Genet. Res., Camb. (1994), **63**, *pp.* 213–227 *With* 3 *text-figures*

The effect of background selection against deleterious mutations on weakly selected, linked variants

BRIAN CHARLESWORTH
Department of Ecology and Evolution, The University of Chicago, 1101 E. 57th St., Chicago, IL 60637-1573, USA

(*Received 21 October 1993 and in revised form 11 January 1994*)

Summary

This paper analyses the effects of selection against deleterious alleles maintained by mutation ('background selection') on rates of evolution and levels of genetic diversity at weakly selected, completely linked, loci. General formulae are derived for the expected rates of gene substitution and genetic diversity, relative to the neutral case, as a function of selection and dominance coefficients at the loci in question, and of the frequency of gametes that are free of deleterious mutations with respect to the loci responsible for background selection. As in the neutral case, most effects of background selection can be predicted by considering the effective size of the population to be multiplied by the frequency of mutation-free gametes. Levels of genetic diversity can be sharply reduced by background selection, with the result that values for sites under selection approach those for neutral variants subject to the same regime of background selection. Rates of fixation of slightly deleterious mutations are increased by background selection, and rates of fixation of advantageous mutations are reduced. The properties of sex-linked and autosomal loci in random-mating populations are compared, and the effects of background selection on asexual and self-fertilizing populations are considered. The implications of these results for the interpretation of studies of molecular evolution and variation are discussed.

1. Introduction

In regions of the genome where recombination is infrequent, the amount of genetic variability at neutral sites can be reduced by selection against linked deleterious alleles, maintained in the population by recurrent mutation at many loci. This process was called 'background selection' by Charlesworth, Morgan & Charlesworth (1993), who showed that its magnitude depends on the total mutation rate to deleterious alleles for the genomic region in question and the frequency of recombination in that region. Given the fact that the mutation rate to alleles with detectable deleterious effects on fitness may be as high as an average of one or more new mutations per zygote per generation in higher organisms (Mukai *et al.* 1972; Crow & Simmons, 1983; Kondrashov, 1988; Houle *et al.* 1992), background selection may severely reduce neutral genetic variation in asexual or highly self-fertilizing species, in which there is effectively little or no recombination across the entire genome. The telomeric and centromeric regions of many randomly mating species also show greatly restricted recombination. Neutral sites in such regions would be expected to show a smaller but significant reduction in variation than in selfing or asexual populations, since they make up only a fraction of the total genome and so are subject to a lower rate of input of deleterious mutation (Charlesworth *et al.* 1993). The process of background selection provides a possible explanation for observations of reduced levels of DNA sequence variability in asexual and self-fertilizing taxa, and in regions of the *Drosophila* genome where recombination is restricted (Charlesworth *et al.* 1993).

The neutral model of molecular evolution (Kimura, 1983) may be viewed as a useful null model, whose statistical predictions can be tested against observed patterns of molecular evolution and variation (Kreitman, 1991; McDonald & Kreitman, 1991). There is still considerable controversy concerning the extent to which nucleotide site variants are subject to natural selection, particularly for substitutions causing amino-acid replacements (Kimura, 1983; Gillespie, 1991; McDonald & Kreitman, 1991). It is thus important to determine whether background selection distorts the predictions of selective versus neutral models.

In addition, if background selection operates differently on sites subject to different kinds of evolutionary forces, differences in patterns of molecular evolution and variation between taxa or genomic regions subject to different levels of background selection could shed light on the nature of the evolutionary forces affecting molecular variation. It is intuitively fairly obvious that both background selection and the related process of hitch-hiking of variants by linked selectively favourable alleles (Maynard Smith & Haigh, 1974; Ohta & Kimura, 1975; Thomson, 1977; Birky & Walsh, 1988; Kaplan, Hudson & Langley, 1989; Stephan, Wiehe & Lenz, 1992; Wiehe & Stephan, 1993) are analogous to a reduction in the effective population number for the region of the genome concerned. We would accordingly expect the level of variability for both neutral and selected variants to be reduced by background selection, although the extent of this reduction will depend on the mode and strength of selection. If effective population number is greatly reduced, the behaviour of weakly selected and neutral sites will become indistinguishable. The rate of substitution of favourable variants should thus be reduced by background selection and hitch-hiking, whereas that for deleterious variants will be increased, and neutral variants will be unaffected. Birky & Walsh (1988) have previously shown this to be true, but did not provide any explicit formulae for the size of the effects.

The main purpose of this paper is to investigate the expected effects of background selection under different assumptions about breeding system and mode of inheritance. Complete linkage is assumed throughout, as this enables explicit formulae to be developed and provides an upper bound to the effect of background selection. Sex-linked as well as autosomal loci will be considered, as both play a prominent role in empirical work on *Drosophila*.

2. A general model

(i) *General considerations*

Background selection is assumed to operate by the elimination of new mutations at the loci of interest if they arise in gametes that carry one or more deleterious mutations (the loci in question will be referred to as *background loci*). Under equilibrium between selection and mutation in a large population, the frequency of gametes free of deleterious mutations is f_0, where f_0 is a function of the diploid mutation rate per genome to deleterious alleles (U), and of the selection coefficient and dominance coefficient for a single deleterious mutant allele at a background locus. These are denoted by s and h respectively, where s is the reduction in fitness to a mutant homozygote and sh is the reduction in fitness experienced by a heterozygous carrier (see Table 1 for a list of symbols). Equal fitness effects of each locus and equal effects on the two sexes, together with multiplicative fitness interactions between deleterious alleles at different background loci, are assumed. The value of f_0 depends on the mating system and mode of inheritance e.g. with autosomal inheritance and random-mating, we have $f_0 = \exp(-U/2hs)$, whereas with sex-linkage $f_0 = \exp(-3U/2s[2h+1])$ (Charlesworth, *et al.* 1993). For chromosome regions of comparable physical size and map length, we would therefore expect larger values of f_0 for X-linked loci, given that deleterious alleles are generally partly recessive (Crow & Simmons, 1983). An approximate f_0 value of 0·2 seems reasonable for an autosomal centromeric region in *D. melanogaster*, 0·6 for the centromeric region of the X, and 0·08 for a highly selfing species (Charlesworth *et al.* 1993).

A general formula for the rate of substitution of weakly selected mutations in a genome influenced by background selection will be given first, followed by a formula for the expected nucleotide-site diversity which they contribute during their sojourn in the population, under the infinite-sites model of molecular variation (Kimura, 1969, 1971). Selection on the mutations studied in this way is assumed to be much weaker than on the background loci subject to recurrent deleterious mutations. This is probably

Table 1. *Definitions of some quantities introduced in the text*

U	Mean number of new deleterious mutations per individual per generation, at loci responsible for background selection (background loci)
f_0	Equilibrium frequency of gametes free of mutations at background loci
s	Selection coefficient against homozygous mutations at background loci
h	Dominance coefficient for fitness effect of a background locus
v_f	Rate of germline mutation per nucleotide site in females
v_m	Rate of germline mutation per nucleotide site in males
σ	Selection coefficient on homozygote for a weakly selected variant
θ	Dominance coefficient for fitness effect of a weakly selected variant
S	Product of σ and effective population number
$\tilde{K}$	Rate of substitution of variants under background selection, relative to the neutral value
$\tilde{\pi}$	Nucleotide site diversity under background selection, relative to the neutral value without background selection
$\tilde{\pi}^*$	Nucleotide site diversity under background selection, relative to the neutral value under background selection
Q	Ratio of substitution rate to nucleotide site diversity, both measured relative to their neutral values under background selection

https://doi.org/10.1017/S0016672300032365 Published online by Cambridge University Press

appropriate for most types of nucleotide site variants (Kimura, 1983; Gillespie, 1991).

The formulae will be applied to a number of special cases, to illustrate the expected effects of background selection on genetic variability at loci subject to weak selection, under a variety of assumptions concerning mating system and mode of inheritance. The effective population number is assumed throughout to be sufficiently large that the distribution of the numbers of deleterious mutations per individual is close to the infinite population equilibrium, for the loci generating background selection.

The method used by Charlesworth *et al.* (1993) for the neutral case will be employed here. This method assumes that, with complete linkage, selection on the background loci is sufficiently strong in relation to population size that there is a negligible probability that a deleterious mutation at a background locus can be fixed by drift over the time taken for a weakly selected or neutral variant to become lost or fixed. This implies that, in the absence of recombination, a weakly selected variant is doomed to loss if it arises initially in a gamete carrying a deleterious mutation at one or more of the background loci, or if it subsequently enters this class as a result of new mutations at background loci. Hence, only a fraction f_0 of new, weakly selected, variants have a non-zero probability of fixation.

Similarly, since the mean time to loss from the population of a given gamete carrying deleterious mutations at the background loci is very short, weakly selected or neutral variants carried in such gametes are rapidly lost, and contribute very little to the overall nucleotide site diversity of the population (Charlesworth *et al.* 1993; Hudson, 1994). The net contribution to such diversity from a weakly selected variant can therefore be approximated by considering its sojourn in the population as being effectively restricted to the gametes that are free of mutations at the background loci. This in turn implies that the effective population number N_e is replaced by $f_0 N_e$ in the standard formulae for allele frequency change under genetic drift (Charlesworth *et al.* 1993). While this assumption is clearly only an approximation, it has been shown to be very accurate in the neutral case when the size of the population is more than a few hundred individuals (Charlesworth *et al.* 1993; Hudson, 1994), and there is no reason why weak selection should affect its accuracy.

(ii) *Fixation probabilities and rates of gene substitution with background selection*

Let the frequency of a new, weakly selected, variant in the gene pool of the breeding adults in the generation in which it arises be p. The value of p depends on the mode of inheritance (autosomal or sex-linked), the sex of the individual in which the mutation arises, the breeding sex ratio of the population, and the breeding system of the population (random-mating, asexual, or self-fertilizing). Further details are given in section (i) of the Appendix.

Since variants that arise in gametes carrying deleterious alleles at background loci are lost from the population with a probability of one (section 2(i) above), we need only consider the variant further if it is carried initially in a gamete free of deleterious mutations. Conditioning on this, the initial frequency of the new variant is

$$p_0 = \frac{p}{f_0}. \tag{1}$$

The corresponding conditional probability of fixation can be obtained from the standard diffusion equation formula (Kimura, 1962):

$$u(p_0) = A \int_0^{p_0} G(x)\,\mathrm{d}x, \tag{2a}$$

where

$$G(x) = \exp -2 \int_0^x \frac{M_{\delta y}(y)}{V_{\delta y}(y)}\,\mathrm{d}y. \tag{2b}$$

$M_{\delta y}(y)$ and $V_{\delta y}(y) = y(1-y)/2f_0 N_e$ are the mean and variance of the change in allele frequency at frequency y, respectively, N_e is the effective population number for the specified mating system in the absence of background selection, and

$$A = 1 \Big/ \int_0^1 G(x)\,\mathrm{d}x. \tag{2c}$$

From the considerations of section 2(i), the net fixation probability of the variant is

$$u = f_0\, u(p_0). \tag{3}$$

If the mutation is neutral, we have $u(p_0) = p_0$. Equations (1) and (3) imply that in this case $u = p$, i.e. it is the same as in the absence of background selection, as expected from the results of Birky & Walsh (1988).

Fixation probabilities obtained from these formulae can be used to obtain rates of gene substitution, under the assumption that each variant that is fixed in evolution arises by mutation as a unique copy (Kimura & Ohta, 1971, pp. 11–12). This requires specification of mutation rate for the sites in question, since the substitution rate per site, K, is equal to the rate of input per site of mutations into the population, multiplied by their probability of fixation. Given the evidence that mutation rates may depend on sex, at least in mammals (Miyata *et al.* 1987; Charlesworth, 1993; Crow, 1993), it is desirable in general to allow for the possibility of sex differences in mutation rates. Let the mutation rate for the female germline be v_f and that for the male germline be v_m. General expressions

https://doi.org/10.1017/S0016672300032365 Published online by Cambridge University Press

for the substitution rates are derived in section (i) of the Appendix.

It is often convenient to express these as ratios of the corresponding values for neutral mutations. Equations (2) of the Appendix imply that background selection does not affect the neutral substitution rates, as expected intuitively. Consistent with the results of Miyata *et al.* (1987), these rates for autosomal and sex-linked loci (with male heterogamety) are respectively

$$K_{An} = \tfrac{1}{2}(v_f + v_m) \tag{4a}$$

$$K_{Xn} = \tfrac{1}{3}(2v_f + v_m). \tag{4b}$$

(iii) *Nucleotide site diversity*

Under statistical equilibrium between mutation and drift in the infinite sites models (Kimura, 1969, 1971), ergodicity implies that the equilibrium nucleotide site diversity π is proportional to the sum H of the diversity measure $2x(1-x)$ over all allele frequencies x experienced by a variant in the interval between its origination by mutation and loss or fixation. The equilibrium value of π is obtained by the same procedure that was used above to obtain rates of substitution from fixation probability: H is multiplied by the number of new mutations that enter the population each generation (e.g. Ewens, 1979, p. 239). By the argument of section 2(i), only mutations that arise in gametes free of deleterious mutations contribute significantly to H.

For a variant with conditional initial frequency p_0, H with background selection is thus approximated by the standard formula for a variant allele in a population of effective population size $f_0 N_e$. By equation (4.24) of Ewens (1979), the expected time spent by such a variant in the frequency range x to $x + dx (x \geqslant p_0)$ is given by

$$t(x, p_0) = \frac{2u(p_0)\int_x^1 G(y)\,dy}{V_{\delta x}(x)\,G(x)}, \tag{5}$$

where the quantities on the right-hand side are defined by equations (2).

Weighting the diversity $2x(1-x)$ by $t(x, p_0)$ and integrating over all values of x (Kimura, 1969, 1971; Ewens, 1979, p. 239), we obtain

$$H \sim 8BN_e f_0 u(p_0), \tag{6a}$$

where

$$B = \int_0^1 G(x)^{-1} \int_x^1 G(y)\,dy\,dx. \tag{6b}$$

This formula can be combined with the expression for fixation probabilities and the rates of input of mutations to yield formulae for the equilibrium diversity levels (Appendix, equations [A 3]). In the case of neutrality, these give the following results for autosomal and sex-linked variants, respectively

$$\pi_{An} \approx 2f_0 N_e(v_f + v_m) \tag{7a}$$

$$\pi_{Xn} \approx \tfrac{4}{3} f_0 N_e(2v_f + v_m). \tag{7b}$$

In the absence of sex differences in mutation rates, equation (7a) is identical with equation (3) of Charlesworth *et al.* (1993).

3. Approximate formulae for weak selection

Except for some special cases, the general equations derived above and in section (i) of the Appendix can only be evaluated numerically. Some useful insights can be obtained for the case when selection is weak, so that higher-order terms in the product of the selection coefficient and $f_0 N_e$ can be neglected. This may well be appropriate for many molecular variants (Kimura, 1983; Gillespie, 1991). Unless there is a strongly non-linear dependence of the quantities of interest on the product of the selection coefficient and $f_0 N_e$, the results derived below should give a useful indication of the nature of the effects of genetic system and background selection in more general cases.

(i) *Autosomal inheritance with random mating*

Assume initially that selection acts equally on the two sexes, and that the population is random-mating. Let the relative fitness of the genotypes AA, Aa and aa be $1+\sigma$, $1+\theta\sigma$ and 1, respectively. A is the mutant allele under consideration, σ is the selection coefficient on the mutant homozygote, such that $\sigma < 0$ implies that the mutation is deleterious when homozygous compared with the wild-type homozygote, and $\sigma > 0$ implies that it is advantageous. θ is the coefficient of dominance. For $0 \leqslant \theta < 0{\cdot}5$, the mutant allele is recessive or partly recessive; $\theta = 0{\cdot}5$ corresponds to the case of intermediate dominance; $0{\cdot}5 < \theta \leqslant 1$ corresponds to a dominant or partly dominant mutation. The cases of $\theta < 0$, $\sigma < 0$ and $\theta > 1$, $\sigma > 0$ correspond to overdominance (heterozygote advantage).

For random-mating populations, standard theory (e.g. Ewens, 1979, p. 138) implies that

$$\frac{M_{\delta x}}{V_{\delta x}} = 2f_0 S\{x(1-2\theta)+\theta\} \tag{8}$$

where $S = N_e\sigma$.

Approximate expressions for K and π are obtained by substituting from equation (8) into the general equations derived above and in the Appendix, and neglecting second- and higher-order terms in $f_0 S$ (Charlesworth, Coyne & Barton, 1987). Division by the appropriate neutral values yields the rate of substitution relative to the neutral value and the equilibrium genetic diversity relative to the neutral value without background selection. These are

https://doi.org/10.1017/S0016672300032365 Published online by Cambridge University Press

Table 2. *Weak selection approximations to rates of substitution and nucleotide site diversity (relative to neutral values without background selection), and to the ratio of substitution rate to diversity (each measured relative to its neutral value with background selection)*

Genetic system	$\tilde{K}$	$\tilde{\pi}$	Q
Autosomal locus (selection on both sexes; random mating)			
General	$1+\frac{2}{3}f_0 S(1+\theta)$	$f_0(1+\frac{2}{3}f_0 S\theta)$	$1+\frac{2}{3}f_0 S$
No sexual selection	$1+\frac{2}{3}f_0 N\sigma(1+\theta)$	$f_0(1+\frac{2}{3}f_0 N\sigma\theta)$	$1+\frac{2}{3}f_0 N\sigma$
Intense sexual selection	$1+\frac{8}{3}f_0 N_m\sigma(1+\theta)$	$f_0(1+\frac{8}{3}f_0 N_m\sigma\theta)$	$1+\frac{8}{3}f_0 N_m\sigma$
Autosomal locus (selection on one sex; random mating)			
General case	$1+\frac{1}{3}f_0 S(1+\theta)$	$f_0(1+\frac{1}{3}f_0 S\theta)$	$1+\frac{1}{3}f_0 S$
No sexual selection	$1+\frac{1}{3}f_0 N\sigma(1+\theta)$	$f_0(1+\frac{1}{3}f_0 N\sigma\theta)$	$1+\frac{1}{3}f_0 N\sigma$
Intense sexual selection	$1+\frac{4}{3}f_0 N_m\sigma(1+\theta)$	$f_0(1+\frac{4}{3}f_0 N_m\sigma\theta)$	$1+\frac{4}{3}f_0 N_m\sigma$
Selfing population	$1+2f_0 S$	$f_0(1+\frac{2}{3}f_0 S)$	$1+\frac{4}{3}f_0 S$
Asexual population	$1+2f_0 S\theta$	$f_0(1+\frac{4}{3}f_0 S\theta)$	$1+\frac{2}{3}f_0 S$
Sex-linked locus (selection on both sexes; random mating)			
General case	$1+\frac{2}{9}f_0 S(5+2\theta)$	$f_0(1+\frac{2}{9}f_0 S[1+2\theta])$	$1+\frac{8}{9}f_0 S$
No sexual selection	$1+\frac{1}{6}f_0 N\sigma(5+2\theta)$	$f_0(1+\frac{1}{6}f_0 N[1+2\theta])$	$1+\frac{2}{3}f_0 N\sigma$
Intense sexual selection	$1+f_0 N_m\sigma(5+2\theta)$	$f_0(1+f_0 N_m\sigma[1+2\theta])$	$1+4f_0 N_m\sigma$
Sex-linked locus (selection on heterogametic sex; random mating)			
General case	$1+\frac{2}{3}f_0 S$	$f_0(1+\frac{2}{9}f_0 S)$	$1+\frac{4}{9}f_0 S$
No sexual selection	$1+\frac{1}{2}f_0 N\sigma$	$f_0(1+\frac{1}{6}f_0 N\sigma)$	$1+\frac{1}{3}f_0 N\sigma$
Intense sexual selection	$1+3f_0 N_m\sigma$	$f_0(1+f_0 N_m\sigma)$	$1+2f_0 N_m\sigma$
Sex-linked locus (selection on homogametic sex; random mating)			
General case	$1+\frac{4}{9}f_0 S(1+\theta)$	$f_0(1+\frac{4}{9}f_0 S\theta)$	$1+\frac{4}{9}f_0 S$
No sexual selection	$1+\frac{1}{3}f_0 N\sigma(1+\theta)$	$f_0(1+\frac{1}{3}f_0 N\sigma\theta)$	$1+\frac{1}{3}f_0 N\sigma$
Intense sexual selection	$1+2f_0 N_m\sigma(1+\theta)$	$f_0(1+2f_0 N_m\sigma\theta)$	$1+2f_0 N_m\sigma$

denoted by $\tilde{K}$ and $\tilde{\pi}$, respectively. In addition, it is useful to consider the ratio of the rate of evolution to the level of genetic diversity (when both are measured relative to their neutral values with background selection), since this has been used to test for selection on DNA and protein sequences (McDonald & Kreitman, 1991; Sawyer & Hartl, 1992). For a region subject to background selection, this ratio is given by $Q = f_0\tilde{K}/\tilde{\pi}$.

The general approximate expressions for $\tilde{K}$, $\tilde{\pi}$, and Q for the autosomal case, with random mating and equal selection on the two sexes, are shown in the first row of Table 2. The second and third rows respectively show the values when there is no sexual selection and a 1:1 primary sex ratio, so that the distribution of completed family size is Poisson with $N_e = N$, and when there is intense sexual selection, so that $N_m \ll N_f$, and $N_e \approx 4N_m$ (Wright, 1969, pp. 213–214).

As expected, the rate of substitution of deleterious alleles is accelerated when f_0 is reduced by background selection, and the rate of substitution of favourable alleles is reduced. The effects are greatest for dominant or overdominant mutations and least for recessives ($\theta = 0$). As might be expected intuitively, weakly selected recessive mutations behave like neutral mutations as far as their contribution to diversity is concerned. Not surprisingly, heterozygote advantage increases diversity over neutral expectation. The effect of background selection is to bring the measure of diversity relative to the neutral value with background selection, $\pi^* = \tilde{\pi}/f_0$, closer to one. As expected, the rate/diversity ratio Q is increased above unity by selection in favour of the mutant alleles, and decreased below unity by selection against them. Reductions in f_0 bring Q closer to unity. Perhaps unexpectedly, Q is independent of the dominance coefficient.

The case when selection acts on only one of the two sexes can be dealt with by treating this as equivalent to the above case with a selection coefficient of $0{\cdot}5\sigma$. The resulting formulae are shown in rows 4–6 of Table 2.

(ii) *Selfing and asexual populations*

These results can also be applied to the cases of a completely selfing population and an asexual population, with the appropriate changes in parameters (formulae for f_0 in these cases are given by Charlesworth *et al.* 1993). Here, there is obviously no distinction of sex. With selfing, the right-hand side of equation (8) is replaced by $2f_0 N_e\sigma$, where N_e is one-half the value for a random-mating population of equivalent size with equal numbers of breeding males and females (Pollak, 1987). The selection term here is equivalent to the random-mating case with a dominance coefficient of one-half and a selection coefficient of 2σ, so that these parameter values can be substituted into the relevant equations in order to obtain the corresponding expressions for a selfing population. The results are shown in the seventh line of Table 2. Note that $S = 0{\cdot}5N\sigma$ when there is a Poisson distribution of family size and a 1:1 sex ratio, so that

the formulae for this case are the same as for the random-mating case with no sexual selection and with intermediate dominance.

With diploid asexuality, the descendant copies of a mutation are confined to one of the two haploid genomes of each individual in the line of descent from the individual who carried the original mutation, and a mutation is considered to be fixed if all individuals in the population come to be heterozygous for it. The right-hand side of equation (8) is thus replaced by $2f_0 N_e \theta\sigma$, where N_e is again one-half the value for a random-mating population of equivalent size. The formulae for this case are therefore equivalent to the random-mating case with a dominance coefficient of one-half, and selection coefficient of $2\theta\sigma$. They are shown in row 8 of Table 2.

(iii) *Sex-linkage with random mating*

The case of a sex-linked locus with selection acting equally on hemizygous males and homozygous females can be considered similarly. With weak selection, equation (8) is replaced by

$$\frac{M_{\delta x}}{V_{\delta x}} = \frac{4S}{3}\{x(1-2\theta)+\theta+\tfrac{1}{2}\}. \tag{9}$$

The general formula for effective population size in the sex-linked case is given by the following equation (Wright, 1969, pp. 213–214)

$$N_e = \frac{9N_m N_f}{4N_m + 2N_f}. \tag{10}$$

If there is no sexual selection, so that the sex ratio of breeding males and females is 1:1, N_e is three-quarters the value for that of an autosomal locus, i.e. $N_e = 0{\cdot}75N$. In this case, N_e is substantially less than the corresponding autosomal value. If there is strong sexual selection, so that $N_f \gg N_m$, N_e approaches $4{\cdot}5N_m$, slightly larger than the corresponding autosomal value (see 3(i) above).

The results for the sex-linked case with random mating are shown in the lower part of Table 2, using the same method as above. With selection on both sexes and no sexual selection, there is a larger magnitude of the deviation of $\tilde{K}$ from the neutral value than for the corresponding autosomal case with the same f_0 value, if the new mutation is recessive or partly recessive ($\theta < 0{\cdot}5$), and a smaller deviation for partly dominant alleles. The same applies to genetic diversity. With intense sexual selection, there is always a larger deviation for $\tilde{K}$ than for the equivalent autosomal case, provided that $\theta < 3{\cdot}5$ (which applies to all cases of directional selection, for which $\theta \leqslant 1$). There is a larger deviation for $\tilde{\pi}$ provided that $\theta < 1{\cdot}5$. With no sexual selection, Q takes the same value as for the autosomal case. With intense sexual selection, there is a 1·5-fold deviation of Q from the neutral value compared with the equivalent autosomal case.

Similar conclusions apply to cases when selection is sex-limited (cf. Charlesworth *et al.* 1987). Only the case of a species with dosage compensation in the heterogametic sex (taken here to be male, for brevity of description) will be considered here. If selection is male-limited, the expression for $M_{\delta x}$ in the sex-linked case is equivalent to that for the standard autosomal equation with a selection coefficient of $2\sigma/3$ and a dominance coefficient of one-half. Comparing this with the result for the autosomal case with male-limited selection, the magnitudes of the deviations of $\tilde{K}$ and $\tilde{\pi}$ from the neutral value are greater for sex-linked loci with partly recessive mutations, and less for partly dominant ones, just as with selection on both sexes. Q is always the same for sex-linked and autosomal loci. With intense sexual selection, magnitudes of the deviations of $\tilde{K}$ and $\tilde{\pi}$ from the neutral value are greater for sex-linked loci with $\theta < 1{\cdot}25$ and $\theta < 0{\cdot}75$ respectively. The deviation of Q from its neutral value is 1·5-fold larger in magnitude than for the equivalent autosomal case.

If selection is limited to the homogametic sex, it is easy to see that the selection term for the sex-linked case is equivalent to that for the autosomal locus without sex-limitation, with the same dominance coefficient and with selection coefficient $2\sigma/3$. It is also easy to see that, with no sexual selection, the values of S are the same for the sex-linked case and the corresponding sex-limited autosomal case, so for the same value of f_0, $\tilde{K}$, $\tilde{\pi}$ and Q are the same as for the equivalent autosomal case. With intense sexual selection, the magnitudes of their deviations from the neutral value are 1·5 times that for the autosomal case.

4. Arbitrary selection intensities

(i) *Intermediate dominance*

Explicit formulae for the relative rate of substitution and diversity for the case of random mating and selection on both sexes, with arbitrary values of S, can only be obtained easily for the case of intermediate dominance ($\theta = 0{\cdot}5$). For an autosomal locus with $\sigma \ll 1$, and assuming that $|S|p_0$ is small, the results of Kimura (1962, 1969) give

$$\tilde{K} = \frac{2f_0 S}{(1-e^{-2f_0 S})} \tag{11}$$

$$\tilde{\pi} = \frac{(2f_0 S + e^{-2f_0 S} - 1)}{S(1-e^{-2f_0 S})} \tag{12}$$

$$Q = \frac{2f_0^2 S^2}{(2f_0 S + e^{-2f_0 S} - 1)}. \tag{13}$$

Similar formulae apply to the cases of completely self-fertilizing or asexual populations, with the appropriate changes in the definition of S (see section 3[ii]).

The dependence of substitution rate and diversity on the direction and strength of selection for the case of intermediate dominance have been discussed by

https://doi.org/10.1017/S0016672300032365 Published online by Cambridge University Press

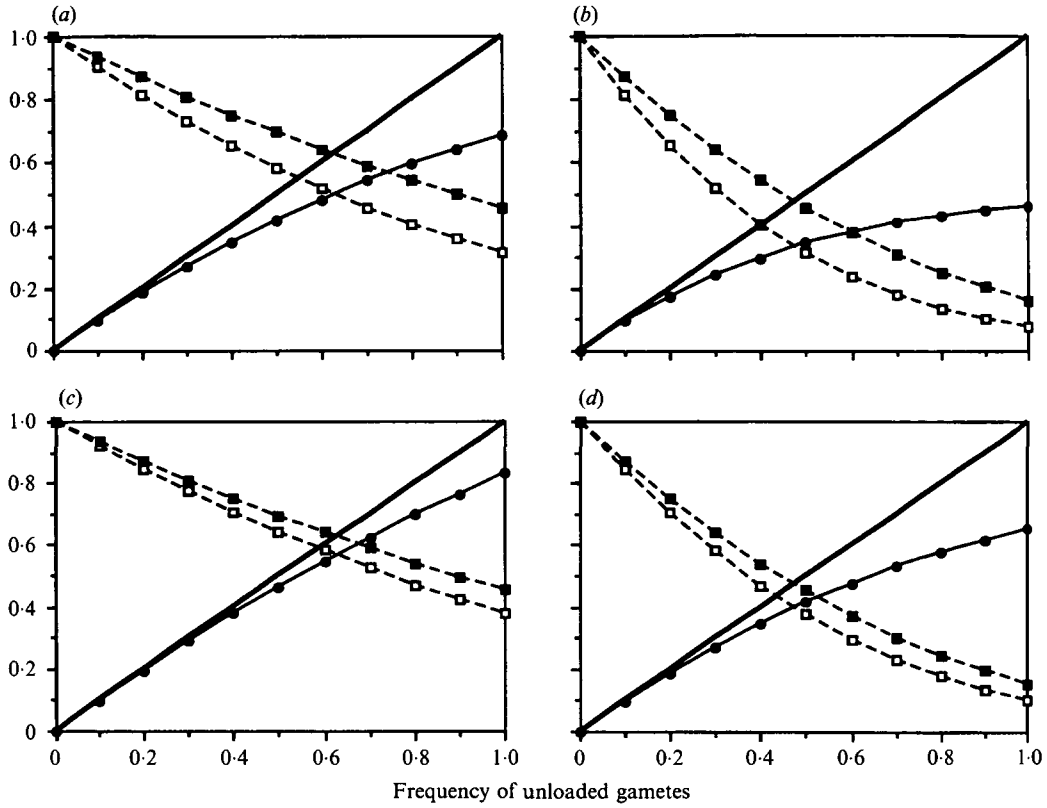

Fig. 1. The effect of background selection on slightly deleterious alleles at an autosomal locus in a random-mating population. A 1:1 sex ratio and equal male and female mutation rates are assumed. The abscissa is the frequency of deleterious-mutation free (unloaded) gametes (f_0) under background selection. The curves show the equilibrium genetic diversity relative to the neutral value with no background selection (●), the rate of substitution of new mutations relative to the neutral value (□), and the ratio of this rate to the genetic diversity relative to the neutral value with background selection (■). ——, genetic diversity for neutral alleles relative to the value without background selection. (*a*) is the case with $S = -1$ and $\theta = 0{\cdot}5$, (*b*) is for $S = -2$ and $\theta = 0{\cdot}5$, (*c*) for $S = -1$ and $\theta = 0{\cdot}2$, and (*d*) for $S = -2$ and $\theta = 0{\cdot}2$. The breeding population consists of 100000 individuals.

Kimura (1983, pp. 44–45). It is evident from the above formulae that background selection reduces the effect of selection on the locus of interest, with respect to both variables, as expected intuitively. The relative fixation probability $\tilde{K}$ depends exponentially on $f_0 S$; for deleterious alleles, it quickly tends to zero as $f_0|S|$ increases above unity. It approaches $2f_0 S$ for strongly selected advantageous alleles. A large reduction in f_0 can have a considerable effect on the rate of evolution, if $|S|$ is large. The relative diversity $\tilde{\pi}$ is less sensitive to $f_0 S$. For deleterious alleles, it tends to $-1/S$ as $f_0|S|$ increases; for advantageous alleles it tends to $2f_0$, i.e. twice the neutral value (cf. Kimura, 1983, p. 239). The rate/diversity ratio Q is a strictly increasing function of $f_0 S$ (see section (ii) of the Appendix for a proof of this statement). Q tends to zero with increasing $f_0|S|$ for deleterious alleles, and to $f_0 S$ for advantageous alleles.

Similar results can be obtained for sex-linked loci. We have

$$\tilde{K} = \frac{8f_0 S}{3(1-e^{-\frac{8}{3}f_0 S})}. \tag{14a}$$

With no sexual selection, this is identical with the value for the corresponding autosomal case, consistent with the finding for small $f_0 S$. With intense sexual selection, it becomes

$$\tilde{K} = \frac{32N_m f_0 \sigma}{3(1-e^{\frac{32}{3}f_0 N_m \sigma})}. \tag{14b}$$

This implies a stronger effect of selection than in the corresponding autosomal case, again as expected from the finding for small $f_0 S$.

We also have

$$\tilde{\pi} = \frac{3(\frac{8}{3}f_0 S + 1 - e^{-\frac{8}{3}f_0 S})}{4S(1-e^{-\frac{8}{3}f_0 S})}. \tag{15}$$

It is easily verified that this expression is identical to that for the autosomal case when there is no sexual selection. With intense sexual selection, $S = 4N_m \sigma$, giving a larger effect of selection than in the equivalent autosomal case.

In addition,

$$Q = \frac{32f_0^2 S^2}{9(\frac{8}{3}f_0 S + 1 - e^{-\frac{8}{3}f_0 S})}. \tag{16}$$

Again, this is identical to the autosomal case when there is no sexual selection. There is a stronger effect

https://doi.org/10.1017/S0016672300032365 Published online by Cambridge University Press

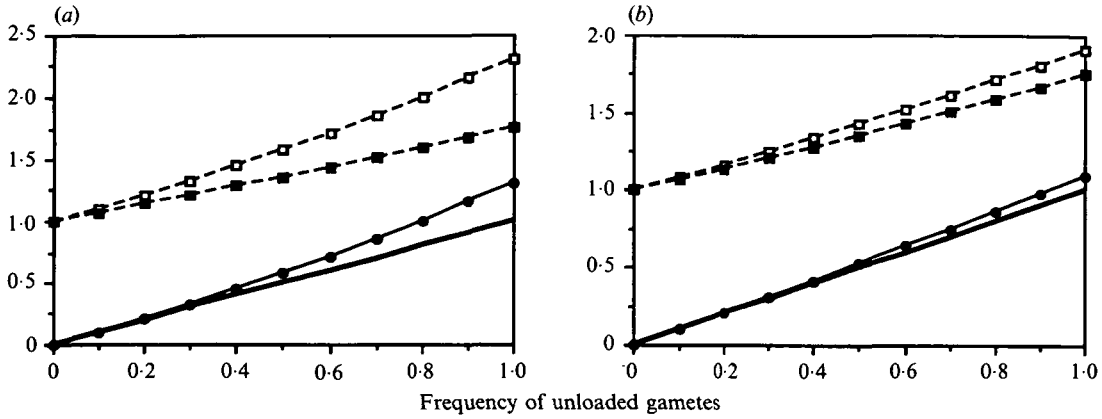

Fig. 2. The effect of background selection on slightly advantageous alleles at an autosomal locus in a random-mating population. (*a*) is the case with $S = 1$ and $\theta = 0{\cdot}5$, (*b*) is for $S = 1$ and $\theta = 0{\cdot}2$. Otherwise, the curves are as in Fig. 1.

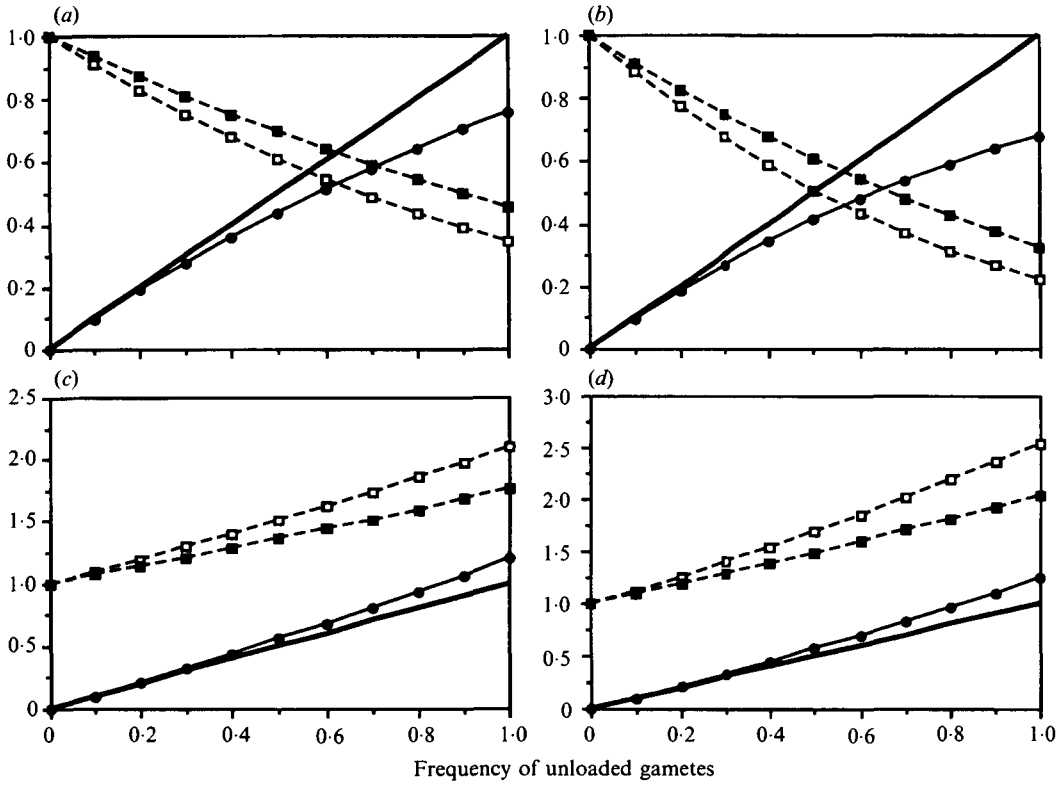

Fig. 3. The effect of background selection on a sex-linked locus in a random-mating population. $\theta = 0{\cdot}2$ in each case. (*a*) and (*c*) are cases of no sexual selection, with $S = -0{\cdot}75$ and $S = 0{\cdot}75$ respectively. (*b*) and (*d*) are cases of fairly strong sexual selection, with $S = -1{\cdot}07$ and $S = 1{\cdot}07$ respectively. Autosomal loci exposed to the sexual selection regimes corresponding to these sex-linked cases give results shown for the $|S| = 1$ cases in Figs 1 and 2. Otherwise, the curves are as in Fig. 1.

of $f_0 S$ with extreme sexual selection. As before, Q is a strictly increasing function of $f_0 S$.

(ii) *Arbitrary dominance*

Results for arbitrary dominance coefficients can be obtained by numerical integration of the relevant equations. Fig. 1 shows the dependence of $\tilde{K}$, $\tilde{\pi}$ and Q on f_0 for the case of deleterious alleles with $S = -1$ (left-hand panels *a* and *c*) and -2 (right-hand panels *b* and *d*), and dominance coefficients of 0·5 (upper panels *a* and *b*) and 0·2 (lower panels *c* and *d*). Equal mutation rates in males and females, and a 1:1 breeding sex ratio are assumed. As expected from the weak-selection approximation discussed above, differences in the dominance coefficient have only a small effect on $\tilde{K}$, and a negligible effect on Q. Considerably more genetic diversity is maintained for moderate or high values of f_0 when there is partial recessivity of the deleterious fitness effects, again in agreement with the weak selection approximation. For low values of f_0 (i.e. when there is a strong effect of background

https://doi.org/10.1017/S0016672300032365 Published online by Cambridge University Press

selection), there is little effect of dominance or of S, since the diversity values for the cases with selection converge on the neutral value with background selection (shown by the heavy straight lines in the figures). There is a much more marked effect of background selection on $\tilde{K}$ and Q when $|S|$ is large than when it is small, e.g. with intermediate dominance, the ratio of Q values for $f_0 = 1$ and $f_0 = 0{\cdot}1$ is 0·49 when $S = -1$, and 0·18 when $S = -2$. This reflects the exponential dependence of $\tilde{K}$ and Q on $f_0 S$ in equations (11) and (13). Background selection can thus have a very large effect in accelerating the rate of evolution by the fixation of deleterious alleles, when these have fitness effects which are as large as or larger than the reciprocal of the effective population size.

Fig. 2 shows comparable results for the case of autosomal sites subject to positive selection, with $f_0 S = 1$ and dominance coefficients of 0·5 and 0·2. Here, the effect of selection is to increase both the rate of evolution and the equilibrium diversity over neutral expectation. Partial recessivity of the favourable fitness effects of the mutations reduces both the rate of evolution and the level of diversity, but Q is almost unaffected by dominance, as expected from the weak selection approximation. Background selection can again have a substantial effect on the rate of evolution, this time by reducing it towards the neutral value.

Fig. 3 displays results for a sex-linked locus with equal selection on the two sexes, for sites subject to partly recessive deleterious mutations (upper panels *a* and *b*) and favourable mutations (lower panels *c* and *d*), for the cases of no sexual selection with $N_e = 0{\cdot}75N$ (left-hand panels *a* and *c*) and fairly strong sexual selection (5% of breeding individuals are males, so that $N_e = 0{\cdot}204N$ (right-hand panels *b* and *d*). The parameter values have been chosen to be comparable with the autosomal cases of Figs. 1 and 2 with $|S| = 1$. The results are qualitatively similar to those for the autosomal case, with somewhat more marked effects of the parameter values, especially for the case of intense sexual selection. For example, the equilibrium diversity (relative to the neutral value) for deleterious mutations with sexual selection is 79% of the corresponding autosomal value when $f_0 = 1$, and the Q value is 66%. The differences between sex-linked and autosomal cases diminish with higher intensities of background selection, e.g. with $f_0 = 0{\cdot}1$, the sex-linked values for diversity and Q are 98 and 97% of the corresponding autosomal values. Correspondingly, there is a somewhat stronger effect on sex-linked loci of the same change in f_0.

5. Discussion

The results described above have a number of implications for understanding patterns of molecular evolution and variation. These will be considered in turn.

(i) *Genetic diversity*

The results derived above show that, as expected intuitively, the equilibrium level of genetic diversity is reduced by background selection under all types of selection acting directly on variants at the sites concerned. If recombination is absent or infrequent and background selection is sufficiently strong, genetic diversities for both neutral and selected sites will be similar and low. We would therefore expect genetic diversity to be quite similar for silent and replacement sites, in regions of reduced recombination. Surveys of molecular variation in *Drosophila* have provided the main basis for the conclusion that genetic variability is severely reduced in such regions, but these are mainly based on restriction mapping studies which usually do not distinguish between silent and replacement site variation (Begun & Aquadro, 1992; Aquadro, Begun & Kindahl, 1994). The limited sequencing studies available to date indicate that both silent and replacement site polymorphisms are infrequent in loci in regions of reduced recombination (Berry, Ajioka & Kreitman, 1991; M. Wayne, in preparation; H. Hilton, R. M. Kliman & J. Hey, in preparation), but it seems premature to conclude that levels of diversity are always similar for both classes of variant.

This conclusion assumes that the intensity of selection at the sites under consideration is much weaker than that operating on the deleterious alleles which are the source of the background selection, so that variants which arise in chromosomes carrying such deleterious alleles are destined to be rapidly lost from the population, even if they confer a selective advantage. Variants subject to sufficiently strong balancing selection in relation to the degree of linkage to surrounding deleterious mutations may, of course, be maintained close to their equilibrium, even in regions subject to significant background selection. This suggests that, if a polymorphism is detected in a region of restricted recombination, where variation at most sites is depleted, serious consideration should be given to the possibility that it is maintained by balancing selection. According to Aquadro *et al.* (1994), allozyme variation is not reduced significantly in regions of reduced recombination in *D. melanogaster*, suggesting that it may be maintained by balancing selection.

(ii) *Rates of molecular evolution*

As shown in section 4(ii), background selection can have a very large effect in accelerating the rate of evolution by the fixation of deleterious alleles, when these have fitness effects which are as large as or larger than the reciprocal of the effective population size. The rate of substitution of weakly selected advantageous alleles is reduced (see also Birky & Walsh, 1988). One might therefore expect a substantially

higher rate of substitution of replacement site variants in regions of restricted recombination under the slightly deleterious alleles model of molecular evolution (Ohta, 1974, 1992), and a lower rate under the rival model of the adaptive evolution of protein sequences (Gillespie, 1991), relative to their values in other regions. Hitch-hiking by favourable mutations will have similar effects (Birky & Walsh, 1988).

Comparisons of rates of amino acid substitution in different regions of the *Drosophila* genome with different recombinational regimes might thus provide a means of distinguishing between these theories. The sequencing data of Hilton *et al.* (in preparation) for five loci in the *melanogaster* species group of *Drosophila* show that there is more inter-species replacement divergence per nucleotide site for two loci (*ci*^{D} and *asense*) that are located in regions of reduced recombination than for the loci subject to normal levels of recombination, suggesting that this possibility may be realized in this case. There is also an indication from restriction mapping surveys of both *D. ananassae* (Stephan & Mitchell, 1992) and *D. melanogaster* (Begun & Aquadro, 1993; Aquadro *et al.* 1994) that there are greater inter-population differences within species for nucleotide-site variants in regions of restricted recombination than in regions where recombination is normal. The effect of background selection in reducing effective population size should promote divergence between populations subject to limited gene flow, even at neutral sites, since $f_0 N_e m$ must replace $N_e m$ in the standard formulae for divergence between populations for variants that survive stochastic loss. The observed differentiation might, however, simply be the consequence of hitch-hiking events that occurred relatively recently in one of the populations surveyed (Stephan & Mitchell, 1992), rather than an effect of background selection on enhancing divergence by drift.

Unfortunately, differences among different proteins in the degree to which sequence evolution is subject to selective constraints makes this test difficult to carry out when only small numbers of sequences in regions of both normal and restricted recombination are available for comparing related species or populations, since the effects of recombination are confounded with functional differences among proteins. When a large bank of gene sequences in sets of populations and related species becomes available, this problem should be less severe, since functional differences should presumably average out.

An alternative method of examining this question is provided by the fact that there are good reasons for believing that the level of codon bias in bacteria, yeast, and *Drosophila* is affected by the chance fixation of silent alterations to third codons, which cause the adoption of sub-optimal codons and which are therefore likely to be slightly deleterious (e.g. Bulmer, 1991; Li & Graur, 1991). One might therefore expect the degree of selective control of codon usage, as evidenced by codon bias, to be less marked in regions of restricted recombination. There is evidence that this is the case in *Drosophila* (Kliman & Hey, 1993). Similar processes may be responsible for the relatively high GC contents of regions of the mammalian and yeast genomes with unusually high levels of recombination (Charlesworth, 1994).

The effect of background selection on the rate of molecular evolution also implies that loci which are in regions of low recombination in some species but not in others, may vary in their rate of molecular evolution. Changes in the recombinational environment of a locus can occur by chromosome rearrangements (e.g. O'Brien, 1993), especially when rearrangements involve breaks close to the centromeric heterochromatin, where recombination is often suppressed. Thus, observed variation in the rate of protein evolution for the same locus among different lineages, which has been used as an argument in favour of adaptive evolution (Gillespie, 1991), could simply reflect variation in the recombinational environment of the locus. This suggests that it is important to consider the recombinational environment of loci, and the breeding system of the species, when conducting comparative studies of rates of molecular evolution. For example, moving a gene from the mid-section of a *D. melanogaster* autosome to the centromeric region could cause an 80% reduction in f_0 (Charlesworth *et al.* 1993). This would result in a factor of 8·6 increase in the rate of evolution for deleterious amino acid replacements with intermediate dominance and an $|S|$ of 2. Even within the *melanogaster* subgroup of *Drosophila* there have been extensive internal rearrangements of gene locations within chromosome arms (Ashburner, 1989, Chap. 36).

(iii) *Tests of neutrality versus selection*

McDonald & Kreitman (1991) have proposed the use of the ratio of the number of nucleotide substitutions separating two species to the level of nucleotide site variability within species, in a test of the neutral theory of molecular evolution. On the pure neutral theory, this ratio should be the same for silent and replacement sites; on a model of adaptive protein sequence evolution, it should be higher for replacement sites than for silent sites, assuming that the latter are close to neutrality. This follows from the fact that diversity is less sensitive to selection than the rate of evolution, i.e. the rate/diversity ratio for selected versus neutral sites, Q, exceeds unity for positively selected sites (see Table 2). McDonald & Kreitman (1991) found a significant difference in the rate/diversity ratio between silent and replacement sites for the *Adh* gene in the *D. melanogaster* subgroup, suggesting that adaptive evolution had occurred at some amino acid sites. Since the rate of replacement site evolution is generally lower than that for silent

sites (Kimura, 1983), indicating negative selection against most amino acid changes, the McDonald–Kreitman test for adaptive evolution with a null hypothesis of $Q = 1$ is in fact rather conservative; because $Q < 1$ for deleterious mutations, the rate/diversity ratio for replacement changes should be less than one in the absence of positive selection.

The considerable effect of f_0 on Q (see Figs. 1–3) implies that comparisons of the rate/diversity ratio for replacement and silent sites among regions of the genome with different levels of recombination should indicate a general tendency towards values close to unity in regions of reduced recombination, with either higher or lower values for amino acid replacements in regions with normal levels, depending on whether or not negative or positive selection for amino acid site variants is more prevalent. It is therefore of importance to take into account recombinational environment and breeding system when interpreting the results of the McDonald/Kreitman test.

The models discussed so far assume that the sites under consideration are all subject to the same selection regime. This is, of course, quite unrealistic. Even if we confine ourselves to the problem of variation and evolution of amino acid changes, there is likely to be a wide distribution of fitness effects, from strongly deleterious to strongly advantageous (Kimura, 1983; Gillespie, 1991). An extreme alternative to the model of fixed selection coefficients is to imagine two classes of replacement substitutions: those subject to strong negative selection, with the product of selection coefficient and N_e equal to S_0, such that $f_0 S_0 \ll -1$, and those subject to strong positive selection, with the product of selection coefficient and N_e equal to S_1, such that $f_0 S_1 \gg 1$. Let the proportion of amino acid sites that fall into these two categories be ϕ_0 and ϕ_1, respectively. Using the results of section 4(i) for an autosomal locus with intermediate dominance, we obtain

$$\tilde{K} \approx -2\phi_0 f_0 S_0 e^{2f_0 S_0} + 2\phi_1 f_0 S_1 \quad (17a)$$

$$\tilde{\pi} \approx -\frac{\phi_0}{S_0} + 2\phi_1 f_0 \quad (17b)$$

$$Q \approx 2\phi_0 f_0^2 S_0^2 e^{2f_0 S_0} + \phi_1 f_0 S_1. \quad (17c)$$

Equation (17c) implies that $Q - 1 < \phi_1 f_0 S_1$. The rate/diversity ratio minus one thus provides an underestimate of a composite of the frequency of positively selected sites, the frequency of unloaded gametes, and the product of selection coefficient and effective population size. Using the data on *Adh* sequences in three species of the *D. melanogaster* subgroup in table 2 of McDonald & Kreitman (1991), and taking the number of segregating sites within species as a crude estimate of diversity for the *Adh* gene, we have $Q = 7 \times 42/2 \times 17 = 8{\cdot}64$. In this case, f_0 is close to unity, since *Adh* is in a region with normal recombination rates (Lindsley & Zimm, 1992), so that $Q - 1$ provides a lower bound to the estimate of $\phi_1 S_1$. At least some amino acid site variants at the *Adh* locus must therefore be subject to positive selection, with selection coefficients several times greater than $1/N_e$, in agreement with the more elaborate analysis of Sawyer & Hartl (1992).

These considerations also bring out a potential bias in the test proposed by Kreitman & Aguadé (1986) and Hudson, Kreitman & Aguadé (1987) for selection on molecular polymorphism. This test uses a comparison of levels of intra- and inter-specific nucleotide site diversity in different regions of the genome. The test assumes that the only differences between regions are in the probabilities that mutations fall into one of the two categories of effectively neutral or deleterious, which implies that both intra-specific and inter-specific diversity are affected proportionately by differences in these probabilities.

But the existence of a range of values of selection coefficients, so that the fate of mutations at many sites is controlled jointly by drift and selection, may lead to lack of such proportionality and hence to a bias in the test. As already discussed, the standing genetic diversity contributed by slightly deleterious alleles during their sojourn in the populations is less sensitive to the magnitude of $f_0 S$ than is the rate at which they become fixed (Kimura, 1983, pp. 44–45). A particular region could thus have a high level of intra-specific diversity for replacement substitutions, relative to the degree of inter-specific divergence, simply because of lower selective constraints in that region. Conversely, a low relative degree of intra-specific diversity in a particular region need not reflect a recent hitch-hiking event (cf. Kreitman & Hudson, 1991). The potential importance of this bias remains to be investigated. It would not, however, produce an increase in variation at silent sites linked to a replacement site polymorphism, of the kind observed at the *Adh* locus (Kreitman & Hudson, 1991).

(iv) *Sex-linked loci*

The comparison of the sex-linked and autosomal cases indicates that, under the slightly deleterious alleles model, levels of genetic variability can be lower for sex-linked loci than for autosomal loci, even if variation is expressed relative to neutral expectation to correct for differences in effective population sizes between the sex-linked and autosomal cases. This assumes that deleterious mutations are recessive or partly recessive, as seems plausible for deleterious alleles (Crow & Simmons, 1983). These effects are large only if mutations are fairly recessive in their effects on fitness (Table 2, Figs. 1–3). The maximum effect is seen when there is intense sexual selection. Similarly, rates of evolution and Q values are likely to be lower for sex-linked than for autosomal loci under the slightly deleterious model. The converse is true for sites subject to positive selection, if favourable

mutations are partly recessive (Charlesworth *et al.* 1987).

Little information is currently available on rates of molecular evolution and levels of genetic diversity for X-linked versus autosomal loci. A recent survey of restriction-site variation in populations of *D. melanogaster* indicates that X-linked loci tend to be less variable than autosomal loci, even after correcting for differences in effective population size between sex-linked and autosomal loci by multiplying diversity levels for sex-linked loci by 4/3 (Aquadro *et al.* 1994). This may indicate some degree of selection against slightly deleterious effects of silent-site changes, consistent with the evidence on codon bias (Kliman & Hey, 1993). Alternatively, hitch-hiking events may be commoner on the X than on the autosomes, if favourable mutations are predominantly recessive or partly recessive in their effects on fitness, thus leading to a higher frequency of selective sweeps on the X compared with the autosomes (Aquadro *et al.* 1994).

The difference between X-linked and autosomal variation noted by Aquadro *et al.* (1994) may even be underestimated by the fact that the standard correction for sex-linkage assumes no sexual selection. But equations (7) imply that this is inappropriate when there is intense sexual selection. With a highly male-biased breeding sex ratio ($N_m \ll N_f$) and assuming no sex differences in mutation rate, we find that the neutral values of π are $16f_0 N_m v$ for the autosomal case and $18f_0 N_m v$ for the sex-linked case. In other words, neutral diversity in the case of sex-linked loci is slightly higher than that for autosomal loci with the same mutation rate, with intense sexual selection. Values for more moderate intensities of sexual selection will be intermediate.

In mammals, mutation rates in males are higher than in females (Miyata *et al.* 1987; Charlesworth, 1993; Crow, 1993). Since X chromosomes in species with male heterogamety spend less time in males than do autosomes, both rates of evolution and levels of diversity should be higher for the autosomes if mutation rates are higher in the male germ line. In *Drosophila* there is, however, no solid evidence for such a sex difference in mutation rate (Woodruff, Slatko & Thomson, 1983; Crow, 1993). Mutation rate differences between the sexes have little or no effect on measures of diversity and evolutionary rates when these are expressed relative to the neutral values, as has been the focus of attention here.

(v) *Asexual and self-fertilizing species*

Very much smaller f_0 values are expected for asexual and predominantly self-fertilizing species than for the relatively small segments of the genome subject to restricted recombination in random-mating populations (Charlesworth *et al.* 1993), so that large differences in both evolutionary rates and levels of diversity from comparable random-mating populations are expected. In selfers, the effects of such small f_0 values on evolutionary rates are likely to be considerably greater than the fact that the homozygous effects of mutations are the target of selection, so that in general the rate of evolution of even partly recessive slightly deleterious mutations will probably be accelerated relative to neutral evolution in highly selfing species. This contrasts with the conclusion of Charlesworth (1992), who only considered loci evolving in the absence of background selection. The converse is true for advantageous mutations. Related taxa with different breeding systems may therefore exhibit considerable differences in rates of molecular evolution, as well as in diversity levels. Correlations between these rates and levels of self-fertilization in hermaphroditic groups could provide a test of the slightly deleterious mutation versus adaptive mutation models. These effects may involve the plastid and mitochondrial genomes, as well as the nuclear genome, since there is effectively complete linkage between the nuclear and maternally transmitted genomes in a completely selfing or asexual population (Charlesworth *et al.* 1993).

Joel Peck (in preparation) has pointed out that the low values of f_0 which are likely to prevail in asexual taxa imply that such groups may have much lower rates of substitution of new favourable mutations than sexual taxa, and that this may contribute to the geologically short duration of many asexual taxa (Maynard Smith, 1978; Bell, 1982).

(vi) *Hitch-hiking versus background selection*

It is intuitively obvious that the hitch-hiking effects of linked favourable mutations that sweep to fixation will have qualitatively similar effects on both the level of genetic diversity and rate of evolution. Birky & Walsh (1988) found this to be true for the rate of evolution due to the fixation of both advantageous and deleterious alleles. There is an extensive literature on the effects of hitch-hiking on neutral genetic variation (Maynard Smith & Haigh, 1974; Ohta & Kimura, 1975; Thomson, 1977; Birky & Walsh, 1988; Kaplan, Hudson & Langley, 1989; Stephan, Wiehe & Lenz, 1992; Wiehe & Stephan, 1993), which shows that such variation can be drastically reduced, just as in the case of background selection (Charlesworth *et al.* 1993).

There may, however, be some quantitatively different patterns in the effects of hitch-hiking and background selection on neutral variation, which would permit them to be distinguished empirically (Charlesworth *et al.* 1993; Aquadro *et al.* 1994). For example, a complete loss of all genetic variability due to the fixation of a favourable mutation that remains completely linked to the region under study is likely to be followed by a long recovery period in which variability restored by mutation is skewed towards rare variants, compared with the neutral equilibrium.

Selective sweeps may thus give significantly negative values of Tajima's (1989) D statistic. With background selection, a negative D is unlikely to be found in a sample of realistic size (Charlesworth *et al.* 1993, in preparation; Hudson, 1994). In general, it seems that *Drosophila* loci in regions of reduced recombination fail to show negative D values, yet their reduction in variation is too great to be entirely caused by background selection (Aguadé, Meyers, Long & Langley, 1994; Aquadro *et al.* 1994). Another possible approach would be to compare the degrees of reduction in variation associated with reduced recombination proximal to the centromere for X-linked versus autosomal loci in *Drosophila*. Since f_0 is expected to be smaller for this region for the autosomal arms than for the X (Charlesworth *et al.* 1993), background selection predicts a smaller reduction for the X chromosome than for the autosomes. If no such pattern, or the opposite relation, were to be observed, the hitch-hiking model would be supported, since a higher rate of selective sweeps can occur on the X chromosome under suitable conditions (Aquadro *et al.* 1994). At present, neither model on its own seems to be capable of fully explaining the data.

It is unclear whether such differences between hitch-hiking and background selection will cause detectable differences in the relative frequencies of selected and neutral alleles, when regions of normal and restricted recombination are compared. For example, a complete loss of variability due to a selective sweep will cause both slightly deleterious and neutral alleles to remain at low frequencies during the period of recovery from the sweep. There is thus likely to be a similar pattern of low but similar levels of diversity for both classes of variant, as with background selection. This question requires further quantitative study.

I thank Deborah Charlesworth, Charles Langley, and Allen Orr for discussions that led to the preparation of this paper. I also thank Charles Aquadro, Jody Hey, Charles Langley, Joel Peck, and Marta Wayne for providing me with their unpublished results. Charles Aquadro, Deborah Charlesworth, Richard Hudson, Charles Langley, Wolfgang Stephan, and Bruce Walsh made helpful comments on the manuscript. I am especially grateful to Richard Hudson for pointing out an error in the original version.

Appendix

(i) *General formulae for rates of substitution and nucleotide site diversities*

In the case of a random-mating population, assume that the numbers of breeding adult females and males in each generation are N_f and N_m, respectively. The total number of breeding adults is $N = N_f + N_m$. With autosomal inheritance, males and females make equal contributions to the gene pool produced by these individuals. With sex-linkage and male heterogamety, females contribute two-thirds of the gene pool and males one-third. Using the argument of section 2(ii), the effective initial frequencies of a new mutation present as a single copy in a female in the autosomal and sex-linked cases, respectively, are

$$p_{Af} = \frac{1}{4f_0 N_f} \qquad \text{(A 1a)}$$

$$p_{Xf} = \frac{1}{3f_0 N_f}. \qquad \text{(A 1b)}$$

The corresponding expressions for males, p_{Am} and p_{Xm}, are obtained by substituting N_m for N_f in these expressions.

In the autosomal case, variants present in males and females have equal probabilities of having arisen in male and female parents. The expected numbers of new mutations per site that appear in breeding females and males are thus $N_f(v_f + v_m)$ and $N_m(v_f + v_m)$ respectively. Substituting the initial frequencies from equation (A 1a) into equations (2a) and (3), we obtain the substitution rate for the autosomal case as

$$K_A = (v_f + v_m) f_0 \{N_f u(p_{Af}) + N_m u(p_{Am})\}. \qquad \text{(A 2a)}$$

The expression for the sex-linked case is slightly different, since mutations carried in males must have arisen in females:

$$K_X = (v_f + v_m) N_f f_0 u(p_{Xf}) + v_f N_m f_0 u(p_{Xm}). \qquad \text{(A 2b)}$$

In the neutral case, the fixation probabilities are equal to the initial frequencies multiplied by f_0, yielding equations (4) of the text.

A similar procedure can be used to obtain expressions for the expected nucleotide site diversities, by substituting from equations (1) into equation (6a) and multiplying the appropriate H values by the numbers of new mutations that arise per generation, discounted by f_0 to allow for the fact that background selection eliminates a fraction f_0 of new variants. The following formulae are obtained for the diversities for autosomal and sex-linked loci:

$$\pi_A \approx 8 f_0^2 B N_e (v_f + v_m) \{N_f u(p_{Af}) + N_m u(p_{Am})\} \qquad \text{(A 3a)}$$

$$\pi_X \approx 8 f_0^2 B N_e \{(v_f + v_m) N_f u(p_{Xf}) + v_f N_m u(p_{Xm})\}. \qquad \text{(A 3b)}$$

In the neutral case, $B = 0{\cdot}5$; using the neutral fixation probabilities we obtain equations (7) of the text.

(ii) *Proof that Q is an increasing function of* $f_0 S$

From equation (13), we can write

$$Q = \frac{x^2}{2(x + e^{-x} - 1)}, \qquad \text{(A 4)}$$

where $x = 2f_0 S$.

Using primes to denote derivatives, we have

$$Q'(x) \propto x\, g(x), \qquad \text{(A 5)}$$

where

$$g(x) = (x-2) + (x+2)e^{-x}.$$

Using a Taylor expansion, we find that $g(0) = 0$, $g(0+\epsilon) > 0$, and $g(0-\epsilon) < 0$, if ϵ is taken sufficiently small. Using primes to denote derivatives, we also have

$$g'(x) = 1-(x+1)e^{-x} \quad \text{(A 6)}$$

It is easily seen that $g'(x) > 0$ for $x > 0$, so that $g(x) > 0$ for all $x > 0$. Similarly, $g''(x) < 0$ for $x < 0$, and so $g'(x) < 0$ for all $x < 0$. Hence, $g(x) < 0$ for all $x < 0$.

This establishes that $Q'(x) > 0$ for all x.

References

Aguadé, M., Meyers, W., Long, A. D. & Langley, C. H. (1994). Reduced DNA sequence polymorphism in the *su(s)* and *su(w^a)* regions of *Drosophila melanogaster* as revealed by SSCP and stratified DNA sequencing. *Proceedings of the National Academy of Sciences, USA* (in press).

Aquadro, C. F., Begun, D. J. & Kindahl, E. C. (1994). Selection, recombination, and DNA polymorphism in *Drosophila*. In *Non-neutral Evolution: Theories and Molecular Data* (ed. G. B. Golding), in press. London: Chapman & Hall.

Ashburner, M. (1989). *Drosophila. A Laboratory Handbook*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Begun, D. J. & Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rate in *D. melanogaster*. *Nature* **356**, 519–520.

Begun, D. J. & Aquadro, C. F. (1993). African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550.

Bell, G. (1982). *The Masterpiece of Nature*. London: Croom-Helm.

Berry, A. J., Ajioka, J. W. & Kreitman, M. (1991). Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129**, 1111–1117.

Birky, 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.

Bulmer, M. G. (1991). The selection–mutation–drift theory of synonymous codon usage. *Genetics* **129**, 897–907.

Charlesworth, B. (1992). Evolutionary rates in partially self-fertilizing species. *American Naturalist* **140**, 126–148.

Charlesworth, B. (1993). More mutations in males. *Current Biology* **3**, 466–467.

Charlesworth, B. (1994). Patterns in the genome. *Current Biology* **4**, 182–184.

Charlesworth, B., Coyne, J. A. & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *American Naturalist* **130**, 113–146.

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

Crow, J. F. (1993). How much do we know about spontaneous human mutation rates? *Environmental and Molecular Mutagenesis* **21**, 122–129.

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

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

Gillespie, J. H. (1991). *The Causes of Molecular Evolution*. Oxford: Oxford University Press.

Houle, D., Hoffmaster, D. K., Assimacopoulos, S. & Charlesworth, B. (1992). The genomic mutation rate for fitness in *Drosophila*. *Nature* **359**, 58–60.

Hudson, R. R. (1994). Gene trees with background selection. In *Non-neutral Evolution: Theories and Molecular Data* (ed. G. B. Golding), in press. London: Chapman & Hall.

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Ikemura, I. & Wada, K. (1991). Evident diversity of codon usage patterns of human genes with respect to chromosome banding patterns and chromosome numbers; relation between nucleotide sequence data and cytogenetic data. *Nucleic Acids Research* **19**, 4333–4339.

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

Kimura, M. (1962). On the probability of fixation of a mutant gene in a population. *Genetics* **47**, 713–719.

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 foundations of population genetics at the molecular level. *Theoretical Population Biology* **2**, 174–208.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

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

Kreitman, M. (1991). Detecting selection at the level of DNA. In *Evolution at the Molecular Level* (ed. R. K. Selander, A. G. Clark and T. S. Whittam), pp. 202–221. Sunderland, MA: Sinauer.

Kreitman, M. & Aguadé, M. (1986). Excess polymorphism at the alcohol dehydrogenase locus in *Drosophila melanogaster*. *Genetics* **114**, 93–110.

Kreitman, M. & Hudson, R. R. (1991). Inferring the evolutionary history of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**, 565–582.

Li, W.-H. & Graur, D. (1991). *Fundamentals of Molecular Evolution*. Sunderland, MA: Sinauer.

Lindsley, D. L. & Zimm, G. G. (1992). *The Genome of Drosophila melanogaster*. San Diego, CA: Academic Press.

Maynard Smith, J. (1978). *The Evolution of Sex*. Cambridge: Cambridge University Press.

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

McDonald, J. H. & Kreitman, M. (1991). Accelerated protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654.

Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. (1987). Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harbor Symposia on Quantitative Biology* **52**, 863–867.

Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. (1972). Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335–355.

O'Brien, S. J. (1993). The genomics generation. *Current Biology* **3**, 395–397.

Ohta, T. (1974). Mutational pressure as main cause of molecular evolution. *Nature* **252**, 351–354.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* **23**, 263–286.

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.

Pollak, E. (1987). On the theory of partially inbreeding populations. I. Partial selfing. *Genetics* **117**, 353–360.

Sawyer, S. A. & Hartl, D. L. (1992). Population genetics of polymorphism and divergence. *Genetics* **132**, 1161–1176.

Stephan, W. & Mitchell, S. J. (1992). Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae*. *Genetics* **132**, 1039–1045.

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.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis. *Genetics* **123**, 585–595.

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 from *Drosophila melanogaster*. *Molecular Biology and Evolution* **10**, 842–854.

Woodruff, R. C., Slatko, B. E. & Thompson, J. N. (1983). Factors affecting mutation rates in natural populations. In *The Genetics and Biology of Drosophila, Vol. 3c* (ed. M. Ashburner, H. L. Carson and J. N. Thompson), pp. 37–124. London: Academic Press.

Wright, S. (1969). *Evolution and the Genetics of Populations*. Chicago, IL: University of Chicago Press.