Genet. Res., Camb. (1996), **67**, *pp.* 159–174 *With* 3 *text-figures* Copyright © 1996 Cambridge University Press

The effect of recombination on background selection*

MAGNUS NORDBORG†, BRIAN CHARLESWORTH
AND DEBORAH CHARLESWORTH
Department of Ecology & Evolution, The University of Chicago, 1101 E. 57th St, Chicago, IL 60637-1573, USA. Tel: (312) 702-1040, Fax: (312) 702-9740, E-mail: magnus@darwin.uchicago.edu

(*Received 3 July 1995 and in revised form 20 October 1995*)

Summary

An approximate equation is derived, which predicts the effect on variability at a neutral locus of background selection due to a set of partly linked deleterious mutations. Random mating, multiplicative fitnesses, and sufficiently large population size that the selected loci are in mutation/selection equilibrium are assumed. Given these assumptions, the equation is valid for an arbitrary genetic map, and for an arbitrary distribution of selection coefficients across loci. Monte Carlo computer simulations show that the formula performs well for small population sizes under a wide range of conditions, and even seems to apply when there are epistatic fitness interactions among the selected loci. Failure occurred only with very weak selection and tight linkage. The formula is shown to imply that weakly selected mutations are more likely than strongly selected mutations to produce regional patterning of variability along a chromosome in response to local variation in recombination rates. Loci at the extreme tip of a chromosome experience a smaller effect of background selection than loci closer to the centre. It is shown that background selection can produce a considerable overall reduction in variation in organisms with small numbers of chromosomes and short maps, such as *Drosophila*. Large overall effects are less likely in species with higher levels of genetic recombination, such as mammals, although local reductions in regions of reduced recombination might be detectable.

1. Introduction

The possible effects of natural selection on variation and evolution at neutral or weakly selected linked loci have recently attracted a good deal of attention (Birky & Walsh, 1988; Kaplan *et al.* 1989; Stephan *et al.* 1992; Charlesworth *et al.* 1993; Wiehe & Stephan, 1993; Barton, 1994, 1995; Charlesworth, 1994; Gillespie, 1994; Hudson, 1994; Hudson & Kaplan, 1994, 1995; Braverman *et al.* 1995; Charlesworth *et al.* 1995; Simonsen *et al.* 1995), 20 yr after the pioneering work on this subject (Maynard Smith & Haigh, 1974; Ohta & Kimura, 1975; Thomson, 1977). In addition, the effect of artificial selection in reducing the effective population size at neutral loci has been studied theoretically (Santiago & Caballero, 1995; Santiago, in preparation). This work has in part been stimulated by data from surveys of DNA variation in natural populations of *Drosophila*, which consistently indicate that genetic variability is lower for loci in regions where genetic recombination is relatively infrequent, compared with regions in which it occurs at higher frequencies (Aquadro *et al.* 1994; Kreitman & Wayne, 1994). In addition, codon bias in *D. melanogaster* appears to be lower in regions of reduced recombination, suggesting that selection at weakly selected sites is less effective when recombination is infrequent (Kliman & Hey, 1993).

Two main hypotheses have been proposed to explain these observations. The 'selective sweep' model (Berry *et al.* 1991) appeals to hitch-hiking of neutral (or weakly selected) variants by favourable mutations that arise at closely linked loci, and which cause a substantial loss in variation as a result of the fixation of surrounding chromosomal regions (Maynard Smith & Haigh, 1974; Kaplan *et al.* 1989; Stephan *et al.* 1992; Wiehe & Stephan, 1993; Barton, 1994, 1995; Braverman *et al.* 1995; Simonsen *et al.*

* This paper is dedicated to Richard Lewontin on the occasion of his 65th birthday.

† Corresponding author.

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

1995; Stephan, 1995). The 'background selection' model involves the loss of neutral or nearly neutral variants as a result of elimination of linked deleterious mutant alleles from the population (Charlesworth *et al.* 1993; Charlesworth, 1994; Hudson, 1994; Hudson & Kaplan, 1994, 1995; Barton, 1995; Charlesworth *et al.* 1995).

To evaluate the relative contributions of these processes to the patterns discerned in the data, it is necessary to develop theoretical models which are sufficiently realistic to allow quantitative predictions about the relations between chromosomal location (i.e. recombinational environment) and genetic variability and rate of evolution at neutral or weakly selected loci. Since genetic recombination occurs at a low rate even in regions of the *Drosophila* genome where its frequency is reduced, models which include the effects of genetic recombination are essential for this purpose. Although useful results can be obtained by computer simulation (Charlesworth *et al.* 1993, 1995; Braverman *et al.* 1995; Hudson & Kaplan, 1995; Simonsen *et al.* 1995), it is clearly valuable to have analytic results, provided that sufficiently accurate approximations to reality can be achieved. Several such models have been developed for the analysis of selective sweeps (Maynard Smith & Haigh, 1974; Kaplan *et al.* 1989; Stephan *et al.* 1992; Wiehe & Stephan, 1993; Stephan, 1995). With respect to background selection, Hudson & Kaplan (1994) have obtained a formula for the reduction in genetic diversity at a neutral locus, taking into account selection at a single partly linked locus subject to recurrent deleterious mutation. Barton (1995) derived a similar result for the fixation probability of a favourable allele. Hudson & Kaplan (1995) derived predictions of the effect of mutation at multiple loci subject to selection on variability at a linked neutral locus, using the simplifying assumptions of one crossover per chromosome, and equal selective effects at each locus. They used their results to predict the pattern of genetic variability as a function of chromosomal location in the genome of *D. melanogaster*.

In this paper, we present an alternative derivation of the effect on variability at a neutral site due to a linked locus subject to deleterious mutations, and show how an arbitrary number of selected loci can be treated without the restrictive assumptions made by Hudson & Kaplan (1995). We also investigate the conditions required for the analytical results to hold, and relate our results to those on the reduction in effective population size due to selection, derived by Santiago & Caballero (1995). We use computer simulations to evaluate the performance of the formulae under a wide range of parameter values. These results have been used to predict patterns of variability in *Drosophila* under less restrictive assumptions than those made by Hudson & Kaplan (1995) (B. Charlesworth, submitted).

2. Analytical results

(i) *Formulation of the model*

The model assumes m autosomal diallelic loci subject to mutation–selection balance, in a randomly mating population. The wild-type allele at the ith locus is denoted by A_i and the mutant allele by a_i; their frequencies are p_i and $q_i = 1-p_i$, respectively. The mutant alleles are assumed to be so rare that terms of order q_i^2 are negligible. Selection can thus be assumed to take place exclusively against heterozygous carriers of mutant alleles. Let the mutation rate from the wild-type to the mutant allele at the ith locus be u_i. The mean number of new mutations per diploid individual is $U = 2\sum_i u_i$. Let the fitness of mutant heterozygotes relative to that of wild-type homozygotes be $1-t_i$ (t is the product of the selection coefficient for mutant homozygotes, s, and the dominance coefficient h). If the fitness effects of different loci are multiplicative, as will be assumed in this section, there is no linkage disequilibrium in an infinite population at equilibrium under mutation and selection (Felsenstein, 1965; Feldman *et al.* 1980; Charlesworth, 1990), and the equilibrium allele frequency at the ith locus is approximately u_i/t_i, provided that $u_i \ll t_i$ (Crow & Kimura 1970, ch. 6). The population size is assumed to be so large that these conditions are satisfied, to a good approximation.

(ii) *The effect of a single selected locus on a linked neutral locus*

We first consider the effect of selection at a single locus, locus i, for which the frequency of recombination with the neutral locus is r_i. If there are sex differences in recombination frequencies, r_i is the appropriate average over the sexes (Crow & Kimura, 1970, p. 50). Let the neutral locus have alleles B and b, and let the frequencies of these alleles within mutant-free gametes with respect to the selected locus be x_0 and $1-x_0$ respectively. Similarly, let the frequency of B in gametes carrying a mutation at locus i be x_1, and write $\delta = x_0 - x_1$. The genetic diversity at locus B, as measured by the probability that two randomly chosen gametes differ in allelic state (Nei, 1987, ch. 8), is

$$G = 2p_i^2 x_0(1-x_0) + 2p_i q_i[x_0(1-x_1) + x_1(1-x_0)] + 2q_i^2 x_0(1-x_0). \quad (1)$$

Rearranging, and neglecting terms of order q_i^2, we obtain

$$G \approx 2x_0(1-x_0) - 2q_i(1-2x_0)\,\delta. \quad (2)$$

It is assumed that neutral variation is produced according to the infinite sites model, with a mutation rate v per site (Kimura, 1969). The neutral locus in the above formulation corresponds to a single such site,

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

segregating for a pair of alleles. Allele B is arbitrarily chosen as the allele which originates by mutation. The expected nucleotide site diversity at statistical equilibrium, π, is thus given by the expectation of G. Knowing the mean and variance terms for the frequencies of B in the different gamete classes, it is straightforward (see Appendix) to apply the linear differential operator method of Ohta & Kimura (1969) in order to obtain an expression for π.

The final expression for the ratio of π to the classical neutral value, π_0, is found to be

$$\frac{\pi}{\pi_0} \approx 1 - \frac{q_i}{(1+\rho_i)^2}, \tag{3}$$

where $\rho_i = \tilde{r}_i/t_i$ and $\tilde{r}_i = r_i(1-t_i)$. This formula is equivalent to equation (3) of Hudson & Kaplan (1995), with the minor difference that their formula involves r_i not $\tilde{r}_i$. It is also similar to the formula derived by Barton (1995) for the probability of fixation of a favourable mutation linked to a locus subject to mutation and selection. It is easily seen that, to the order of the approximations used here, equation (3) with no recombination is equivalent to the expression $\pi = f_0 \pi_0$ (where $f_0 = 1 - q_i$ in this case) derived previously for the case of an arbitrary number of loci (Charlesworth *et al.* 1993).

Some modification is needed for the case of X-linked mutations, which are selected against in the hemizygous state in males (assuming male heterogamety). The largest difference from the autosomal case is likely to be when all deleterious mutations act in both males and females. The equilibrium frequency of a mutant allele at locus i is then $q_i = u_i/\tilde{t}_i$, where $\tilde{t}_i = (2t_i + s_i)/3$, and s_i is its homozygous or hemizygous effect on fitness (Haldane, 1927). This can be substituted into equation (3), replacing t_i by $\tilde{t}_i$. If selection acts only in males, $\tilde{t}_i = s_i/3$; if its acts only on females, $\tilde{t}_i = 2t_i/3$. The averaging of recombination rates across the sexes must be done differently from the autosomal case, since an X-linked gene spends two-thirds of its time in the homogametic sex, compared with one-half for autosomes. The effective population size for the classical neutral case must also be adjusted appropriately (Charlesworth 1994; Caballero, 1995; Nagylaki, 1995).

(iii) *The effects of multiple loci*

The above results can be generalized (see Appendix) to incorporate the effects of multiple selected loci. With small effects of each locus, we have

$$\frac{\pi}{\pi_0} \approx \exp - \sum_i \frac{q_i}{(1+\rho_i)^2}. \tag{4}$$

This is similar to equation (5) of Hudson & Kaplan (1995), which was derived on the assumption of equal selective effects of each locus, and a single chromosome with at most one crossover per chromosome. As with their formula, the summation can be replaced by integration over the genomic region in question, if selected loci are sufficiently densely packed into chromosomes that a chromosome can be treated as a continuum.

We can then assume that loci are distributed uniformly along the physical map of the genomic region in question, and denote by a variable z the physical position of a locus subjects to mutation and selection. The frequency of recombination between a selected locus at position z and the neutral site under consideration is denoted by $r(z)$. Let u be the frequency of new mutations per unit physical distance for a haploid genome. The joint probability density of u and the selection coefficient t (i.e. the 'mutation spectrum') is $\phi(u, t)$. This allows for the possibility that there may be different rates of mutation to alleles with different effects on fitness (Crow & Simmons, 1983; Keightley, 1994). We assume that ϕ is independent of location. If the size of the region in terms of physical position in some unit of measurement (e.g. kilobases) is R, the densities of loci and of new mutations per physical distance unit are m/R and $U/2R$, respectively. Equation (4) can then be replaced by

$$\frac{\pi}{\pi_0} \approx \exp - \int_R \int_0^1 \int_0^1 \frac{u\phi(u,t)}{t[1+\rho(t,z)]^2} \, dz \, dt \, du, \tag{5}$$

where $\rho(t, z) = r(z)(1-t)/t$.

A further simplification is possible when all loci have the same mutation density u and selection coefficient t, such that $u = U/2R$. We then have

$$\frac{\pi}{\pi_0} \approx \exp - \frac{U}{2Rt} \int_R \frac{dz}{[1+\rho(z)]^2}, \tag{6}$$

where $\rho(z) = r(z)(1-t)/t$. Consider the case of a single chromosome. If a proportion P of the region is located to the left of the neutral locus and a proportion $Q = 1 - P$ is to the right, equation (6) becomes

$$\frac{\pi}{\pi_0} \approx \exp - \frac{U}{2Rt} \left\{ \int_0^{PR} \frac{dz}{[1+\rho(z)]^2} + \int_0^{QR} \frac{dz}{[1+\rho(z)]^2} \right\}. \tag{7}$$

If the map distance (in Morgans) M over the region R is short enough for double crossovers to be ignored, we have a linear relationship between distance z and recombination rate,

$$r(z) = z, \tag{8}$$

as assumed by Hudson & Kaplan (1995). The exponent in equation (7) then becomes

$$E = \frac{U[t + 2PQM(1-t)]}{2[t + QM(1-t)][t + PM(1-t)]}. \tag{9}$$

For a very short map ($M \ll t$), this gives $E \approx U/2t$, as expected from the general result for no recombination (Charlesworth *et al.* 1993). For a map that is substantially longer than the selection coefficient

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

($M \gg t$), and for loci which are not close to the ends (so that $PQM \gg t$), we have

$$E \approx \frac{U}{M(1-t)}, \tag{10}$$

i.e. the proportional effect of background selection is approximately the same as the density of new mutations per map unit, as pointed out previously by Hudson & Kaplan (1994, 1995) and Barton (1995) for the case of a neutral locus located in the centre of a block of selected loci. Remarkably, the effect of background selection with $M \gg t$ is virtually independent of the position of the neutral locus, except at the extreme ends of the chromosome, so that this result holds for almost any chromosomal region where the recombination rate per nucleotide is approximately independent of position. There is, however, an edge effect, as may be seen by comparing the value of E for the case of a neutral locus at the extreme left end of the chromosome ($PQ = 0$, $E = U/\{2[t+M(1-t)]\}$) with the value for a locus in the middle ($P = Q = 1/2$, $E \approx U/[M-(1-t)]$). This implies that there is a weaker effect of background selection on a neutral site located at the end of a chromosome than on one in the middle, as would be expected from the fact that it experiences the effects of selection only from one side.

(iv) *The effect of the strength of selection*

It is useful to look more closely at the qualitative behaviour of the results derived in the previous sections. Equation (3), the background selection effect due to a single selected locus, can be further approximated by

$$\frac{\pi}{\pi_0} \approx 1 - \frac{u_i}{t_i(1+r_i/t_i)^2}, \tag{11}$$

using $q_i \approx u_i/t_i$ and assuming small t_i. As is intuitively expected, diversity decreases with lower r_i and higher u_i. The effect of the selection coefficient is slightly more complicated. Differentiating (11) with respect to t_i, we see that the effect on diversity increases as t_i decreases, up to a maximum at $t_i = r_i$, and decreases thereafter. In other words, a weakly selected locus can cause strong background selection if it is tightly linked, but its importance declines rapidly with increasing r_i, whereas a strongly selected locus causes weaker background selection (i.e. greater π/π_0), but can do so from a greater distance.

It is thus clear that only mutations within a certain distance will matter when we sum over all loci to obtain the total effect. From the above discussion, it is also clear that this distance depends on t. For weakly selected mutations, the total effect comes from summing over a few closely linked loci, each causing strong background selection, whereas for strongly selected loci the total effect comes from summing over a larger fraction of the chromosome, each locus causing relatively weak background selection. If there is a distribution of mutation effects, weakly selected mutations may cause strong background selection in regions of reduced recombination but be unimportant in other parts of the genome. Also, since the region important for background selection is smaller when mutations are weaker, the resulting effect is expected to follow any variation in recombination rate more closely if the mutations are weak rather than strong. This is illustrated in Fig. 1. The expected values of π/π_0 from equation (7) are shown in Fig. 1*b* for two selection coefficients, for a chromosome with an approximate map length of one Morgan with reduced recombination in the centre (Fig. 1*a*). Fig. 1 also illustrates the edge effect referred to in the previous section. Near the tips of the chromosome, π/π_0 increases sharply. Notice that the region affected is much smaller under weak than under strong selection, in accordance with the argument just given: when selection is weak, only closely linked loci matter.

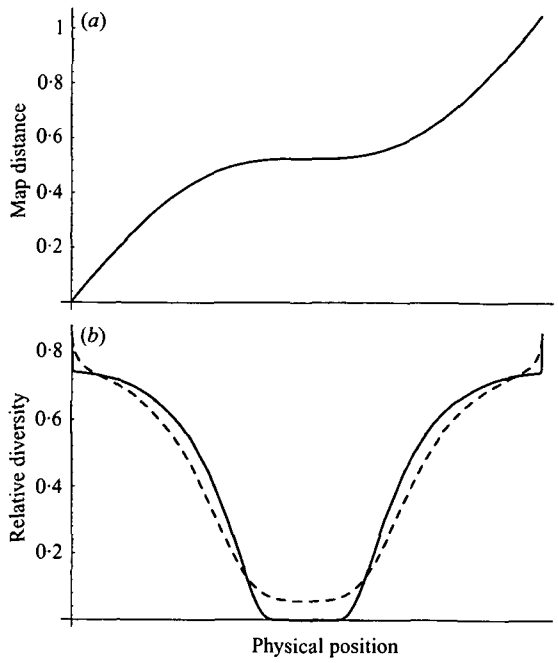

Fig. 1. The effect of background selection on a chromosome of approximate map length 1 Morgan with reduced recombination in the centre. Plot (*a*) shows the relationship assumed between physical distance and map distance [$r(z) = (z+\sin z)/6$]. Plot (*b*) shows π/π_0 with $U = 0{\cdot}1$ for weak (——) ($t = 0{\cdot}0003$) and strong (– – –) ($t = 0{\cdot}03$) selection. Strong selection gives a smoother relationship between the recombination rate and the effect of background selection. Note also that whether strongly or weakly selected mutations result in stronger background selection depends on the position on the map.

3. Simulation methods

The approximations derived in Section 2 were applied to results from Monte Carlo simulations of small populations. The program was designed to follow

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

closely the Wright–Fisher sampling model (e.g. Ewens, 1979, pp. 16–19). Populations of N diploid individuals were simulated, following the general method of Felsenstein & Yokoyama (1976) and Charlesworth *et al.* (1993). Pseudo-random numbers were generated using ACG from the GNU C++Library (Free Software Foundation, 1992). Details of the program and source code can be obtained from the authors.

Each chromosome in a haploid genome was assumed to have a large number of loci subject to deleterious mutations (assumed to be the same for all chromosomes when genomes with more than one chromosome were modelled). The exact number of loci is not important as long as it is large enough (compared to the mutation rate and the expected survival time of a deleterious mutation) for the probability of double mutations (i.e. a gamete carrying a mutant allele at a certain locus experiencing another mutation at this locus) to be negligible. When double mutations did occur, the resulting alleles were given a selective disadvantage equal to the sum of those of the two mutations. This occurred too rarely to be important. If a mutation became fixed (again, this happened with extremely low probability), it was treated as a new wild-type allele.

Fitnesses were calculated according to the fitness function in use (see below). These values were then divided by the greatest individual fitness present (almost always 1). An individual was then drawn at random from the population, and used as the first parent if its fitness was greater than a uniform variate in the interval [0, 1). If not, then a new individual was drawn (possibly the same), and the process was repeated. Using the parent, a gamete was formed as follows. A number of recombination events per chromosome was generated as a Poisson variate with mean equal to the total map length of the chromosome, and positions uniformly distributed over the chromosome (i.e. no interference). Finally, mutations were added to the new gamete, the number being determined by a Poisson variate with mean equal to the haploid mutation rate, and the positions again being uniformly distributed. A second new gamete was created by the same method (and could originate from the same parent as the first one). This procedure was repeated until N new individuals had been created.

Several different mutation schemes were used. In the simplest case, all mutations were identical, but the program could also model weaker selection on every nth site, or could pick the selective effects from a gamma distribution. The fitness of an individual with n heterozygous mutations, w_n, was calculated by one of three methods, namely:

(i) *Multiplicative*. Fitness was given by the product of the single-locus fitnesses. For instance, in the case of n identical mutations with heterozygous selection coefficient t, $w_n = (1-t)^n$.

(ii) *Additive*. Fitness was equal to one minus the sum of the single-locus selection coefficients, or zero, if the sum is negative. For instance, in the case of n identical mutations with heterozygous selection coefficient t, $w_n = \max(1-nt, 0)$.

(iii) *Synergistic epistasis*. Each additional mutation had a progressively larger effect, following the exponential quadratic model described by Charlesworth (1990). This model is defined only for mutations of equal effect, for which $w_n = \exp -n(\alpha + n\beta/2)$. The fitness of an individual with a mixture of heterozygous and homozygous mutations was determined by the method described by Charlesworth *et al.* (1991).

To monitor the effects of the selective background on neutral loci, the method of Charlesworth *et al.* (1993) was used. The neutral mutation process in itself was not modelled, but neutral alleles were introduced, one at a time, and observed until either fixed or lost. The number of generations until either fixation or loss, T, was recorded, as was the quantity

$$H = 2\sum_{i=1}^{T} x_i(1-x_i), \tag{12}$$

where x_i is the frequency of the allele. The genetic diversity under background selection, relative to the classical neutral value, was calculated as $\pi/\pi_0 = H/2$ (Charlesworth *et al.* 1993).

A number of neutral loci were placed at evenly spaced intervals from the centre of each chromosome, and the simulation was run until a given number of alleles (usually 16000) had been fixed or lost at each site. When more than a single neutral allele was to be introduced in the population, they were introduced in different individuals. The purpose of this was to speed up the simulations by allowing several simultaneous observations of different loci. Although the values of H observed for different sites are not independent, the mean properties we are measuring should not be affected. However, as is shown in Section 4(vii), the position of a neutral site can sometimes influence its distribution, and the simulation method used provides a convenient way of investigating this. Thus, when results from neutral sites are pooled in what follows, only sites deemed to be free of such effects were used. Confidence intervals for π/π_0 were calculated assuming that the means of the $H/2$ values over replicates were approximately normally distributed. Although the individual values of $H/2$ are far from normally distributed, this procedure is justified by the very large sample sizes.

Except where otherwise noted, the predicted means were calculated from equation (7), with $P = Q = 1/2$. We used two extremes for $r(z)$, either the linear one [equation (8)], or Haldane's mapping function (Haldane, 1919),

$$r(z) = \frac{1-e^{-2z}}{2}, \tag{13}$$

which, like the simulations, assumes no interference. In addition to making it possible to evaluate the

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

integral (7) explicitly, these two functions serve as useful limits. Because the map distance increases fast with physical distance when there are no double crossovers, π/π_0 is always smaller when (13) is used than with (8). However, as most of the effect is contributed by closely linked loci [cf. Section 2(iv)], the precise choice of mapping function rarely matters greatly. Other mapping functions, which allow for some degree of positive interference among crossovers (e.g. McPeek & Speed, 1995), will give results intermediate between the above.

Except where noted, we used a population size of $N = 3200$, and a single chromosome. As in previous studies (Charlesworth *et al.* 1993; Hudson & Kaplan, 1995), approximations for the effect of background selection derived under the assumption of large population size were found to work remarkably well for small N (see Section 5). The number of neutral loci per chromosome varied, as did the required number of observations (introductions followed by loss or fixation) per locus, but all simulations recorded at least a total of one and a half million such observations. It was verified that the frequency of fixations was not significantly different from the expected $1/2N$. We assumed a dominance coefficient of $h = 0{\cdot}2$, so that only values of t are given below. The main variables investigated in the simulations were the map length, M, the diploid mutation rate, U, and the selection scheme (the distribution of effects of mutations and their interaction). The parameter values were chosen partly to test the model with simulations using parameters comparable with previous work (Charlesworth *et al.* 1993, 1995; Hudson & Kaplan, 1994, 1995), and partly because they (barring N) lie within biologically reasonable bounds.

4. Simulation results

(i) *Identical mutations, multiplicative selection, short maps*

The approximate equation (7) for the effect of background selection in a region with recombination was derived assuming multiplicative interactions between loci of equal effect, and should therefore accurately predict the outcome of simulations incorporating this assumption. Cases 1–3 of Table 1 show that this expectation is fulfilled, regardless of the mapping function used for the theoretical expectations. The results agree with those obtained by Hudson & Kaplan (1995), who simulated the same cases, allowing only a single crossover per chromosome.

(ii) *Effect of map length*

The choice of mapping function is expected to be most important for longer maps. Cases 4 and 5 of Table 1 include the results of two simulations with long maps ($M = 1$), the first with a low mutation rate ($U = 0{\cdot}1$), and the second with a high mutation rate ($U = 0{\cdot}4$). In the latter case, when background selection is strong, the prediction using Haldane's mapping function is significantly better than that obtained using equation (8), but even here the difference is not great, illustrating the conclusion already noted that, because most of the background selection effect is contributed by closely linked loci, the recombination function does not matter much.

The most important effect of map length, however, is that, *ceteris paribus*, the shorter the map, the stronger the background selection. This is illustrated by cases 5–8 in Table 1 where we decrease the map length from $M = 1$ to $M = 0{\cdot}1$, while keeping everything else constant (see also Fig. 2).

(iii) *Weak selection*

The analysis in Section 2 assumes that the sojourn time of a deleterious allele is so short that a linked neutral allele contributes nothing to nucleotide diversity (see Section 5). If selection is weak, this latter assumption will not hold, and the approximations should overestimate the effect of background selection. Cases 9–12 of Table 1 show the results of simulations with weak selection. When a strong effect of background selection is predicted (cases 11–12), the predictions consistently overestimate the effect.

(iv) *Mutations with unequal effects*

When the selective effects varied across loci, the expected effect can be calculated using the relevant form of equation (5), namely,

$$\frac{\pi}{\pi_0} \approx \exp - \frac{U}{2R} \int \frac{1}{t} \times \left\{ \int_0^{PR} \frac{\mathrm{d}z}{[1+\rho(z)]^2} + \int_0^{QR} \frac{\mathrm{d}z}{[1+\rho(z)]^2} \right\} \phi(t)\,\mathrm{d}t. \quad (14)$$

To test this equation, two different distributions of mutation effects were used. In the first, every third site was subject to weaker selection (t_3) than the rest ($t_{1,2}$). Table 2 (case 1) shows the good agreement with the prediction from equation (14).

In the second test, the selective effects were drawn from a gamma distribution with parameters $\alpha = 0{\cdot}70$, $\beta = 0{\cdot}032$, which yields a mean t of 0·022, and fits the observations from *Drosophila* well (Keightley, 1994). Note, however, that this distribution will yield some mutations whose effect is too small to satisfy the assumption discussed above. We therefore used a gamma distribution truncated below at $Nt = 5$, as well as the full distribution, in the simulations. Since mutations of very small effect are not expected to

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

Table 1. *Summary of simulation results for multiplicative selection with mutations of equal effect*

Case	Selection	U	M	π/π_0 Observed (95% c.i.)	Expected
	Multiplicative				
1	$t = 0{\cdot}005$	0·08	0·16	0·64 (0·60–0·69)	0·62
2	$t = 0{\cdot}01$	0·08	0·16	0·61 (0·56–0·66)	0·63–0·64
3	$t = 0{\cdot}03$	0·08	0·16	0·72 (0·66–0·78)	0·68–0·69
4	$t = 0{\cdot}02$	0·1	1	0·93 (0·85–1·00)	0·90
5	$t = 0{\cdot}02$	0·4	1	0·61 (0·56–0·66)	0·65–0·68
6	$t = 0{\cdot}02$	0·4	0·75	0·55 (0·51–0·59)	0·57–0·60
7	$t = 0{\cdot}02$	0·4	0·5	0·46 (0·43–0·50)	0·45–0·47
8	$t = 0{\cdot}02$	0·4	0·1	0·069 (0·055–0·072)	0·052–0·055
	Multiplicative, weak selection				
9	$t = 3{\cdot}125 \times 10^{-4}$	0·025	0·16	0·93 (0·87–1·00)	0·86
10	$t = 6{\cdot}25 \times 10^{-5}$	0·005	0·16	1·00 (0·93–1·08)	0·97
11	$t = 10^{-3}$	0·08	0·16	0·71 (0·66–0·76)	0·61
12	$t = 10^{-3}$	0·16	0·32	0·67 (0·62–0·71)	0·61

The results are discussed in Sections 4(i)–(iii). The expected values were calculated from equation (7), using both Haldane's mapping function (13) and the linear function (8). If the two functions gave the same prediction only a single value is given, otherwise the low value comes from using Haldane's function and the high one from the linear function.

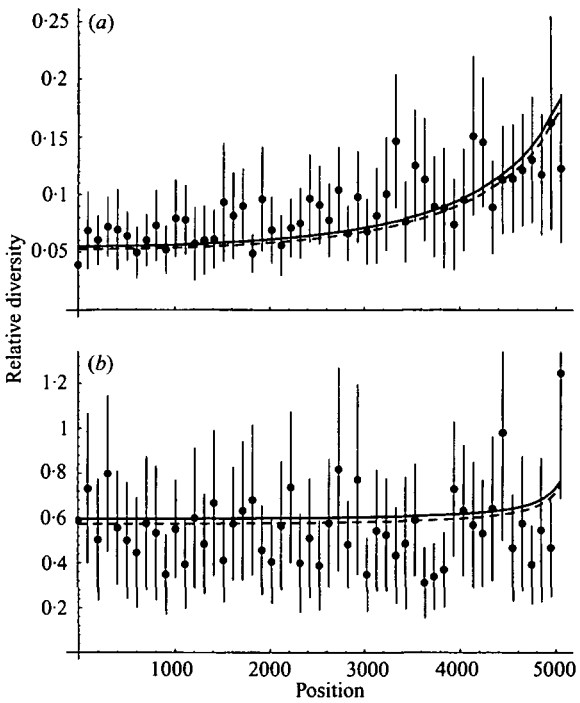

Fig. 2. The effect of background selection (means and 95% confidence intervals) along chromosomes of map lengths (*a*) $M = 0{\cdot}1$ and (*b*) $M = 0{\cdot}75$ (cases 8 and 6 of Table 1). The mutation rate is $U = 0{\cdot}4$ and selection is multiplicative with $t = 0{\cdot}02$. Neutral sites were placed at positions 0, $\pm 101, \pm 202, \ldots, \pm 5005$, but, since the chromosomes are symmetric, statistics were computed by combining results from the two halves. The numbers on the abscissa therefore represent loci counting from the centre (0). Means with 95% confidence intervals are shown. The expected values were calculated from equation (7), using both Haldane's mapping function (---) (13) and the linear function (——) (8).

cause much background selection, the simulation results with truncation should show stronger effect of background selection than those without truncation. Table 2 (cases 2–9) shows that the simulation results agree with this prediction. The predictions based on the full distribution might be expected to overestimate the effect of background selection, but this effect seems to be too small to detect except for the case with very tight linkage (case 5), when the truncated result and prediction also disagree significantly (case 9). We repeated the simulations of cases 7 and 9 with the gamma distribution truncated at $Nt = 25$. As expected (see Section 5), this makes no difference when the map is long (case 10), but significantly improves the prediction when the map is short (case 11). It is not clear what causes these disagreements but it is probably due to the small population size (see Section 5).

(v) *Non-multiplicative selection*

The effects of deviations from the assumption of multiplicativity were tested using two models of non-multiplicative interactions, additivity and synergistic epistasis (both described in Section 3). Given the small selective values used, and the observation that $(1-t)^n \approx 1-nt$ for small values of t, additive interactions are expected to give very similar results to multiplicative interactions. This is upheld for the additive case by the simulation results (Table 2, cases 12 and 13). The same is true for synergistic epistasis with $\beta = 0$ (case 14) which should be identical to multiplicative selection (Charlesworth,

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

Table 2. *Summary of simulation results examining the sensitivity of the predictions to the distribution of selective effects and the selection scheme*

Case	Selection	U	M	π/π_0 Observed (95% c.i.)	π/π_0 Expected
	Multiplicative, unequal effects				
1	$t_{1,2} = 0{\cdot}04, t_3 = 0{\cdot}01$	0·4	1	0·67 (0·62–0·69)	0·65–0·68
	Multiplicative, gamma-distributed effects				
2	$\alpha = 0{\cdot}70, \beta = 0{\cdot}032$	0·08	0·16	0·68 (0·63–0·72)	0·65–0·66
3	$\alpha = 0{\cdot}70, \beta = 0{\cdot}032$	0·4	1	0·67 (0·62–0·72)	0·66–0·68
4	$\alpha = 0{\cdot}70, \beta = 0{\cdot}032$	0·4	0·5	0·5 (0·46–0·53)	0·46–0·47
5	$\alpha = 0{\cdot}70, \beta = 0{\cdot}032$	0·4	0·1	0·16 (0·15–0·17)	0·045–0·047
	Multiplicative, gamma-distributed effects, truncated at $Nt = 5$				
6	$\alpha = 0{\cdot}70, \beta = 0{\cdot}032$	0·08	0·16	0·68 (0·63–0·73)	0·66–0·67
7	$\alpha = 0{\cdot}70, \beta = 0{\cdot}032$	0·4	1	0·63 (0·59–0·68)	0·66–0·68
8	$\alpha = 0{\cdot}70, \beta = 0{\cdot}032$	0·4	0·5	0·43 (0·40–0·46)	0·46–0·47
9	$\alpha = 0{\cdot}70, \beta = 0{\cdot}032$	0·4	0·1	0·12 (0·11–0·13)	0·051–0·053
	Multiplicative, gamma-distributed effects, truncated at $Nt = 25$				
10	$\alpha = 0{\cdot}70, \beta = 0{\cdot}032$	0·4	1	0·64 (0·59–0·69)	0·65–0·68
11	$\alpha = 0{\cdot}70, \beta = 0{\cdot}032$	0·4	0·1	0·088 (0·081–0·096)	0·070–0·072
	Multiplicative				
12	$t = 0{\cdot}02$	0·08	0·16	0·65 (0·60–0·70)	0·66–0·67
	Additive				
13	$t = 0{\cdot}02$	0·08	0·16	0·65 (0·60–0·70)	0·66–0·67
	Synergistic epistasis, $\alpha = 0{\cdot}02$				
14	$\beta = 0$	0·08	0·16	0·68 (0·63–0·73)	0·66–0·67
15	$\beta = 0{\cdot}01$ ($\bar{n} = 2{\cdot}067$)	0·08	0·16	0·75 (0·69–0·81)	0·70–0·71
16	$\beta = 0{\cdot}1$ ($\bar{n} = 0{\cdot}7393$)	0·08	0·16	0·79 (0·73–0·85)	0·80
17	$\beta = 0{\cdot}01$ ($\bar{n} = 6{\cdot}111$)	0·4	0·5	0·47 (0·43–0·50)	0·49–0·51
15	$\beta = 0{\cdot}1$ ($\bar{n} = 2{\cdot}199$)	0·4	0·5	0·59 (0·54–0·64)	0·57–0·60

The results are discussed in Section 4(iv)–(v). The expected values were calculated from equation (7) or (14) as appropriate, using both Haldane's mapping function (13) and the linear function (8) as in Table 1. Under synergistic epistasis, the expected effect was calculated by the method explained in the text. $\bar{n}$ is the mean number of mutations per individual.

1990). For $\beta \neq 0$, we estimated the 'effective' t as $U/\bar{n}$ (Charlesworth, 1990, p. 204), where $\bar{n}$ is the mean number of deleterious mutations per individual, calculated numerically by the method of Kimura & Maruyama (1966), assuming segregation but no recombination. The resulting value was then used as t in equation (7) to calculate the prediction for π/π_0. The calculated values of $\bar{n}$, the simulation results, and the predictions are shown in Table 2 (cases 15–18). Again, the agreement is clearly satisfactory, even with strong epistasis ($\beta \gg \alpha$).

(vi) *Multiple chromosomes*

The approximations derived in Section 2 can easily be generalized to yield predicted values for the case of multiple chromosomes, simply by letting $r(z) = 1/2$ for all loci that lie on chromosomes other than the one carrying the neutral locus under consideration. This was tested using four chromosomes, each with $U = 0{\cdot}2$ and $M = 0{\cdot}25$. Multiplicative interactions between mutations of equal effect, $t = 0{\cdot}02$, and a population size of $N = 1600$ were assumed. Simulation of this model gave $\pi/\pi_0 = 0{\cdot}46$, with a 95% confidence interval of (0·44–0·48). The predicted value using Haldane's mapping function (13) is 0·47 and that using the linear mapping function (8) is 0·48. The effect of the additional three chromosomes is extremely small; the predictions for a single chromosome with $U = 0{\cdot}2$ would be 0·48 and 0·50 for the Haldane and linear mapping functions respectively. The correction for multiple chromosomes can also be calculated from the effect of unlinked loci on effective population size [cf. Appendix (iii)].

(vii) *Background selection close to chromosome tips*

As we have seen, the effect of background selection should be less pronounced close to the tips of chromosomes. Figure 3 shows this very clearly. The results shown are those from the simulation with four chromosomes described in Section 4(vi). There were a total of 12007 loci, with seven neutral loci at positions $\{-6003, -4002, -2001, 0, 2001, 4002, 6003\}$ on each chromosome. To calculate confidence intervals for each position, results from equivalent positions (e.g. -4002 and 4002 from all four chromosomes) were pooled. The predicted values were obtained from equation (7) with the correction for multiple chromo-

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

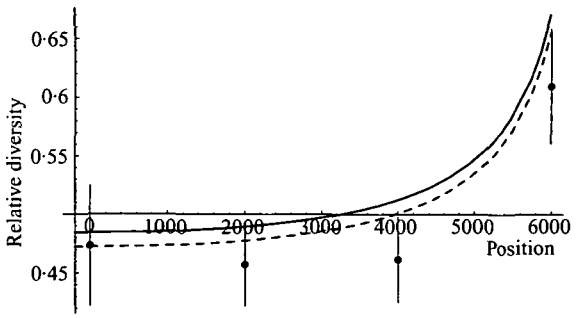

Fig. 3. An illustration of the decreased effect of background selection close to tips of chromosomes. The numbers on the abscissa represent loci counting from the centre (0). The dots are means with 95% confidence intervals. The broken and whole lines show the predicted π/π_0 using Haldane's mapping function (---) (13) and the linear function (——) (8), respectively. Parameters are $U = 0{\cdot}2$, $M = 0{\cdot}25$, and $t = 0{\cdot}02$. See Section 4(vii) for details.

somes given in Section 4(vi). The prediction is extremely good for the centre, but seems to overestimate the edge effect slightly.

5. Discussion

The present results, and the results of Hudson & Kaplan (1995) for short maps, show that a relatively simple formula [equation (4)] can predict the effects of background selection on genetic diversity at a neutral locus in a random-mating population, under a wide range of models of selection and map lengths. The major assumptions used to obtain this result are that: (*a*) the population size is sufficiently large that the frequencies of mutant alleles are close to their equilibrium values for an infinite population; (*b*) there is a negligible contribution to genetic diversity from neutral variants associated with deleterious mutations [see Appendix, Section (ii)]; (*c*) different loci act multiplicatively to determine fitness, and are independently distributed. Our simulations violate (*a*) grossly, and it is therefore rather surprising that the results agree with the theoretical predictions at all. The likely reason for the agreement is that, in the simulations, a given point on the chromosome is surrounded by a large group of closely linked selected loci. Since $Nt_i \gg 1$, the total number of mutant alleles in the group will be close to the expected number for large populations, even though each locus is far from being at deterministic equilibrium. A neutral locus elsewhere on the chromosome thus perceives the group as if it were a single locus in mutation–selection equilibrium. Given this, our simulation results indicate that the equation (4) applies both when each locus is subject to the same selection intensity, and when there is a wide distribution of selection coefficients (Tables 1 and 2). It even seems to work well when there is substantial synergistic epistasis for the fitness effects of mutant alleles (Table 2), which is a potentially important feature of systems with multiple loci subject to mutation–selection balance (Kimura & Maruyama, 1966; Crow, 1970; Feldman *et al.* 1980; Kondrashov, 1988; Charlesworth, 1990).

An important question concerning the applicability of these results is whether very weak selection is often likely to cause the failure of predictions based on equation (4). Assumption (*b*) can fail even if $N_e t$ is of the order of 100 or more, as indicated by our previous simulation results, leading to overestimates of the effect of background selection on genetic diversity (see Charlesworth *et al.* 1993, table 2). One reason for this effect appears to be that the mean time to loss of a very weakly deleterious allele in a finite population is not very different from the neutral value (Kimura & Ohta, 1969). A neutral variant associated with a chromosome carrying one or more weakly deleterious mutations therefore persists in the population almost as long as if there is no selection, if it does not recombine onto a mutation-free background, even though it is ultimately destined for elimination. In addition, as pointed out by Charlesworth *et al.* (1993), when the mean number of deleterious mutations in a non-recombining chromosomal region is substantially greater than one, a chromosome with a small number of deleterious mutations is actually at a selective advantage relative to the overall population. If U is small, such a chromosome can persist in the population for a long time before it accumulates further deleterious mutations and becomes selectively disadvantageous. Both of these factors can lead to failure of the assumption that neutral variants contribute little to diversity if they are associated with mutant-carrying chromosomes. Coalescent-process simulations of the case with no recombination indicate that rather weak selection (with t of the order of 10^{-3} or less) is required for large deviations from the predicted effects of background selection, if the population size is larger than 500000 (Charlesworth *et al.* 1995). Bigger deviations from the predictions are found in small populations than in large populations with the same selection coefficient, although the effect is not very sensitive to population size after a value of a few tens of thousands is reached, except if the expected reduction in diversity is very large (Table 3).

While the slowness of our Monte Carlo simulations with recombination and large population size precludes an exhaustive analysis of the conditions under which the theoretical predictions breakdown, the examples we have studied suggest that the above conclusions are qualitatively valid when there is some recombination (Table 1). However, the argument given in the Appendix [Section (ii)] suggests that assumption (*b*) is less likely to be violated when recombination occurs, consistent with the results in Table 3, where the largest discrepancies due to weak selection are observed with tight linkage. Of course, if population size and selection coefficients are both so small that $N_e t$ is of the order of one or less, then the fundamental assumption that the number of

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

Table 3. *Dependence of background selection effects on selection coefficient and population size, when there is weak selection and no recombination*

Theoretical $\pi/\pi_0(\times 10^{-1})$:	0·82	0·24	0·07
$t = 2\times 10^{-3}$			
U	$1{\cdot}0\times 10^{-2}$	$1{\cdot}5\times 10^{-2}$	$2{\cdot}0\times 10^{-2}$
Simulated $\pi/\pi_0(\times 10^{-1})$:			
$N = 5000$	1·797±0·018	1·354±0·013	1·255±0·014
$N = 25000$	1·050±0·004	0·478±0·014	0·330±0·010
$N = 100000$	0·846±0·027	0·318±0·014	0·136±0·008
$N = 500000$	0·831±0·027	0·251±0·014	0·081±0·008
$N = 1000000$	0·789±0·029	0·223±0·014	0·071±0·006
$t = 2\times 10^{-4}$			
U	$1{\cdot}0\times 10^{-3}$	$1{\cdot}5\times 10^{-3}$	$2{\cdot}0\times 10^{-3}$
Simulated $\pi/\pi_0(\times 10^{-1})$:			
$N = 5000$	8·684±0·072	8·721±0·071	8·672±0·076
$N = 25000$	2·735±0·045	2·391±0·041	2·265±0·020
$N = 100000$	1·311±0·031	0·823±0·020	0·698±0·018
$N = 500000$	0·969±0·030	0·336±0·014	0·230±0·018
$N = 1000000$	0·989±0·035	0·314±0·016	0·130±0·008

The simulated values of π/π_0 were obtained from the means of 2000 replicates of the coalescent process simulation procedure of Charlesworth *et al.* (1995), with a sample size of 100 haploid genomes per replicate.

deleterious mutations per individual is close to equilibrium will be invalid even in the multi-locus case, and background selection will be ineffective (Charlesworth *et al.* 1993). Although there is evidence for variation in the strength of selection against new mutations affecting viability in *D. melanogaster* (Keightley, 1994), the fact that the harmonic mean of t is about 0·02 (Crow & Simmons, 1983) means that only about 2% of mutations would have $t < 10^{-4}$ under the gamma distribution parameters fitted in Section 4(iv). Even with an N_e of 20000, as estimated for a local population of *D. melanogaster* by Mukai & Yamaguchi (1974), a t value of 10^{-4} would be sufficient to prevent fixation of deleterious alleles (Crow & Kimura, 1970).

These considerations suggest that we can be fairly confident that investigations based on the validity of equation (4) and its corollaries, such as that of Hudson & Kaplan (1995), will provide useful predictions of the background selection effects of all but the most weakly selected loci, at least as far as *Drosophila* populations are concerned. This provides a justification for using the analytical results based on equation (4) to make detailed predictions about the effect of background selection on the relation between chromosomal location and DNA variation for *D. melanogaster*, as has been done by Hudson & Kaplan (1995) and B. Charlesworth (submitted). These predictions may help to distinguish between alternative explanations of the empirical association between low local recombination rates and low genetic variation (Aquadro *et al.* 1994; Kreitman & Wayne, 1994; Braverman *et al.* 1995; Charlesworth *et al.* 1995; Hudson & Kaplan, 1995; Simonsen *et al.* 1995).

Some significant implications about the effects of background selection in the presence of recombination follow from equation (4) and its corollaries. First, provided that recombination does not vary with map position and that the map length is less than one Morgan or so, equation (10) can be applied to neutral loci that are not too close to the tips of the chromosome. This equation implies that the reduction in genetic diversity for such loci is a negative exponential function of the ratio of the diploid mutation rate U for the chromosome to its map length in Morgans. This result was also obtained by a different method by Hudson & Kaplan (1995), and a similar formula was derived by Barton (1995) for the effect of background selection on the probability of fixation of a selectively favourable mutation. For *D. melanogaster*, a U of 0·4 for one of the major autosomes is probably close to the true value, or even somewhat conservative (Keightley, 1994). Given that there is no recombination in males, the effective map length of an autosome is about 0·5 Morgans, so that a reduction in neutral diversity to about 45% of the classical neutral value would be predicted. More accurate calculations, taking into account the suppression of crossing over near the centromeres and higher frequencies of exchange elsewhere, give a value close to 62% for loci in the most freely recombining sections of the chromosome, with much smaller values in the centromeric regions (Hudson & Kaplan, 1995; B. Charlesworth, submitted). A somewhat smaller effect (70% of the neutral value) is predicted for the X chromosome. This suggests that all loci on the major chromosomes of *D. melanogaster* experience the effects of background selection from genes on the same

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

chromosome as themselves [the results of the Appendix, Section (iii), indicate that unlinked loci are likely to have little effect].

Drosophila species with the basic complement of four acrocentric autosomes, rather than the two metacentrics derived from fusions of acrocentrics, which are characteristic of the *melanogaster* species subgroup (Ashburner, 1989, ch. 37), would be expected to have substantially smaller effects of background selection, since the mutation rate per autosome is then half the value for *D. melanogaster*. Additional effects will be produced by higher rates of recombination per nucleotide site. The combined effect of these factors could be substantial. For example, with $U/M = 0{\cdot}2/1$ instead of $0{\cdot}4/0{\cdot}5$, the expected relative diversity, π/π_0, is 82% instead of 45%. An autosomal neutral locus in a species such as *D. virilis* or *D. subobscura*, which have much larger map lengths per chromosome arm than *D. melanogaster* and four major autosomes instead of two (Alexander, 1976; Krimbas, 1993), might thus have nearly twice as much neutral variation than a mid-arm locus of *D. melanogaster*, other things being equal. Differences among members of the *melanogaster* subgroup are discussed by B. Charlesworth (submitted).

It is difficult to predict the likely effect of background selection in other taxa, in the absence of estimates of U values for higher organisms. Currently, the only direct estimates of this parameter are for *Drosophila* (Crow & Simmons, 1983; Keightley, 1994). While there are still considerable uncertainties associated with these estimates, it seems likely that the value of U for *Drosophila* is approximately 1, yielding the value of 0·4 for the major autosomes of *D. melanogaster* used above. This would presumably apply to other taxa of higher insects. An indirect approach for highly self-fertilizing species of plants yields a similar value (Charlesworth *et al.* 1990; Charlesworth *et al.* 1994; Johnston & Schoen, 1995). A method for determining U for mammals from data on rates of molecular evolution has recently been proposed, but no results have yet been reported (Kondrashov & Crow, 1993). The mutation rate per locus for visible mutations in mammals seems to be similar to that for *Drosophila* (Kondrashov & Crow, 1993), but there are more than four times as many genes (Bird, 1995). This suggests that U for mammals may be at least 4, which we may adopt as a working estimate. In humans, with 23 chromosomes, the total sex-averaged map length is about 40 Morgans (Morton, 1991), so that the mean map length per chromosome is approximately 1·75 Morgans, and the mean U per chromosome is 0·17. With the Kosambi mapping function (Kosambi, 1944), which is commonly used in mammalian mapping studies, the expected nucleotide site diversity relative to the classical neutral value would be 0·90 for a centrally located gene, and 0·95 for a distal one, assuming that $t = 0{\cdot}02$. For the mouse, with a much shorter total map length of about 14 Morgans (Dietrich *et al.* 1992), the central and distal relative diversity values would be predicted to be 0·77 and 0·87 respectively. This suggests that recombinational differences could contribute to differences in genetic diversity and rates of molecular evolution among mammalian species. No data are currently available to determine whether or not such differences are observed. In plants, map lengths of about 1–2 Morgans per chromosome have been found in genome mapping projects (Tanksley *et al.* 1992; Ahn & Tanksley, 1993). With a mutation rate of 1 per genome, and 10 chromosomes with a map length of 1·2, as in tomatoes (Tanksley *et al.* 1992), this would yield central and distal relative diversity values of 0·92 and 0·96, suggesting a rather weak overall effect of background selection.

The above predictions ignore regional differences in recombination within chromosomes. In humans, there is evidence for pericentric reductions but telomeric increases in exchange rates per nucleotide, compared with the intervening regions of chromosome arms (NIH/CEPH Collaborative Mapping Group, 1992). In plants, there is often strong centromeric and telomeric suppression of recombination, extending over large sections of the chromosomes (Neuffer & Coe, 1974; Tanksley *et al.* 1992). Background selection and selective sweeps may thus produce significant within-chromosome structuring of patterns of genetic diversity and molecular evolution in these species, as in *Drosophila*. This possibility should be taken into consideration in future studies of molecular evolution and variation.

As pointed out in Section 2(iv), the background selection effects of weakly selected loci are much more sensitive to recombination than those of strongly selected loci, although (for the same mutation rate) a weakly selected locus can have a larger background selection effect than a strongly selected locus if linkage is very tight. As displayed in Fig. 1, if there is regional variation in the frequency of recombination per nucleotide, weakly selected loci will contribute more to regional differences in the level of neutral variability than strongly selected loci. It has been suggested that there may be a long tail of deleterious mutations with very small effects among the viability mutations detected in the *Drosophila* mutation accumulation experiments (Keightley, 1994; Lande, 1994). In addition, there is evidence that the abundant transposable elements found in natural *Drosophila* populations have very slightly deleterious fitness effects, with mean selection coefficients of approximately 2×10^{-4} (Charlesworth *et al.* 1992). As suggested by Hudson (1994), they could make a substantial contribution to the reduction in variability in regions of low recombination. This possibility will be explored in more detail in a subsequent paper (B. Charlesworth, submitted).

Finally, it is interesting to note that similar formulæ

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

for the influence of linked loci subject to deleterious mutations have emerged from the study of their effects on the fixation probabilities of favourable mutations with selection coefficients which are sufficiently large that branching process theory can be used (Barton, 1995), and on nucleotide site diversity at neutral sites (Hudson, 1994; Hudson & Kaplan, 1994, 1995). Related results have also been derived for the effect on neutral loci of directional selection on a quantitative trait (Santiago & Caballero, 1995; Santiago, in preparation), and for the effects of selective sweeps and temporally fluctuating selection coefficients (Barton, 1995). But, except for the case of complete linkage (Charlesworth, 1994), we are currently lacking results on the effects of background selection on weakly selected mutations, for which branching process theory cannot be used. With free recombination, all approaches lead to the conclusion that the effects of selection can be approximately represented by a reduction in effective population size N_e. The factor by which N_e is divided to obtain the relevant expression in each case is approximately equal to one plus four times the additive genetic variance in family size, if fitness is measured relative to the population mean. More complex formulæ obviously apply with linkage (cf. equation [4]). But it is important to note that N_e is not a sufficient descriptor of the effects of selection at linked sites. As shown by Charlesworth *et al.* (1993) and Hudson (1994) for background selection, and by Braverman *et al.* (1995) and Simonsen *et al.* (1995) for selective sweeps, there may be very different effects on the nucleotide site diversity and on the number of segregating sites. This would not be expected if these processes could be described simply in terms of a reduction in N_e. Rather, both the genealogical structure of the population and the expected time to coalescence between a pair of genes (which is controlled by N_e in the classical neutral model), are affected by selection at linked sites. It is much more difficult to obtain useful analytical results for this effect than for the reduction in nucleotide site diversity, or the fixation probability.

Appendix

The application of the linear diffusion operator method (Ohta & Kimura, 1969) to the case of a neutral locus and a single selected locus will be presented here.

(i) *Diffusion coefficients*

Deterministic equations for the changes in the x_i can easily be written down on the assumption that the locus under selection is in equilibrium. Second-order terms in q_i, t_i and u_i will be neglected in what follows. To the assumed order of approximation, the mean fitness of the population is $1-2u_i$ (Crow & Kimura 1970, Chap. 6), the marginal fitness of a gamete carrying a wild-type allele at the selected locus is $1-q_i t_i \approx 1-u_i$, and the marginal fitness of a mutant-carrying gamete is $1-t_i-u_i$. Let the frequencies of the gametes $A_i B$, $A_i b$, $a_i B$, and $a_i b$ be $y_1 = p_i x_0$, $y_2 = p_i(1-x_0)$, $y_3 = q_i x_1$, and $y_4 = q_i(1-x_1)$, respectively. The coefficient of linkage disequilibrium is $y_1 y_4 - y_2 y_3 \approx q_i \delta$.

These expressions can be substituted into the standard equations for two loci (Crow & Kimura, 1970, ch. 5), with the addition of terms to included the effect of mutation from A_i to a_i, in order to obtain expressions for the deterministic changes in x_0 and x_1. Writing M_i for the changes in x_i from selection, mutation and recombination, letting $\tilde{r}_i = r_i(1-t_i)$, and noting that the net frequencies of the alleles at the selected locus remain unchanged over a generation, we obtain

$$M_0 \equiv \frac{\Delta y_1}{p_i} \approx \frac{1}{p_i(1-2u_i)} \times [(1-u_i)y_1 - u_i y_1 - \tilde{r}_i q_i \delta - (1-2u_i)y_1], \quad \text{(A 1)}$$

and

$$M_1 \equiv \frac{\Delta y_3}{q_i} \approx \frac{1}{q_i(1-2u_i)} \times [(1-t_i-u_i)y_3 + u_i y_1 + \tilde{r}_i q_i \delta - (1-2u_i)y_3], \quad \text{(A 2)}$$

which, upon neglecting second-order terms and utilizing the relations $u \approx q_i t_i$ and $u_i \ll t_i$, can be further simplified to yield

$$M_0 \approx -q_i \tilde{r}_i \delta, \quad \text{(A 3)}$$

and

$$M_1 \approx p_i(t_i + \tilde{r}_i)\delta. \quad \text{(A 4)}$$

The assumption of equilibrium at the selected locus further means that the covariance between x_0 and x_1 is zero, and that the effects of genetic drift on the x_i can be represented by sampling within the corresponding gamete classes, so that the sampling variances for the changes in the x_i over one generation are given by

$$V_0 \approx \frac{x_0(1-x_0)}{2p_i N_e} \quad \text{(A 5)}$$

and

$$V_1 \approx \frac{x_1(1-x_1)}{2q_i N_e}, \quad \text{(A 6)}$$

where N_e is the variance effective population size (Crow & Kimura, 1970, ch. 8). The validity of the diffusion approximation used here requires $2q_i N_e \gg 1$ (Ewens, 1979, ch. 4). Since the assumption that q_i is close to its equilibrium value requires $2u_i N_e \gg 1$

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

(Crow & Kimura, 1970, pp. 443–444), this condition should automatically be met whenever the model is appropriate.

(ii) *Genetic diversity at statistical equilibrium*

For details of the method used in this section, see Ewens (1979, section 4.10), Kimura & Ohta (1971, appendix 3), or Stephan *et al.* (1992). For a function $g(x_0, x_1)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}\tau}E_\tau(g) = E_\tau\left[M_0\frac{\partial g}{\partial x_0}+M_1\frac{\partial g}{\partial x_1}+\frac{1}{2}\left(V_0\frac{\partial^2 g}{\partial x_0^2}+V_1\frac{\partial^2 g}{\partial x_1^2}\right)\right], \tag{A 7}$$

where E_τ denotes expectation taken at a given point τ in time. The object is to find a set of equations in functions g_i that result in a soluble set of equations for the expectations. For the set

$$g_0 = 2x_0(1-x_0), \tag{A 8a}$$

$$g_1 = 2q_i(1-2x_0)\,\delta, \tag{A 8b}$$

$$g_2 = \delta^2, \tag{A 8c}$$

we have the following equations for the effects of selection, recombination and drift:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}E(g_0) \approx -\frac{E(g_0)}{2p_i N_e}-\tilde{r}_i E(g_1), \tag{A 9a}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}E(g_1) \approx -\frac{q_i E(g_0)}{p_i N_e}-[q_i\tilde{r}_i+p_i(t_i+\tilde{r}_i)] \times E(g_1)+4q_i^2\tilde{r}_i E(g_2), \tag{A 9b}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}E(g_2) \approx \frac{E(g_0)}{4p_i q_i N_e}-\frac{E(g_1)}{4q_i^2 N_e} -\left[\frac{1}{2q_i N_e}+2q_i\tilde{r}_i+2p_i(t_i+\tilde{r}_i)\right]E(g_2), \tag{A 9c}$$

where all expectations are taken at time τ.

At statistical equilibrium, the sum of the contribution from mutation at the neutral sites to the expected change in each function and the change in the expectation of the function due to drift and the other deterministic forces [represented by the appropriate member of equations (A 9)] must be equal to zero. The mutational term for a given function is equal to the change in the value of the function due to a single new mutation that arises in a non-segregating population, times the rate per generation at which mutations arise (Kimura & Ohta, 1971, pp. 186–187). If the total breeding population size is N, the change in g_0 due to a single new mutation that arises in a mutant-free chromosome is $1/(Np_i)$ (neglecting terms of order $1/N^2$). The expected number of mutations that arise per generation is $2Nv$, and the probability that each arises in a mutant-free chromosome is p_i, so that the mutational change in the expectation of g_0 is $2v$. It is easily seen that the mutational change in the expectation of δ is zero, so that the change in the expectation of g_1 is of order $1/N^2$. The change in the expectation of g_2 is $v/2Nq_i$, which is negligible for sufficiently small v/N. Thus we have

$$-2v \approx -\frac{E(g_0)}{2p_i N_e}-\tilde{r}_i E(g_1), \tag{A 10a}$$

$$0 \approx -\frac{q_i E(g_0)}{p_i N_e}-[q_i\tilde{r}_i+p_i(t_i+\tilde{r}_i)] \times E(g_1)+4q_i^2\tilde{r}_i E(g_2), \tag{A 10b}$$

$$0 \approx \frac{E(g_0)}{4p_i q_i N_e}-\frac{E(g_1)}{4q_i^2 N_e} -\left[\frac{1}{2q_i N_e}+2q_i\tilde{r}_i+2p_i(t_i+\tilde{r}_i)\right]E(g_2), \tag{A 10c}$$

where all expectations are now taken at statistical equilibrium. Equations (A 10) form a linear system of equations. Their approximate solutions can be found as follows.

The leading term in the factor of $E(g_2)$ in equation (A 10*c*) is $2p_i(t_i+\tilde{r}_i)$ provided that $4q_i(t_i+\tilde{r}_i)N_e \gg 1$. The neglected term $1/2q_i N_e$ only contributes to equation (A 10*b*) if $r_i \neq 0$, and its contribution to the final result is negligible if $4q_i N_e(t_i+\tilde{r}_i) \gg 1$. Neglecting the other terms, equation (A 10*c*) gives

$$E(g_2) \approx \frac{E(g_0)/p_i - E(g_1)/q_i}{8p_i q_i(t_i+\tilde{r}_i)N_e}. \tag{A 11}$$

Substitution of this into equation (A 10*b*), again retaining only the leading term in $E(g_1)$ and neglecting second-order terms, gives

$$E(g_1) \approx -\left(1-\frac{\tilde{r}_i}{2(t_i+\tilde{r}_i)}\right)\frac{q_i E(g_0)}{(t_i+\tilde{r}_i)N_e}. \tag{A 12}$$

$E(g_0)$ can now be obtained by substituting this expression into equation (A 10*a*). Using this in conjunction with equation (A 12), we have

$$E(g_0) \approx \frac{4p_i N_e v}{1-\dfrac{2q_i\tilde{r}_i}{t_i+\tilde{r}_i}\left(1-\dfrac{\tilde{r}_i}{2(t_i+\tilde{r}_i)}\right)[1+O(t_i)]}. \tag{A 13}$$

Write π for the expectation of the genetic diversity, G. Then π in the absence of background selection is $\pi_0 = 4N_e v$ (Kimura, 1969). Rearranging equation (A 13), and using equations (2) and (A 11), we obtain

$$\frac{\pi}{\pi_0} = \frac{E(g_0)-E(g_1)}{4N_e v} = \frac{1+O\left(\dfrac{q_i}{(t_i+\tilde{r}_i)N_e}\right)}{1+\dfrac{q_i[1+O(t_i)]}{(1+\rho_i)^2}}, \tag{A 14}$$

where $\rho_i = \tilde{r}_i/t_i$ measures the frequency of recombination relative to the selection coefficient for the ith locus.

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

We can therefore write

$$\frac{\pi}{\pi_0} \approx \frac{1}{1+\dfrac{q_i}{(1+\rho_i)^2}}, \tag{A 15}$$

provided that $q_i \ll (t_i+\tilde{r}_i)N_e$. This condition can be violated if $(t_i+\tilde{r}_i)N_e$ is sufficiently small. For fixed q_i, this is most likely to occur with complete linkage, weak selection, and small population size. In such cases, $E(g_1)$ contributes significantly to π, and the sign of $E(g_1)$ from equation (A 12) implies that π is underestimated by ignoring it. In biological terms, this means that there is a contribution to π from neutral variants associated with deleterious alleles, which is neglected in equation (A 15). This term corresponds to that arising from variants initially associated with deleterious alleles, which was similarly neglected in the formula for the multi-locus case with no recombination obtained by Charlesworth *et al.* (1993).

(iii) *Effective population size with background selection*

With free recombination ($r_i = 1/2$), equation (A 15) gives

$$\frac{\pi}{\pi_0} \approx \frac{1}{1+4q_i t_i^2}. \tag{A 16}$$

The additive genetic variance in fitness, V_{Gi}, due to the ith selected locus is equal to $2q_i t_i^2$ (Mukai *et al.* 1974), so that the nucleotide site diversity is equivalent to the formula for genetic diversity in the classical neutral case, but with the effective size of the population divided by a factor of $1+2V_{Gi}$. This is consistent with the general theory of the effect of heritable variation in family size on effective population size. According to this theory, selection at unlinked loci contributes a term of four times the additive genetic variance in family size to the denominator of the formula for N_e, if family size is measured relative to the population mean (Robertson, 1961; Nei & Murata, 1966; Santiago & Caballero, 1995). The additive genetic value of a full-sib family is equal to the mean of the additive values of the two parents, so that the additive variance in family fitness is one-half V_G as defined here. The assumption of a single locus means that the additive variance for absolute fitness is nearly the same as that for fitness scaled relative to mean fitness, so that the two results are equivalent.

This suggests that the effect of background selection with arbitrary linkage can also be deduced from the general theory of effective population size. The detailed argument is as follows. Consider a neutral locus linked to a given selected locus i. The number of deleterious alleles at this locus in an individual can be treated as an additive trait. The breeding value of the progeny of a gamete with a value g_i due to its genotype at the selected locus is expected to be $g_i(1-t_i-\tilde{r}_i)$ after one generation, since there is a reduction $g_i t_i$ due to selection, and a further reduction $g_i \tilde{r}_i$ due to recombination. After a large number of generations, the sum of the contributions of this gamete to the breeding value of its descendants approaches $g_i/(t_i+\tilde{r}_i)$. The contribution to the divisor of N_e is given by the variance in family size associated with those terms (Robertson, 1961; Nei & Murata, 1966; Santiago & Caballero, 1995). This is equal to $q_i t_i^2/(t_i+\tilde{r}_i)^2$, which is identical with the corresponding term in equation (A 15). π/π_0 is thus equal to the ratio of N_e with background selection to the classical neutral value of N_e.

(iv) *The effects of multiple loci*

The above result for the effect of background selection on N_e suggests the following heuristic argument for the case of multiple selected loci. In general, the ratio of N_e with selection to the classical neutral value is $1/(1+V)$, where V is the heritable variance in family size, scaled relative to the population mean (Robertson, 1961; Nei & Murata, 1966; Santiago & Caballero, 1995). $1+V$ is equivalent to the expectation of the squares of the scaled family sizes contributed by selective differences among genotypes. With multiplicative fitnesses, and in the absence of linkage disequilibrium among the selected loci, this expectation is simply the product of the expectations contributed by each locus, i.e. $\prod_i [1+q_i t_i^2/(t_i+\tilde{r}_i)^2]$. With small effects at each locus, we thus have

$$\pi/\pi_0 = \prod_i [1+q_i t_i^2/(t_i+\tilde{r}_i)^2]^{-1} \approx \prod_i \left[1-\frac{q_i t_i^2}{(t_i+\tilde{r}_i)^2}\right], \tag{A 17}$$

which yields equation (4).

We wish to thank Henrik Nordborg for help with the simulations, Nick Barton, Armando Caballero, Bill Hill, Dick Hudson, and Wolfgang Stephan for helpful discussions and comments on the manuscript. This work was supported by National Science Foundation grant DEB9217683, the Darwin Trust of Edinburgh, the Underwood fund of the Biotechnology and Biological Sciences Research Council (UK), and the Sweden–America Foundation.

References

Ahn, S. & Tanksley, S. D. (1993). Comparative linkage maps of the rice and maize genomes. *Proceedings of the National Academy of Sciences, USA* **90**, 7980–7984.

Alexander, M. L. (1976). The genetics of *Drosophila virilis*. In *The Genetics of Drosophila* (ed. M. Ashburner and E. Novitski), vol. 1c, pp. 1365–1627. London: Academic Press.

Aquadro, C. F., Begun, D. J. & Kindahl, E. C. (1994). Selection, recombination, and DNA polymorphism in *Drosophila*. In *Non-Neutral Evolution: Theories and*

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

Molecular Data (ed. G. B. Golding), pp. 46–56. London: Chapman and Hall.

Ashburner, M. (1989). *Drosphila. A Laboratory Handbook*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Barton, N. H. (1994). The reduction in fixation probability caused by substitutions at linked loci. *Genetical Research* **64**, 199–208.

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

Berry, A. J., Ajioka, J. W. & Kreitman, M. (1991). Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. *Genetics* **129**, 1085–1098.

Bird, A. P. (1995). Gene number, noise reduction and biological complexity. *Trends in Genetics* **11**, 77–117.

Birky, Jr., C. W. & Walsh, J. B. (1988). Effects of linkage on rates of molecular evolution. *Proceedings of the National Academy of Sciences, USA* **85**, 6414–6418.

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphism. *Genetics* **140**, 783–796.

Caballero, A. (1995). On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* **139**, 1007–1011.

Charlesworth, B. (1990). Mutation–selection balance and the evolutionary advantage of sex and recombination. *Genetical Research* **55**, 199–221.

Charlesworth, B. (1994). The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetial Research* **63**, 213–227.

Charlesworth, B., Charlesworth, D. & Morgan, M. T. (1990). Genetic loads and estimates of mutation rates in highly inbred plant populations. *Nature* **347**, 380–382.

Charlesworth, B., Charlesworth, D. & Morgan, M. T. (1991). Multilocus models of inbreeding depression with synergistic selection and partial self-fertilisation, *Genetical Research* **57**, 177–194.

Charlesworth, B., Lapid, A. & Canada, D. (1992). The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genetical Research* **60**, 103–114.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

Charlesworth, D., Charlesworth, B. & Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics* **141**, 1619–1632.

Charlesworth, D., Lyons, E. E. & Litchfield, L. B. (1994). Inbreeding depression in two highly inbreeding populations of *Leavenworthia*. *Proceedings of the Royal Society, London*, B **258**, 209–214.

Crow, J. F. (1970). Genetic loads and the cost of natural selection. In *Mathematical Topics in Population Genetics* (ed. K. Kojima), pp. 128–177. Berlin: Springer-Verlag.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper & Row.

Crow, J. F. & Simmons, M. J. (1983). The mutation load in *Drosophila*. In *The Genetics and Biology of* Drosophila (ed. H. L. Carson, M. Ashburner and J. N. Thomson), vol. 3e, pp. 1–35. London: Academic Press.

Dietrich, W., Katz, H., Lincoln, S. E., Shin, H.-S., Friedman, J., Dracopoli, N. C. & Lander, E. S. (1992). A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131**, 423–447.

Ewens, W. J. (1979). *Mathematical Population Genetics*. Berlin: Springer-Verlag.

Feldman, M. W., Christiansen, F. B. & Brooks, L. D. (1980). Evolution of recombination in a constant environment. *Proceedings of the National Academy of Sciences, USA* **77**, 4838–4841.

Felsenstein, J. (1965). The effect of linkage on directional selection. *Genetics* **52**, 349–363.

Felsenstein, J. & Yokoyama, S. (1976). The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* **83**, 845–859.

Free Software Foundation (1992). GNU C++ Library. Publicly available via *ftp://prep.ai.mit.edu/*.

Gillespie, J. H. (1994). Alternatives to the neutral theory. In *Non-Neutral Evolution: Theories and Molecular Data* (ed. G. B. Golding), pp. 1–17. London: Chapman and Hall.

Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distance between loci of linked factors. *Journal of Genetics* **8**, 299–309.

Haldane, J. B. S. (1927). A mathematical theory of natural and artifical selection. Part V. Selection and mutation. *Proceedings of the Cambridge Philosophical Society* **23**, 838–844.

Hudson, R. R. (1994). How can the low levels of *Drosophila* sequence variation in regions of the genome with low levels of recombination be explained? *Proceedings of the National Academy of Sciences, USA* **91**, 6815–6818.

Hudson, R. R. & Kaplan, N. L. (1994). Gene trees with background selection. In *Non-Neutral Evolution: Theories and Molecular Data* (ed. G. B. Golding), pp. 140–153. New York: Chapman & Hall.

Hudson, R. R. & Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics* **141**, 1605–1617.

Johnston, M. O. & Schoen, D. J. (1995). Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. *Science* **267**, 226–229.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989). The 'hitch-hiking' effect revisited. *Genetics* **123**, 887–899.

Keightley, P. D. (1994). The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* **138**, 1315–1322.

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. & Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351.

Kimura, M. & Ohta, T. (1969). The average number of generations until extinction of an individual mutant gene in a population. *Genetics* **63**, 701–709.

Kimura, M. & Ohta, T. (1971). *Theoretical Aspects of Population Genetics*. Princeton: Princeton University Press.

Kliman, R. M. & Hey, J. (1993). Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Molecular Biology and Evolution* **10**, 1239–1258.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440.

Kondrashov, A. S. & Crow, J. F. (1993). A molecular approach to estimating the human deleterious mutation rate. *Human Mutation* **2**, 229–234.

Kosambi, D. D. (1944). The estimation of map distance from recombination values. *Annals of Eugenics* **12**, 172–175.

Kreitman, M. & Wayne, M. L. (1994). Organization of genetic variation at the molecular level: lessons from *Drosophila*. In *Molecular Ecology and Evolution: Approaches and Applications* (ed. B. Schierwater, B. Streit, G. P. Wagner and R. DeSalle), pp. 157–184. Basel: Birkhäuser.

Krimbas, C. B. (1993). *Drosophila subobscura*. Hamburg: Verlag Dr Kovač.

Lande, R. (1994). Risk of population extinction from

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press

fixation of new deleterious mutations. *Evolution* **48**, 1460–1469.

Maynard Smith, J. & Haigh, J. (1974). The hitchhiking effect of a favourable gene. *Genetical Research* **23**, 23–35.

McPeek, M. S. & Speed, T. P. (1995). Modeling interference in genetic recombination. *Genetics* **139**, 1031–1044.

Morton, N. E. (1991). Parameters of the human genome. *Proceedings of the National Academy of Sciences, USA* **88**, 7474–7476.

Mukai, T., Cardellino, R. K., Watanabe, T. K. & Crow, J. F. (1974). The genetic variance for viability and its components in a population of *Drosophila melanogaster*. *Genetics* **78**, 1195–1208.

Mukai, T. & Yamaguchi, O. (1974). The genetic structure of natural populations of *Drosophila melanogaster*. XI. Genetic variability in a local population. *Genetics* **76**, 339–366.

Nagylaki, T. (1995). The inbreeding effective population number in dioecious populations. *Genetics* **139**, 473–485.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei, M. & Murata, M. (1966). Effective population size when fertility is inherited. *Genetical Research* **8**, 257–260.

Neuffer, M. G. & Coe, E. H. (1974). Corn (maize). In *Handbook of Genetics* (ed. R. C. King), vol. 2, pp. 3–30. New York: Plenum.

NIH/CEPH Collaborative Mapping Group (1992). A comprehensive genetic linkage map of the human genome. *Science* **258**, 67–86.

Ohta, T. & Kimura, M. (1969). Linkage disequilibrium due to random genetic drift. *Genetical Research* **13**, 47–55.

Ohta, T. & Kimura, M. (1975). The effect of a selected locus on heterozygosity of neutral alleles (the hitch-hiking effect). *Genetical Research* **25**, 313–326.

Robertson, A. (1961). Inbreeding in artificial selection programmes. *Genetical Research* **2**, 189–194.

Santiago, E. & Caballero. A. (1995). Effective size of populations under selection. *Genetics* **139**, 1013–1030.

Simonsen, K. L., Churchill, G. A. & Aquadro, C. F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**, 413–429.

Stephan, W. (1995). An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorphism data. *Molecular Biology and Evolution* **12**, 959–962.

Stephan, W., Wiehe, T. H. E. & Lenz, M. W. (1992). The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theoretical Population Biology* **41**, 237–254.

Tanksley, S. D., Ganal, M. W., Prince, J. P., deVicente, M. C., Bonierbale, M. W., Broun, P., Fulton, T. M., Giovannoni, J. J., Grandillo, S., Martin, G. B., Messeguer, R., Miller, J. C., Miller, L., Paterson, A. H., Pineda, O., Roder, M. S., Wing, R. A., Wu, W. & Young, N. D. (1992). High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**, 1141–1160.

Thomson, G. (1977). The effect of a selected locus on linked neutral loci. *Genetics* **85**, 753–788.

Wiehe, T. H. E. & Stephan, W. (1993). Analysis of a genetic hitchhiking model and its application to DNA polymorphism data. *Molecular Biology and Evolution* **10**, 842–854.

https://doi.org/10.1017/S0016672300033619 Published online by Cambridge University Press