, & z_6 &= y_6 + r_1 D_2 + r_2 D_7, \\ z_3 &= y_3 - r_1 D_3 - r_2 D_6, & z_7 &= y_7 + r_1 D_3 - r_2 D_8, \\ z_4 &= y_4 - r_1 D_4 + r_2 D_6, & z_8 &= y_8 + r_1 D_4 + r_2 D_8,\end{aligned}$$

where

$$\begin{aligned}D_1 &= y_1(y_6 + y_7 + y_8) - y_5(y_2 + y_3 + y_4), \\ D_5 &= y_1(y_4 + y_6 + y_8) - y_2(y_3 + y_5 + y_7), \\ D_2 &= y_2(y_5 + y_7 + y_8) - y_6(y_1 + y_3 + y_4), \\ D_6 &= y_3(y_2 + y_6 + y_8) - y_4(y_1 + y_5 + y_7), \\ D_3 &= y_3(y_5 + y_6 + y_8) - y_7(y_1 + y_2 + y_4), \\ D_7 &= y_5(y_2 + y_4 + y_8) - y_6(y_1 + y_3 + y_7), \\ D_4 &= y_4(y_5 + y_6 + y_7) - y_8(y_1 + y_2 + y_3), \\ D_8 &= y_7(y_2 + y_4 + y_6) - y_8(y_1 + y_3 + y_5).\end{aligned}$$

3. Mutation:

$$\begin{aligned}x'_1 &= (1 - u)z_1, & x'_5 &= z_5 + uz_1, \\ x'_2 &= (1 - u)z_2, & x'_6 &= z_6 + uz_2, \\ x'_3 &= (1 - u)z_3, & x'_7 &= z_7 + uz_3, \\ x'_4 &= (1 - u)z_4, & x'_8 &= z_8 + uz_4.\end{aligned}$$

APPENDIX B

In the simulation of model 2, one mutant, M , is introduced at $T = L$ with probability δ or at $T = \lambda$ with probability $1 - \delta$, as described above. Its contribution to heterozygosity at $T = 0$ is $2y(1 - y)$, where y is the frequency of M at $T = 0$. The simulation measures the average value of this contribution by introducing many M 's independently. Mean heterozygosity contributed by one M , π^* , is predicted to be

$$E[\pi^*] = \delta \sum_{i=1}^{2N-1} \frac{1}{i} C\left[\frac{i}{2N}, L\right] / \left(\sum_{i=1}^{2N-1} \frac{1}{i}\right) + (1 - \delta) \sum_{j=0}^{L-1} \frac{1}{L} C\left[\frac{1}{2N}, j\right], \quad (\text{B1})$$

where $C[p, t]$ denotes the expected heterozygosity at $T = 0$ contributed by a mutant whose frequency is p at $T = t$. At neutral equilibrium, expected heterozygosity decays as a function of time and effective population size; *i.e.*, $C[p, t] = 2p(1 - p)\exp[-t/2N]$ (Crow and Kimura 1970). Incorporating these expressions in (B1) and changing the summation to integration, we obtain

$$E[\pi^*] \approx \frac{\delta}{a_{2N}} \int_0^1 2y(1 - y) e^{-L/2N} dy + \frac{1 - \delta}{L} \int_0^L 2 \frac{1}{2N} \left(1 - \frac{1}{2N}\right) e^{-t/2N} dt, \\ \approx \frac{\delta}{a_{2N}} e^{-L/2N} + \frac{2(1 - \delta)}{L} (1 - e^{-L/2N}), \\ = 4N\mu = \theta, \quad (\text{B2})$$

as $\mu = (1 - \delta)/2NL = 1/(4Na_{2N} + 2NL)$.