The Role of Background Selection in Shaping Patterns of Molecular Evolution and Variation: Evidence from Variability on the *Drosophila X* Chromosome

Brian Charlesworth[1]
Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom

ABSTRACT In the putatively ancestral population of *Drosophila melanogaster*, the ratio of silent DNA sequence diversity for *X*-linked loci to that for autosomal loci is approximately one, instead of the expected "null" value of 3/4. One possible explanation is that background selection (the hitchhiking effect of deleterious mutations) is more effective on the autosomes than on the *X* chromosome, because of the lack of crossing over in male *Drosophila*. The expected effects of background selection on neutral variability at sites in the middle of an *X* chromosome or an autosomal arm were calculated for different models of chromosome organization and methods of approximation, using current estimates of the deleterious mutation rate and distributions of the fitness effects of deleterious mutations. The robustness of the results to different distributions of fitness effects, dominance coefficients, mutation rates, mapping functions, and chromosome size was investigated. The predicted ratio of *X*-linked to autosomal variability is relatively insensitive to these variables, except for the mutation rate and map length. Provided that the deleterious mutation rate per genome is sufficiently large, it seems likely that background selection can account for the observed *X* to autosome ratio of variability in the ancestral population of *D. melanogaster*. The fact that this ratio is much less than one in *D. pseudoobscura* is also consistent with the model's predictions, since this species has a high rate of crossing over. The results suggest that background selection may play a major role in shaping patterns of molecular evolution and variation.

MEAN silent site DNA sequence diversities in the putatively ancestral East African populations of *Drosophila melanogaster* seem to be approximately the same for the *X* chromosome (*X*) and autosomes (*A*) (Andolfatto 2001; Hutter *et al.* 2007; Singh *et al.* 2007), despite the fact that the "null" expectation for the ratio of the effective population sizes (N_e) of *X* and *A* is 3/4 for the case of a 1:1 sex ratio and purely random variation in offspring number in both sexes (Wright 1931). There is little evidence for an *X vs. A* difference in mutation rate at silent sites in *Drosophila*, after possible differences in the intensity of selection and mutational biases on silent sites for *X* and *A* are taken into account (Bauer and Aquadro 1997; Hutter *et al.* 2007; Keightley *et al.* 2009; Zeng and Charlesworth 2010; Haddrill *et al.* 2011) (but see Bachtrog 2008). This observation therefore suggests an equality of N_e values for *X* and *A*, since neutral diversity under the infinite sites model is equal to the product of $4N_e$ and the neutral mutation rate per site (Kimura 1971), where N_e is defined as one-half of the expected coalescent time for a pair of alleles at a given locus (Charlesworth and Charlesworth 2010, p. 217).

This equality of N_e values for *X* and *A* could reflect a highly female-biased sex ratio and/or a very high variance in male reproductive success (Hedrick 2007; Hutter *et al.* 2007; Ellegren 2009; Vicoso and Charlesworth 2009a), both of which reduce the effective population size of males relative to females. This reduction would have a smaller effect on *X* than on *A*, since the *Drosophila X* spends two-thirds of its time in females and only one-third of its time in males, whereas an autosome spends half of its time in each sex. But this difference between *X* and *A* also affects the population-effective rate of recombination for *X vs. A*, which controls the rate of breakdown of linkage disequilibrium. *Drosophila* males lack recombinational exchange between homologous

doi: 10.1534/genetics.111.138073
Manuscript received December 23, 2011; accepted for publication February 26, 2012
[1] Address for correspondence: Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Ashworth Laboratories, King's Bldgs., W. Mains Rd., Edinburgh EH9 3JT, United Kingdom. E-mail: brian.charlesworth@ed.ac.uk

chromosomes (Ashburner *et al.* 2005); the population-effective recombination rate for a given rate of recombination r in females between two loci is therefore $0.5r$ for A and $0.667r$ for X (Charlesworth and Charlesworth 2010, p. 381). This suggests that hitchhiking effects may have less influence on variability at typical X loci compared with A loci, given similar selection intensities for X and A mutations, consistent with the observation that X and A loci with similar population-effective recombination rates appear to have relative levels of silent site variability that are close to the null expectation (Vicoso and Charlesworth 2009b); a contrary effect can be produced by recurrent selective sweeps of partially recessive, positively selected mutations (Aquadro *et al.* 1994; Betancourt *et al.* 2004; Ellegren 2009).

An alternative explanation was proposed by Hutter *et al.* (2007), who suggested that a recent population expansion in the Zimbabwe population of *D. melanogaster* has differentially affected X-linked and autosomal variability. However, recent analyses of synonymous site variability in this population have cast doubt on the reality of such an expansion, when selection on codon usage is taken into account (Zeng and Charlesworth 2009, 2010). Similarly, the fact that the X/A diversity ratio is close to one for an African *D. simulans* population, which lacks inversions, suggests that a reduction in autosomal diversity caused by hitchhiking effects of autosomal inversion polymorphisms in *D. melanogaster* (Andolfatto 2001; Singh *et al.* 2007) cannot be a general explanation for this effect.

The purpose of this article is to investigate whether the process of background selection, the hitchhiking of neutral or nearly neutral variability by linked deleterious mutations (Charlesworth *et al.* 1993; Charlesworth 2012), can account for these apparently equal X and autosomal N_e values or at least contribute to an X/A ratio that differs substantially from 3/4, as was suggested earlier by Aquadro *et al.* (1994). To do this, a model of the effect of background selection (BGS) on variability across a large, normally recombining region of a chromosome is needed. A previous investigation of this problem in *D. melanogaster* by Charlesworth (1996) used phenotypic estimates of the strength of selection against deleterious mutations and the overall mutation rate to deleterious alleles, which have now been superseded by estimates based on DNA sequence data (Loewe and Charlesworth 2006; Loewe *et al.* 2006; Haag-Liautard *et al.* 2007; Keightley and Eyre-Walker 2007; Eyre-Walker and Keightley 2009; Keightley *et al.* 2009; Schneider *et al.* 2011; Wilson *et al.* 2011). In addition, the study by Charlesworth (1996) assumed that sites subject to purifying selection were spread uniformly along the chromosome, whereas in reality they are clustered into coding sequences and blocks of functional noncoding sequences (Misra *et al.* 2002).

The present study attempts to remedy these deficiencies, with specific reference to the effect of BGS on the X/A ratio of effective population sizes or neutral diversities under the infinites sites model, both of which are proportional to the expected coalescent time for a pair of alleles. Equations 4–9 of Nordborg *et al.* (1996) provide formulas for the effect of BGS on the coalescent time at a given nucleotide site. Comparisons with the results of computer simulations have shown that these formulas are accurate for the case of a single chromosome, provided that the strength of selection is sufficiently strong in comparison with the effect of genetic drift that the frequencies of deleterious mutant alleles can be treated as though they are in deterministic equilibrium (Nordborg *et al.* 1996). The model developed here takes into account the fact that some noncoding sequences in *Drosophila* (both intergenic and intronic) are strongly conserved, so that deleterious mutations affecting them probably have similar selective effects to nonsynonymous mutations, whereas another large class of noncoding sequences is subject to weaker or no selective constraints (Haddrill *et al.* 2005; Halligan and Keightley 2006; Casillas *et al.* 2007; Sella *et al.* 2009).

This article is concerned with the effect of BGS over a whole chromosome or chromosome arm. To simplify the analysis, I assume that we are dealing with a site located in the middle of a *Drosophila* chromosome arm and that recombination rates per unit physical distance are uniform across the arm. This will somewhat underestimate the overall effect of BGS on variability in a *Drosophila* population, since recombination rates are lower at the telomeres and centromeres than in the middle of an arm (Ashburner *et al.* 2005), but in practice most loci used in resequencing studies in *Drosophila* come from regions with high levels of recombination, which constitute the majority of the genome (Charlesworth 1996). In addition, the effect of lower variability in low recombination regions is partly counteracted by the weaker effect of BGS at the ends of a chromosome arm, caused by the lack of adjacent genes compared with the sites in the middle of a chromosome (Nordborg *et al.* 1996), so that consideration of the properties of the middle of a chromosome arm should provide a reasonably good picture of the typical level of variability in the presence of BGS in *Drosophila*, except at the extreme ends of a chromosome or chromosome arm.

Theory and Methods

Model assumptions: types of sequence and their selection coefficients

Predictions are developed using several different levels of approximations, with the overall goal of evaluating the sensitivity of the predicted X/A ratio of neutral diversity to the assumptions of the models. In line with the facts described above, three classes of selected site are modeled: strongly selected nonsynonymous sites, strongly selected noncoding sites, and weakly selected noncoding sites. Probability distributions of the fitness effects of newly arising deleterious mutations are assumed, such that the probability density of selection coefficient s against a homozygous mutation at a given nucleotide site is $\phi_k(s)$ for the kth class of selected site, where $k = 1$ for nonsynonymous sites, 2 for strongly selected noncoding sites, and 3 for weakly selected noncoding sites. For reasons given below, both nonsynonymous and strongly selected noncoding sequences are assumed here to share the same distribution, so that $\phi_1(s) = \phi_2(s)$, although this assumption is easy to relax. Sex differences in selection coefficients are also ignored in most of the results presented below; again, reasons are given later as to why this is unlikely to be important.

The different classes of sites are organized into n_g coding sequences of length l_g bp, separated by "intergenic" sequences, which are divided into n_s strongly selected noncoding sequences of length l_{is} and $n_s + 1$ weakly selected noncoding sequences of length l_{iw} (see Figure 1). Each coding sequence is flanked by weakly selected noncoding sequences, except at the telomere and centromere, where any nongenic DNA beyond the last coding sequences at the ends of the arm is ignored. The total length of noncoding sequence separating a pair of coding sequences is thus $l_i = n_s l_{is} + (n_s + 1)l_{iw}$. By increasing the number of coding sequences and decreasing their length accordingly, a chromosome arm with genes that includes long introns, containing a mixture of strongly and weakly selected sites, can easily be modeled. For simplicity, the account below mostly refers to noncoding sequences as intergenic.

The focal neutral site for which the effect of BGS is to be calculated is assumed to be located in the center of the chromosome arm, in the middle of a weakly selected intergenic sequence, so that the distance to the nearest coding sequence is $0.5(l_i - 1)$, assuming l_i to be odd. To accommodate this assumption, n_g and n_s are assumed to be even, and l_{iw} is assumed to be odd.

Deleterious mutations involved in BGS effects are assumed to be under such strong selection that they are kept at low frequencies. For nonrecessive autosomal mutations, this means that the equilibrium frequency of a deleterious mutation at a given site in an infinitely large randomly mating population is determined by the ratio of the mutation rate, u, to a deleterious variant at a site and the heterozygous selection coefficient against the deleterious variant, t_s (Charlesworth and Charlesworth 2010, p. 161). In a large finite population, the mean frequency of deleterious mutations over a collection of sites with the same mutation and selection parameters is close to this equilibrium, provided that recombination is sufficiently frequent (Nordborg *et al.* 1996). With dominance coefficient h and selection coefficient s against homozygotes, the effective selection coefficient against the mutation is thus $t_s = hs$ ($h > 0$). For X-linked mutations with equal selection on the two sexes, the corresponding effective selection coefficient is given approximately by $t_s = (2h + 1)s/3$ (Charlesworth and Charlesworth 2010, p. 98); corresponding formulas can be obtained for the case of sex-specific effects on fitness.

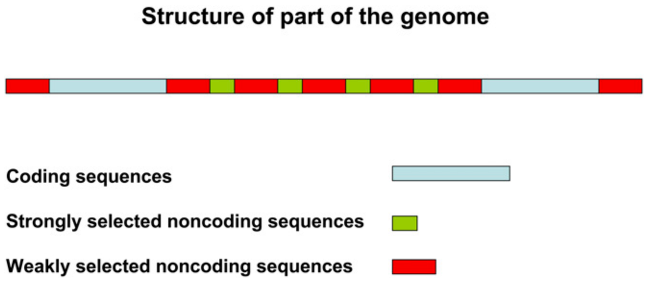
Structure of part of the genome
Coding sequences
Strongly selected noncoding sequences
Weakly selected noncoding sequences

Figure 1 The organization of noncoding sequences around two coding sequences (blue), into blocks of strongly selected sequence (green) and weakly selected sequences (red).

Model assumptions: mutation rates

For most of the results presented here, only a single chromosome arm is considered, on the assumption that the dissipation of the effects of BGS with recombinational distance is sufficiently strong that sites on one chromosome arm have little effect on another (evidence to support this assumption is presented in *Results and Discussion*). We define the diploid deleterious mutation rate for the kth class of selected site on a chromosome arm as U_k, where U_k is equal to the sum of $2u$ over all sites subject to purifying selection of this type of site on the chromosome arm in question. The standard numerical values for the U_k used in most of the calculations below were arrived at in the following way. The net deleterious diploid mutation rate for *D. melanogaster* was estimated by Haag-Liautard *et al.* (2007) to be 1.2 mutations per generation. The corresponding deleterious mutation rate, U_D, for a typical chromosome arm contributing ~20% of the euchromatic genome is equal to 0.24. With an average of 2800 genes per arm and an average total coding sequence length per gene of 1500 bp (Misra *et al.* 2002), the assumption that 70% of coding sites are nonsynonymous (Loewe and Charlesworth 2007) gives ~2.94 Mb of sites capable of generating nonsynonymous mutations out of a total of 4.2 Mb coding sequence. In a chromosome arm of 20 Mb, this leaves ~15.8 Mb of noncoding sequences. The majority of these are subject to some level of selective constraint (Halligan and Keightley 2006); ~25% are strongly conserved sequences with an average length of ~40 bp (Casillas *et al.* 2007).

The remaining 75% of noncoding sequences are here assumed to be under weak purifying selection. In the absence of introns, a total length $l_i = 5659$ bp of sequence between a pair of coding sequences is allowed, which is divided into $n_s = 36$ strongly selected sequences of length 39 bp and $n_s + 1 = 37$ weakly selected noncoding sequences of length 115 bp, organized as described above (these numbers are reduced proportionately if n_g is increased and l_g is reduced, to allow for long introns separating exons within genes). This gives a total of $(n_g - 1)n_s l_s = 2799 \times 36 \times 39 = 3.93$ Mb of strongly selected noncoding sequence and $(n_g - 1)(n_s + 1)l_w = 2799 \times 37 \times 115 = 11.91$ Mb of weakly selected noncoding sequence. Together with the nonsynonymous sites, this gives a total of 18.78 Mb of sequence that is potentially under significant selection [selection on codon usage acting at synonymous sites is ignored here, since it is too weak to have significant BGS effects (Zeng and Charlesworth 2009; Zeng 2010)].

To assess the contributions of these types of sequence to the deleterious mutation rate, we also need estimates of their respective levels of selective constraint, *i.e.*, the fraction of mutations in each class that are sufficiently deleterious to ensure their elimination from the population (Halligan and Keightley 2006). Table 2 of Casillas *et al.* (2007) suggests that the divergence per site between *D. melanogaster* and *D. simulans* for strongly conserved noncoding sequences is comparable to that for nonsynonymous sites (see Table 1 of Sella *et al.* 2009), and their fit of a gamma distribution to the distribution of selection coefficients against mutations in these sequences gave a similar value of the shape parameter a to published estimates for nonsynonymous mutations (Keightley and Eyre-Walker 2007; Eyre-Walker and Keightley 2009; Haddrill *et al.* 2010).

There is also evidence that similar values of the fraction α of fixed differences between species caused by positive selection apply to strongly selected noncoding and nonsynonymous mutations (Casillas *et al.* 2007; Sella *et al.* 2009); these need to be removed from estimates of between-species divergence before calculating selective constraint values, which apply only to sites subject to purifying selection. This component of the ratio of divergence at strongly selected sites relative to that for putatively neutral sites is taken here to be equal to 0.078, consistent with the observed ratio of nonsynonymous to synonymous divergence between *D. melanogaster* and *D. simulans* of ~0.13 and a somewhat conservative α-value of 0.6, yielding a constraint value for strongly selected sites of $c_s = 0.922$, for both class 1 and class 2 sites. For weakly constrained noncoding sites, a constraint value of $c_w = 0.572$ is used here, which is slightly higher than the value indicated by Table 2 of Casillas *et al.* (2007) and Table 1 of Sella *et al.* (2009), to accommodate a small fraction of positively selected mutations. If a shape parameter of the gamma distribution of 0.3 is assumed, consistent with the evidence just mentioned, then mean s values for strongly selected and weakly selected sites that are consistent with these constraint values can be calculated by the method described in the Appendix of Haddrill *et al.* (2010); assuming a dominance coefficient of 0.5, these are found to be $\bar{s}_1 = \bar{s}_2 = 2.5 \times 10^{-3}$ (strongly selected sites) and $a_3 = 0.3$, $\bar{s}_3 = 8 \times 10^{-6}$ (weakly selected sites), assuming $N_e = 10^6$.

Let the proportion of strongly selected sites among all sites potentially under significant selection be $x_s = x_{sc} + x_{sn}$, where x_{sc} and x_{sn} are the proportions of nonsynonymous sites and strongly selected noncoding sites among potentially selected sites; the proportion of weakly selected sites is $x_w = 1 - x_s$. From the way in which U_D was estimated (Haag-Liautard *et al.* 2007), the overall mutation rate for a chromosome arm contributed by these sequences is $U_A = U_D/(x_s\ c_s + x_w\ c_w)$. We have $x_{sc} = 2.94/18.78 = 0.157$, $x_{sn} = 3.93/18.78 = 0.210$, $x_s = 0.367$, and $x_w = 0.633$. Use of these numerical values gives $U_A = 0.24/(0.367 \times 0.922 + 0.633 \times 0.572) = 0.343$. We then obtain the following values of the deleterious mutation rates for each class: $U_1 = U_A x_{sc} c_s = 0.050$, $U_2 = U_A x_{sn} c_s = 0.066$, and $U_3 = U_A x_w c_w = 0.124$.

These values need to be reduced by removing mutations that fall below the threshold value for which the formulas for BGS used below are likely to be accurate (Nordborg *et al.* 1996), giving truncated mutation rates of U_{T_k} for each class of site. In the numerical results presented below, the distributions for both the strongly selected and the weakly selected sites were truncated at the lower end at $s_T = 5 \times 10^{-6}$, corresponding to an $N_e s$ value of 5 in a population of effective size 10^6. This procedure will lead to an underestimate of the effect of BGS, as the mutations that fall below this threshold will exert some effects, although not as large as predicted by the formulas used below (Zeng and Charlesworth 2011). Using the gamma distribution and the mutation rate parameters described above, the truncated deleterious mutation rates are $U_{T_1} = 0.044$, $U_{T_2} = 0.058$, and $U_{T_3} = 0.044$, giving a total truncated deleterious mutation rate $U_T = 0.146$. The truncated mutation rates for the nonsynonymous and strongly selected sites are only slightly lower than the untruncated values, whereas ~65% of the weakly selected noncoding sites are treated as neutral.

Exact model of BGS

For a focal neutral site, the expected effect of BGS caused by a given type of selected site is parameterized by the ratio of the coalescent time for this site to its "neutral" value in the absence of BGS (Hudson and Kaplan 1994, 1995; Nordborg *et al.* 1996), denoted here by B_k for the kth class of site. For simplicity, the mutation rate at a site in class k is assumed in the following analyses to be independent of s for these sites, but only mutations with an s above the truncation point s_T described above are included in the calculations. Because different organizations of sites apply to nonsynonymous, strongly selected noncoding sites and weakly selected noncoding sites, each of these must be considered separately.

For all sites included in a given class k, we have

$$B_k \approx \exp - u \int_{s_T}^{1} \sum_i \frac{t_s \phi_k(s) ds}{(t_s + r_i[1 - t_s])^2}, \qquad (1)$$

where u is the mean haploid mutation rate per base pair, and r_i is the frequency of recombination between the focal site and the ith site of class k (Nordborg *et al.* 1996).

We need to relate r_i for each site to the physical distance from the focal site to the ith site under selection. Let the total map length in females of the chromosome arm be M M, so that the map distance per base pair is $\rho = M/(n_g l_g + [n_g - 1] l_i)$. The simplest mapping function is a linear relation between map distance and recombination rate. Taking into account the lack of crossing over in male *Drosophila* (Ashburner *et al.* 2005), and averaging recombination rates across the two sexes as explained in the Introduction, this model yields population-effective recombination rates for a distance of d bp between the focal site and a given selected site of $r_{d_A} = 0.5\rho d$ and $r_{d_X} = 0.667\ \rho d$, for A and X sites, respectively. More generally, the recombination rate can be related to the map distance $z = \rho d$ by a mapping function that allows for the occurrence of double crossovers. Here, the "standard" mapping function of Charlesworth (1996) is used, which was shown by Cobbs (1978) to provide a good fit to *Drosophila* data. The population-effective recombination rates for map distance z are 0.25{1 – cos(2z)exp(–2z)} and 0.333{1 – cos (2z)exp(–2z)}, for the *A* and the *X*, respectively.

Given the values of the parameters described above, and the distribution of s, it is straightforward in principle to evaluate B_k for a given k by applying numerical integration over $\phi_k(s)$ to the summation in Equation 1. This requires specification of the distances of each selected site from the focal sites, as outlined in the *Appendix* for noncoding sites. The summation in Equation 1 proceeds along the chromosome arm in one direction, starting at the centrally located focal site, and the final result is doubled to estimate the sum for the whole arm.

For calculations involving nonsynonymous sites, all third coding positions are treated as neutral to accommodate synonymous sites and are skipped over when summing along a coding sequence. The mutational density u for these sites is $u = U_1/(1.4n_g l_g)$, since the total number of nonsynonymous sites on a chromosome arm is $\sim 0.7n_g l_g$. For strongly selected noncoding sites, we have $u = U_2/(2[n_g - 1]n_s l_s)$. The same distributions of s are used for nonsynonymous and strongly selected sites in the numerical results shown below. For weakly selected sites, $u = U_3/(2[n_g - 1][n_s + 1]l_w)$, which is substantially smaller than u for the strongly selected sites (reflecting the lower level of selective constraint in this case), and different parameters for the distribution of s are used.

The only difficulty with this procedure is that the large number of nucleotide sites ($\sim$20 million) on a typical *Drosophila* chromosome arm makes numerical integration over all sites very slow. For this reason, the summation formula over sites was used, averaging the contribution from each site over a grid of 1000 points taken over the range of a truncated gamma distribution, with s derived from a single-parameter gamma distribution with shape parameter a_k and mean $\bar{s}_k$ before truncation of values of $s < s_T$ (the mean after truncation is higher). To avoid inaccuracies of numerical integration with very small s, for values of $x < 0.001$ the analytical formula for the integral of $x^{a_k - 1}$ was used instead of x^{a_k-1}exp(–x) in the formula for the gamma distribution (where $x = s\ a_k/\bar{s}_k$). Integration over the remainder of the distribution was continued up to a value of $x = 10$, with constant increments of x on a logarithmic scale. To avoid selection coefficients greater than one, all values with $s > 1$ were reset to 1.

This approach yields the "model 1" results for the case of the standard mapping function and "model 2" results for the case of a linear map. The features of the different types of model are summarized in Table 1.

Approximations

Results can be obtained more rapidly using several different approximations to Equation 1. The simplest is that introduced by Hudson and Kaplan (1994) and Barton (1995), which assumes a linear map of length M M in females. Selected sites are distributed uniformly along it, with a population-effective map length for the chromosome arm of M_e (0.5M and 0.667M, for X and A, respectively), with M_e assumed to be much greater than any value of t_s drawn from the distribution. Replacing the summation in Equation 1 by integration along a continuum, these assumptions give $B_k \approx$ exp(–U_T/M_e) (see Equation 10 of Nordborg *et al.* 1996), where U_T is the sum of the truncated deleterious mutation rates U_{T_k} over all classes k. This yields the "model 3" results.

An approximation that should in principle be more accurate can be obtained as follows, retaining the assumption of a linear map. For nonsynonymous sites, the following procedure is followed. The physical distances from the focal site to the start and end of the jth coding sequence to its right or left are $\sim l_i(j - 0.5) + l_g(j - 1)$ and $l_i(j - 0.5) + l_g j$, respectively. Using the relations described above, these distances can be translated into population-effective recombination frequencies of r_{j1} and r_{j2}, respectively. Summation along the length of a coding sequence is replaced by integration, and a deleterious mutational density of $U_1/(2 \times 0.7 \times n_g l_g)$ per site is assumed, allowing as before for the fact that an average of 70% of coding sequence mutations are nonsynonymous. Using Equation 9 of Nordborg *et al.* (1996) with some rearrangement of terms (see *Appendix*, Equations A1 and A2), for a given value of t_s we obtain the following net contribution to the negative of the exponent in Equation 1 from the jth pair of coding sequences to the right and left of the focal site

$$E_{1j}(t_s) \approx \frac{U_1 t_s}{n_g\big(t_s + r_{j1}[1 - t_s]\big)\big(t_s + r_{j2}[1 - t_s]\big)}. \qquad (2)$$

The assumption that selected sites are uniformly distributed along a coding sequence is of course inaccurate because of the presence of synonymous sites. However, this is likely to cause only a minor error, since the mutational density per nonsynonymous site is $U_1/(1.4n_g l_g)$ rather than $U_1/(2n_g l_g)$, and the mean frequency of recombination between adjacent nonsynonymous sites is 1/(0.7) higher than assumed in this expression, if 30% of sites are neutral. The derivation given in the *Appendix* shows that these two effects cancel out in the final expression.

The overall exponent can then be obtained by integration over the truncated distribution of selection coefficients for nonsynonymous sites and summing over all j from 1 to $n_g/2$. This gives

$$B_1 \approx \exp - \int_{s_T}^{1} \sum_j E_{1j}(t_s)\phi_1(s)ds. \qquad (3)$$

The procedures for noncoding sites are similar, with summation over the integrals for each set of l_{is} strongly selected

Table 1 Models of background selection with selection on nonsynonymous, weakly selected, and strongly selected noncoding sites

Model 1	Standard mapping function with summation over all sites and integration over the distributions of selection coefficients
Model 2	Linear mapping function with summation over all sites and integration over the distributions of selection coefficients
Model 3	Approximations using integration over a continuum of sites (with a linear mapping function)
Model 4	Approximations using summation of the integrals over each cluster of sites with the same selection regime (with a linear mapping function) and integration over the distributions of selection coefficients
Model 5	Approximations using integration along the genome of the integrals over clusters of sites (with a linear mapping function), and first and second moments of the distributions of selection coefficients

and l_{iw} weakly selected sites, respectively (*Appendix*, Equations A4 and A10). Together with the results for the nonsynonymous sites, the expressions for B_2 and B_3 obtained in this way describe "model 4".

Apart from model 3, these calculations are all dependent on the properties of the distribution of s. An alternative approximation that avoids using the details of this distribution can be obtained by using the approach used for model 4, but replacing summation over coding sequence or blocks of noncoding sites by integration with respect to a continuous variable representing the index of the coding sequence or noncoding block in question. The calculation is further simplified by assuming that the distribution of s is such that $t_s \ll 1$ for most sites, so that the terms involving $1 - t_s$ in the above expressions can be replaced by 1.

For nonsynonymous sites, use of the above expressions for the distances from the focal site to the start and end of a coding sequence yields the following approximation to the sum over all j of $E_{1_j}(t_s)$,

$$E_1\ (t_s) \approx \frac{U_1 t_s}{n_g} \int_1^{(1/2)n_g} \times \frac{dx}{\{t_s + \tilde{\rho}(l_i[x - 1/2] + l_g[x-1])\}\{t_s + \tilde{\rho}(l_i[x - 1/2] + l_g x)\}}, \tag{4}$$

where $\tilde{\rho}$= 0.5ρ for A and 0.667ρ for X, and ρ is the gradient of the map distance in female meiosis with respect to the number of base pairs separating a pair of sites.

Elementary integration reduces this expression to

$$E_1(t_s) \approx \frac{U_1 t_s}{n_g \tilde{\rho}^2 l_g (l_i + l_g)} \ln\left\{\frac{(a' + (1/2)bn_g)(a + b)}{(a + (1/2)bn_g)(a' + b)}\right\}, \tag{5}$$

where $a = t_s - 0.5\tilde{\rho}\ l_i$, $a' = t_s - \tilde{\rho}\ (0.5l_i + l_g)$, and $b = \tilde{\rho}(l_i + l_g)$.

We have $(a + b)/(a' + b) = 1 + \tilde{\rho}\ l_g/(a' + b)$ and $(a + 0.5bn_g)/(a' + 0.5bn_g) = 1 + \tilde{\rho} l_g/(a' + 0.5bn_g)$, where $a' + b = t_s + 0.5\tilde{\rho}\ l_i$ and $a' + 0.5bn_g = t_s - \tilde{\rho}\ (0.5l_i + l_g) + 0.5\tilde{\rho}\ (l_i + l_g)n_g \approx t_s + 0.5\tilde{\rho}\ (l_i + l_g)n_g$. We can reasonably assume that $l_i \gg l_g$, given the typical length of an intergenic sequence compared with a coding sequence in *Drosophila* (Misra *et al.* 2002). The logarithmic expression in Equation 5 can then be well approximated by its leading term $\tilde{\rho} l_g\{(1/[t_s + \beta]) - (1/[t_s + \gamma])\}$, where $\beta = 0.5\tilde{\rho} l_i$ and $\gamma = 0.5\ \tilde{\rho}\ (l_i + l_g)n_g$.

An approximation to the expectation of $E_1(t_s)$ over the truncated distribution of s can then be obtained by representing $1/(t_s + \beta)$ and $1/(t_s + \gamma)$ by their Taylor series in the deviation of t_s from its mean, δt_s, and taking the expectation of the resulting expression for $E_1(t_s)$, ignoring terms of higher order than $(\delta t_s)^2$. This yields the approximation

$$B_1 \approx \exp\ -\frac{U_{1_T}}{n_g \tilde{\rho}(l_i + l_g)} \times \left\{\frac{1}{(\bar{t}_1 + \beta)}\left[\bar{t}_1 - \frac{\beta V_1}{(\bar{t}_1 + \beta)^2}\right] - \frac{1}{(\bar{t}_1 + \gamma)}\left[\bar{t}_1 - \frac{\gamma V_1}{(\bar{t}_1 + \gamma)^2}\right]\right\}, \tag{6}$$

where U_{1_T} is the deleterious mutation rate for nonsynonymous sites after truncation of the distribution of selection coefficients, and t_1 and V_1 are the mean and variance of t_s, respectively, taken over the truncated distribution of s for nonsynonymous sites.

The similar but more complex procedures for noncoding sites are described in the *Appendix*. Together with the results for nonsynonymous sites, these approximations for the B_k yield the "model 5" results. All these formulas were implemented in FORTRAN programs, which are available on request.

Results and Discussion

BGS on the X and A in D. melanogaster

Table 2 shows the results of calculations of the expected coalescent times under background selection relative to neutral expectation (B), based on the above formulas and assumptions and using selection and mutation parameters that are probably fairly realistic for the *D. melanogaster X* chromosome (which is a single arm) and an arm of a major *D. melanogaster* autosome. The model assumes that a chromosome is organized into blocks of coding sequences that are uninterrupted by introns, but are separated by blocks of noncoding sequence containing a mixture of weakly selected and strongly selected sites (Figure 1). Note that the truncation of very weakly selected mutations means that ~65% of the sites in the weakly selected sequences are treated as neutral, with the standard selection parameters used here. A diploid deleterious mutation rate $U = 0.24$ for this genomic region was assumed, on the basis of the genome-wide estimate of 1.2 from Haag-Liautard *et al.* (2007), which includes all types of deleterious mutations. The results for both intermediate dominance ($h = 0.5$) and partial recessivity ($h = 0.2$) are shown. There is good evidence that many slightly deleterious mutations are partially recessive ($h < 0.5$) (Crow and Simmons 1983; Garcia-Dorado and Caballero 2000), although very weakly selected mutations, such as most of those generated by the gamma distributions assumed

Table 2 The effects of background selection on *D. melanogaster* autosomal and *X* chromosomal genes

	Model 1	Model 2	Model 3	Model 4	Model 5
B values for autosomes					
Effects of strongly selected sites	0.684	0.684	0.665	0.680	0.638
	0.687	0.687	0.665	0.690	0.606
Effects of weakly selected noncoding sites	0.814	0.815	0.839	0.791	0.831
	0.806	0.806	0.839	0.757	0.867
Effects of all sites	0.556	0.557	0.558	0.538	0.530
	0.554	0.554	0.558	0.523	0.525
B values for *X* chromosome					
Effects of strongly selected sites	0.789	0.789	0.775	0.788	0.752
	0.790	0.790	0.775	0.790	0.746
Effects of weakly selected noncoding sites	0.878	0.878	0.896	0.859	0.895
	0.876	0.876	0.896	0.851	0.904
Effects of all sites	0.693	0.693	0.695	0.677	0.673
	0.692	0.692	0.695	0.672	0.675
Adjusted *X/A* diversity ratio for all sites	0.935	0.935	0.934	0.944	0.952
	0.937	0.937	0.934	0.964	0.964

See Table 1 for the meaning of the different models. *B* is the ratio of the effective population size under background selection to the neutral value. The parameters of the gamma distributions of selection coefficients are $a_1 = a_2 = 0.3$, $\bar{s}_1 = \bar{s}_2 = 2.5 \times 10^{-3}$ (strongly selected sites) and $a_3 = 0.3$, $\bar{s}_3 = 8 \times 10^{-6}$ (weakly selected sites). Results for the dominance coefficient $h = 0.5$ are shown in the top part of each row, and results for $h = 0.2$ are shown in the bottom part. A diploid deleterious mutation rate of $U_D = 0.24$ is assumed for both the autosomal arm (*A*) and the *X* chromosome. Map lengths of 0.5 and 0.6 M in female meiosis are assumed for *A* and *X*, respectively. The number of coding sequences in an arm (n_g) is 2800; the length of a coding sequence is 1500 bp. The noncoding regions between coding sequences are divided into 36 strongly selected sequences of length 39 bp and 37 weakly selected sequences of length 115 bp.

here, are likely to approach additivity (Wright 1934; Kacser and Burns 1981). It is thus not clear *a priori* which of these h values is likely to be more realistic, but an h value much less than 0.5 seems unlikely.

The map lengths of the X chromosome and the autosomal arm in female meiosis were set to 0.6 M and 0.5 M, respectively, which approximate the standard values for *D. melanogaster* (Ashburner *et al.* 2005). The genes in this model include only coding sequences. The relative values of X and A coalescent times are displayed after multiplying the B value for the X by 3/4, which is the ratio expected in the absence of BGS and with a 1:1 sex ratio and random variation in offspring number in both sexes (Wright 1931). This adjusted ratio provides a baseline prediction for the ratio of X/A neutral diversity values; an excess variance in male reproductive success due to sexual competition, or a female-biased sex ratio, would cause an even higher value (Hedrick 2007; Hutter *et al.* 2007; Vicoso and Charlesworth 2009a).

The predictions for B for autosomal loci when all sites are taken into account vary from ~0.52 to 0.56, and for X -linked loci from 0.67 to 0.69, depending on the model and the value of h. The adjusted X/A ratio varies from 0.93 to 0.96; this is the parameter of most interest for the purpose of this article, so that its relatively small range is encouraging. The large data set of Hutter *et al.* (2007) on variability in noncoding sequences in the Zimbabwe population of *D. melanogaster* gave an X/A diversity ratio of 0.90 after correcting for effects of GC content on diversity and divergence, which is in good agreement with the predictions of Table 2 and highly significantly different from the null value of 0.75. Note, however, that these predictions ignore possible effects of selection on the variants in the relatively long noncoding sequences involved, so the exact value of this ratio is still somewhat uncertain.

Model 1 involves the least approximations, but model 2 (which assumes a linear map) gives almost identical results. Model 3, which simply assumes a uniform density of selected sites across the chromosome, gives a remarkably good approximation to the model 1 results; models 4 and 5, somewhat surprisingly, give a slightly worse fit to the model 1 results than model 3 and overpredict the effects of BGS, although the differences are probably not meaningful for the purpose of comparisons with data. The bulk of the effect of BGS comes from the strongly selected sites, but the weakly selected sites make a significant contribution to increasing the adjusted X/A diversity ratio away from 3/4. For example, model 1 with $h = 0.5$ and no weakly selected sites gives an adjusted X/A diversity ratio of ~0.86 instead of 0.93.

For models 1 and 2, a smaller value of h gives a slightly larger effect of BGS; however, the effect of dominance appears to be negligible in all cases, consistent with the good performance of model 3 as an approximation, which is independent of h. However, it should be noted that the mean and threshold s values were kept unchanged from the $h = 0.5$ case, to isolate the effect of h. The distribution of mutational effects in *Drosophila* as estimated from population genetic data in reality involves t_s not s, since nonrecessive autosomal mutations are largely selected against on the basis of their heterozygous effects and X-linked mutations on the basis of a weighted average of their heterozygous effects on females and their hemizgygous effects on males (Charlesworth and Charlesworth 2010, p. 161), so that the s values with $h = 0.2$ should be adjusted to give the same distribution of t_s values as for the $h = 0.5$ case. This implies that changing

Table 3 The effects of background selection on *D. pseudoobscura* autosomal and *X* chromosomal genes

	Model 1	Model 2	Model 3	Model 4	Model 5
B values for autosomes					
Effects of strongly selected sites	0.855	0.855	0.843	0.856	0.812
	0.859	0.859	0.843	0.865	0.800
Effects of weakly selected noncoding sites	0.914	0.914	0.929	0.891	0.941
	0.910	0.910	0.929	0.868	0.961
Effects of all sites	0.782	0.782	0.784	0.763	0.765
	0.782	0.782	0.784	0.751	0.769
B values for *X* chromosome					
Effects of strongly selected sites	0.897	0.897	0.889	0.899	0.860
	0.898	0.898	0.889	0.900	0.864
Effects of weakly selected noncoding sites	0.940	0.940	0.950	0.922	0.960
	0.938	0.938	0.951	0.916	0.966
Effects of all sites	0.843	0.843	0.845	0.829	0.834
	0.843	0.843	0.845	0.825	0.835
Adjusted *X*/*A* diversity ratio for all sites	0.809	0.809	0.808	0.824	0.814
	0.809	0.809	0.808	0.823	0.814

See Table 1 for the meaning of the different models. The parameters are the same as for Table 2, except that map lengths of 1.2 and 1.3 M are assumed for *A* and *X*, respectively.

h should have no effect on the results, provided that the distribution of t_s is held constant, other than through rounding errors in the numerical results. This was verified by recalculating the results after multiplying the mean values of *s* for the strongly and weakly selected sites, as well as the threshold value of *s*, by $0.5/h$ and $3 \times 0.5/(2h + 1)$ for autosomal and *X*-linked loci, respectively. For example, with model 1 and the *D. melanogaster* parameters, the overall *B* values for *A* and *X* are 0.557 and 0.692, respectively, yielding an adjusted *X/A* diversity ratio of 0.93. The same argument can be applied to other modifications to the selection model, such as female- or male-specific selective effects, implying that the results should be robust to these changes.

BGS on the X and A in D. pseudoobscura

It is of interest to compare the results with those for *D. pseudoobscura* and its relatives, which have a two-arm *X* chromosome but single-arm autosomes and a much higher frequency of crossing over per base pair than *D. melanogaster* (Sturtevant and Tan 1937; Bachtrog and Andolfatto 2006; Kulathinal *et al.* 2008; Stevison and Noor 2010). The total map lengths of the chromosome arms are not known precisely; values of 1.2 M and 1.3 M have been assumed here for an autosome and an *X* chromosome arm, respectively, on the basis of Kulathinal *et al.* (2008) and Stevison and Noor (2010). Table 3 shows the results for these map lengths, with the other parameters being the same as for *D. melanogaster*. The adjusted *X/A* ratios are always substantially smaller than their counterparts in Table 2, reflecting the greater dissipation of BGS by the higher frequencies of recombination on both chromosomes.

Haddrill *et al.* (2010) found that, after removing loci that deviated significantly from neutrality, the estimated mean synonymous site diversities *X* and *A* for *D. pseudoobscura* were 0.0149 (SE = 0.0018) and 0.0230 (SE = 0.0021) for *D. pseudoobscura*, giving a value of 0.65 (SE = 0.26) for the *X/A* ratio, which is not significantly different from 0.75 but is significantly different from one and equal to the ratio of *X/A* effective population sizes estimated by Haddrill *et al.* (2011) after taking the recent population expansion in this species into account. The corresponding estimate for the close relative *D. miranda*, for which there is little evidence for a recent expansion, was 0.79. The observed ratios for these two species are thus statistically consistent with the values of $\sim$0.81 shown in Table 3.

Robustness of the results

The results for a focal site in the middle of an arm are largely insensitive to linkage to another chromosome arm of similar size to the one being considered. For example, under model 1 with the parameters used in Tables 2 and 3 with $h = 0.5$, the *B* values for *D. melanogaster* when an additional arm is present are 0.557 for *A* and 0.693 for *X*, with an adjusted *X/A* ratio of 0.93, *i.e.*, a very slight decrease over the Table 2 results. There is no effect at all for *D. pseudoobscura*. Complete insensitivity to the size of the chromosome is necessarily the case for model 3, which uses the result that, for a site that is not too close to the end of a chromosome, the effect of BGS depends only on the ratio of the total mutation rate to the map length (see derivation of the model 3 prediction above). Increasing both the mutation rate and the map length by the same factor, as would happen if the influence of an additional arm with similar mutational parameters were considered, thus has no effect on *B*.

Another important feature of the models is the length of the coding sequences *vs.* intergenic sequences. The model on which the above results are based ignores the fact that, as mentioned in the Introduction, most *Drosophila* genes have introns, many of which are several hundred base pairs or more in length and contain some selectively highly constrained sequences (*e.g.*, Sella *et al.* 2009). This can be crudely modeled

Table 4 Background selection with a large number of short coding sequences

	Model 1	Model 2	Model 3	Model 4	Model 5
B values for autosomes					
Effects of strongly selected sites	0.641	0.641	0.667	0.679	0.660
	0.833	0.833	0.845	0.855	0.829
Effects of weakly selected noncoding sites	0.820	0.822	0.838	0.797	0.822
	0.916	0.917	0.929	0.894	0.940
Effects of all sites	0.525	0.527	0.559	0.541	0.542
	0.763	0.764	0.785	0.764	0.780
B values for *X* chromosome					
Effects of strongly selected sites	0.758	0.758	0.774	0.786	0.763
	0.881	0.881	0.889	0.897	0.879
Effects of weakly selected noncoding sites	0.882	0.882	0.896	0.858	0.902
	0.941	0.941	0.951	0.924	0.960
Effects of all sites	0.668	0.668	0.695	0.673	0.689
	0.829	0.829	0.845	0.829	0.843
Adjusted *X/A* diversity ratio for all sites	0.954	0.950	0.932	0.932	0.953
	0.814	0.814	0.807	0.810	0.811

The top and bottom parts of each row show the results for *D. melanogaster* and *D. pseudoobscura*, respectively. See Table 1 for the meaning of the different models. A dominance coefficient $h = 0.5$ is assumed; the other selection parameters are the same as for Tables 2 and 3, except that 5600 coding sequences of length 750 bp, separated by 18 strongly selected 39-bp noncoding sequences and 19 weakly selected 113-bp noncoding sequences, are assumed.

by increasing the number of coding sequences, while holding their total length constant. The length of intergenic sequence is decreased proportionately, keeping the lengths of individual blocks of weakly selected and strongly selected noncoding sequences approximately constant.

It would be expected that dividing the chromosome arm into a larger number of shorter functional sequences, for the same total size and map length, would increase the effects of BGS and hence the *X/A* diversity ratio, since the average density of selected sites in relation to the frequency of recombination is reduced. Table 4 shows results that are otherwise comparable with the $h = 0.5$ results for *D. melanogaster* and *D. pseudoobscura*, for twice the number of coding sequences as before (some minor adjustments to the numbers and lengths of the noncoding sequences were made, to meet the assumptions about the organization of the chromosome).

As expected, the effects of BGS due to strongly selected sites under models 1 and 2 are enhanced, resulting in a slightly higher *X/A* diversity ratio than before. The effects of weakly selected noncoding sites are slightly diminished, presumably reflecting the fact that there are smaller clusters of blocks of these sites. The overall effect of BGS is greater than before, and the *X/A* diversity ratio for *D. melanogaster* is predicted to be >0.95, whereas that for *D. pseudoobscura* is barely changed at 0.81. Doubling the number of coding sequences again produces further effects in the same direction (results not shown), with the predicted *X/A* diversity ratios under models 1 and 2 for *D. melanogaster* and *D. pseudoobscura* becoming $\sim$0.98 and 0.82, respectively. The predictions of model 3 are no longer as close to the model 1 results as previously, mainly reflecting the increased effect of strongly selected sites in model 1, whereas the model 3 results change only marginally because of the adjustments in the parameters mentioned above. Model 4 performs only slightly better than model 3, mainly because it underpredicts the effects of strongly selected sites. Model 5 gives results that are closer to model 1, despite being an approximation to model 4. Overall, the results suggest that the *X/A* ratios are relatively insensitive to the way in which the chromosome arm is divided among coding and noncoding sequences, with a finer subdivision into strongly selected coding sequences leading to slightly larger effects of BGS.

The sensitivity of the results to the parameters of the distribution of selection coefficients was also examined. Since model 3 generally provides a good approximation, provided that the threshold value of s is kept constant, little effect of changing these parameters is expected, except by altering the proportion of deleterious mutations that fall below the threshold, thereby reducing the net truncated mutation rate U_T. It would therefore be expected that, for a given shape parameter a, the effect of BGS should be greater, the larger the mean selection coefficient; similarly, for a given mean selection coefficient, the effect of BGS should be greater, the larger the value of a, since this reduces the coefficient of variation of the distribution.

The effects of changing the selection parameters for strongly selected sites were investigated, since these contribute the most to the effects of BGS. The theoretical expectations were confirmed, but the effects on the adjusted *X/A* diversity ratio were relatively minor. For example, changing $\bar{s}_1 = \bar{s}_2$ from 2.5×10^{-3} to 0.01 or to 0.5×10^{-3} caused the adjusted *X/A* ratio predicted by model 1 for *D. melanogaster* (with shape parameter $a = 0.3$) to change from 0.935 to 0.936 and 0.922, respectively. Changing a from 0.3 to 0.6 or to 1.2 (with $\bar{s}_1 = \bar{s}_2 = 2.5 \times 10^{-3}$) caused this ratio to change to 0.950 and 0.952, respectively.

Conclusions

Overall, it seems that the results are fairly robust to the details of the selection parameters for deleterious mutations, for a given deleterious mutation rate. However, they are very sensitive to the deleterious mutation rate and the amount of recombination. Using model 3 (Hudson and Kaplan 1994; Barton 1995), which generally gives a reasonable approximation to the more exact results, the adjusted *X/A* ratio with a truncated deleterious mutation rate for an arm of U_T is equal to 0.75 × $\exp\{U_T(M_{eX} - M_{eA})/M_{eX}M_{eA}\}$, where M_{eX} and M_{eA} are the population-effective map lengths of the *X* chromosomal and autosomal arms, respectively. This ratio changes almost linearly from 0.79 to 0.93, over the range from $U_T = 0.0365$ to 0.146 (the value assumed above) with $M_{eX} = 0.40$ and $M_{eA} = 0.25$, the *D. melanogaster* values, and from 0.76 to 0.81 with $M_{eX} = 0.87$ and $M_{eA} = 0.60$, the *D. pseudoobscura* values. A higher U_T of 0.25 per chromosome arm gives values of 1.09 and 0.85, for the *D. melanogaster* and *D. pseudoobscura* map lengths, respectively. Such a high mutation rate seems implausible, however, given the size of the *Drosophila* genome and our current estimates of the mutation rate in *D. melanogaster* (Haag-Liautard *et al.* 2007; Keightley *et al.* 2009), so that the lower range of mutation rates used in these calculations is more likely to apply.

As might be expected intuitively, longer map lengths lead to smaller *X/A* diversity ratios and a lower sensitivity to the deleterious mutation rate. Species like *D. melanogaster*, with a small number of chromosomes and relatively short map lengths, are thus most likely to show an effect of BGS on the overall *X/A* diversity ratio. As discussed by Charlesworth (2012), BGS will tend to reduce rather than increase the *X/A* diversity ratio in taxa like mammals, where crossing over occurs on the autosomes in males (the same applies to the ratio of *Z* chromosome to autosomal diversity in birds, but Lepidoptera should behave like *Drosophila* because of their lack of crossing over in females), but the effect is likely to be fairly small because of the large number of chromosomes and the correspondingly low deleterious mutation rate per chromosome. Even for a *Drosophila* species, whether or not the ratio of *X* to autosomal neutral variability for a gene in the middle of the relevant chromosome arms is substantially greater than the null expectation of 3/4 is highly dependent on the deleterious mutation rate and the map lengths in question. More accurate knowledge of these parameters will help to resolve the question of whether the observations on *X/A* variability ratios in different populations and species can be accounted for solely by BGS or whether the other factors mentioned in the Introduction need to be invoked.

A role for selective sweeps rather than BGS in producing this effect cannot, of course, be ruled out, although these have usually been invoked to explain the *X/A* silent diversity ratio of <3/4 in non-African populations of *D. melanogaster* and *D. simulans* (Begun and Aquadro 1993; Aquadro *et al.* 1994; Begun and Whitley 2000; Andolfatto 2001; Harr *et al.* 2002; Hutter *et al.* 2007; Singh *et al.* 2007; Stephan 2010; Mackay *et al.* 2012), on the basis of a faster rate of adaptive evolution on *X* than on *A* (Charlesworth *et al.* 1987; Vicoso and Charlesworth 2009a) in response to the novel out-of-Africa environment. If this hypothesis is correct, then it seems unlikely that selective sweeps could be the cause of the *X/A* variability ratio of near one in East African populations of *D. melanogaster*, given that the theoretical study of the effect of selective sweeps on the *X/A* diversity ratio in *Drosophila* by Betancourt *et al.* (2004) showed that the fixation of partially recessive favorable mutations reduces the value of this ratio below 3/4.

However, the question of the cause of the much lower *X/A* diversity ratio in non-African populations of *D. melanogaster* and *D. simulans* remains undecided, since purely demographic explanations have been proposed as an alternative or supplement to the selective sweep model (Charlesworth 2001; Wall *et al.* 2002; Pool and Nielsen 2007, 2008). If the ancestral population had a high value of this ratio, this would seem to rule out demographic explanations based on a greater sensitivity of *X* than of *A* to a population bottleneck that require a value close to three-quarters (Pool and Nielsen 2007, 2008). It is possible, however, that demographic effects could interact with those of BGS to contribute to the reduced *X/A* diversity. A severe reduction in population size would be expected to reduce the effect of BGS, since a larger proportion of deleterious mutations will fall below the threshold for validity of the model used here. If the effectiveness of BGS for a population at equilibrium with greatly reduced effective population size is examined by increasing the threshold selection coefficient s_T in inverse proportion to the population size, the reduction can be quite large—for twofold and fourfold reductions below the values shown in Table 2, the adjusted *X/A* diversity ratios become 0.81 and 0.80, respectively. Not surprisingly, however, the ratio always remains above 3/4, in contrast to the observed ratio of <0.60 for the non-African sample of *D. melanogaster* in the meta-analysis in Table 4 of Singh *et al.* (2007).

However, the question of how background selection would interact with changing population size to affect the dynamics of the *X/A* diversity ratio remains to be studied; the computer algorithm for modeling BGS with recombination that has recently been developed by Zeng and Charlesworth (2011) should be helpful in this regard, since it can be modified to allow for changing population size. In this context, it is interesting to note that Singh *et al.* (2007) found that the level of polymorphism for intergenic noncoding sequences was similar for *X* and *A* in a sample from a U.S. population; both *X* and *A* showed a similar reduction in diversity compared with an African sample, suggesting that selective sweeps may be implicated in the greater reduction in variability for *X* than for *A* for sequences obtained from genes. The genome-wide surveys of diversity that are becoming available in *Drosophila* (*e.g.*, Sackton *et al.* 2009; Mackay *et al.* 2012) should help to resolve these questions.

Acknowledgments

I thank Laurence Loewe, Tim Connallon, Kai Zeng, and two reviewers for their helpful comments on this article. This work was supported by grant BB/G003076/1 from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

Literature Cited

Andolfatto, P., 2001 Contrasting patterns of *X*-linked and autosomal nucleotide variation in African and non-African populations of *Drosophila melanogaster* and *D. simulans*. Mol. Biol. Evol. 18: 279–290.

Aquadro, C. F., D. J. Begun, and E. C. Kindahl, 1994 Selection, recombination, and DNA polymorphism in Drosophila, pp. 46–56 in *Non-neutral Evolution: Theories and Molecular Data*, edited by B. Golding. Chapman & Hall, London.

Ashburner, M., K. G. Golic, and R. S. Hawley, 2005 *Drosophila. A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bachtrog, D., 2008 Evidence for male-driven evolution in Drosophila. Mol. Biol. Evol. 25: 617–619.

Bachtrog, D., and P. Andolfatto, 2006 Selection, recombination and demographic history in *Drosophila miranda*. Genetics 174: 2045–2059.

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

Bauer, V. L., and C. F. Aquadro, 1997 Rates of DNA sequence evolution are not sex biased in *Drosophila melanogaster* and *D. simulans*. Mol. Biol. Evol. 14: 1252–1257.

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

Begun, D. J., and P. Whitley, 2000 Reduced *X*-linked nucleotide polymorphism in *Drosophila simulans*. Proc. Natl. Acad. Sci. USA 97: 5960–5965.

Betancourt, A. J., Y. Kim, and H. A. Orr, 2004 A pseudohitchhiking model of X *vs.* autosomal diversity. Genetics 168: 2261–2269.

Casillas, S., A. Barbadilla, and C. M. Bergman, 2007 Purifying selection maintains highly conserved noncoding sequences in Drosophila. Mol. Biol. Evol. 24: 2222–2234.

Charlesworth, B., 1996 Background selection and patterns of genetic diversity in *Drosophila melanogaster.* Genet. Res. 68: 131–150.

Charlesworth, B., 2001 The effect of life-history and mode of inheritance on neutral genetic variability. Genet. Res. 77: 153–166.

Charlesworth, B., 2012 The effects of deleterious mutations on evolution at linked sites. Genetics 190: 5–22.

Charlesworth, B., and D. Charlesworth, 2010 *Elements of Evolutionary Genetics*. Roberts & Co., Greenwood Village, CO.

Charlesworth, B., J. A. Coyne, and N. H. Barton, 1987 The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. 130: 113–146.

Charlesworth, B., M. T. Morgan, and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289–1303.

Cobbs, G., 1978 Renewal approach to the theory of genetic linkage: case of no chromatid interference. Genetics 89: 563–581.

Crow, J. F., and M. J. Simmons, 1983 The mutation load in Drosophila, pp. 1–35 in *The Genetics and Biology of Drosophila*, Vol. 3c, edited by M. Ashburner, H. L. Carson, and J. N. Thompson. Academic Press, London.

Ellegren, H., 2009 The different levels of genetic diversity in sex chromosomes and autosomes. Trends Genet. 25: 278–284.

Eyre-Walker, A., and P. D. Keightley, 2009 Estimating the rate of adaptive mutations in the presence of slightly deleterious mutations and population size change. Mol. Biol. Evol. 26: 2097–2108.

Garcia-Dorado, A., and A. Caballero, 2000 On the average coefficient of dominance of deleterious spontaneous mutations. Genetics 155: 1991–2001.

Haag-Liautard, C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan *et al.*, 2007 Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. Nature 445: 82–85.

Haddrill, P. R., B. Charlesworth, D. L. Halligan, and P. Andolfatto, 2005 Patterns of intron sequence evolution in Drosophila are dependent upon length and GC content. Genome Biol. 6: R67.

Haddrill, P. R., L. Loewe, and B. Charlesworth, 2010 Estimating the parameters of selection on nonsynonymous mutations in *Drosophila pseudoobscura* and *D. miranda*. Genetics 185: 1381–1396.

Haddrill, P. R., K. Zeng, and B. Charlesworth, 2011 Determinants of synonymous and nonsynonymous variability in three species of Drosophila. Mol. Biol. Evol. 28: 1731–1743.

Halligan, D. L., and P. D. Keightley, 2006 Ubiquitous selective constraints in the Drosophila genome revealed by a genome-wide sequence comparison. Genome Res. 16: 875–884.

Harr, B., M. Kauer, and C. Schloetterer, 2002 Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA 99: 12949–12954.

Hedrick, P. W., 2007 Sex differences in mutation, recombination, selection, gene flow, and genetic drift. Evolution 61: 2750–2771.

Hudson, R. R., and N. L. Kaplan, 1994 Gene trees with background selection, pp. 140–153 in *Non-neutral Evolution: Theories and Molecular Data*, edited by B. Golding. Chapman & Hall, London.

Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics 141: 1605–1617.

Hutter, S., H. P. Li, S. Beisswanger, D. De Lorenzo, and W. Stephan, 2007 Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosome-wide nucleotide polymorphism data. Genetics 177: 469–480.

Kacser, H., and J. A. Burns, 1981 The molecular basis of dominance. Genetics 97: 639–666.

Keightley, P. D., and A. Eyre-Walker, 2007 Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics 177: 2251–2261.

Keightley, P. D., M. Trivedi, M. Thomson, F. Oliver, S. Kumar *et al.*, 2009 Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res. 19: 1195–1201.

Kimura, M., 1971 Theoretical foundations of population genetics at the molecular level. Theor. Popul. Biol. 2: 174–208.

Kulathinal, R. J., S. M. Bennett, C. L. Fitzpatrick, and M. A. F. Noor, 2008 Fine-scale mapping of recombination rate in Drosophila refines its correlation to diversity and divergence. Proc. Natl. Acad. Sci. USA 10: 10051–10056.

Loewe, L., and B. Charlesworth, 2006 Inferring the distribution of mutational effects on fitness in *Drosophila*. Biol. Lett. 2: 426–430.

Loewe, L., and B. Charlesworth, 2007 Background selection in single genes may explain patterns of codon bias. Genetics 175: 1381–1393.

Loewe, L., B. Charlesworth, C. Bartolomé, and V. Nöel, 2006 Estimating selection on nonsynonymous mutations. Genetics 172: 1079–1092.

Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* genetic reference panel. Nature 482: 173–178.

Misra, S., M. A. Crosby, C. J. Mungall, B. B. Matthews, K. S. Campbell *et al.*, 2002 Annotation of the Drosophila melanogaster euchromatic genome: a systematic review. Genome Biol. 3: Research0083.

Nordborg, M., B. Charlesworth, and D. Charlesworth, 1996 The effect of recombination on background selection. Genet. Res. 67: 159–174.

Pool, J. E., and R. Nielsen, 2007 Population size changes reshape genomic patterns of diversity. Evolution 61: 3001–3006.

Pool, J. E., and R. Nielsen, 2008 The impact of founder events on chromosomal variability in multiply mating species. Mol. Biol. Evol. 25: 1728–1736.

Sackton, T. B., R. J. Kulathinal, C. M. Bergman, A. R. Quinlan, E. B. Dopman *et al.*, 2009 Population genomic inferences from sparse high-throughput sequencing of two populations of *Drosophila melanogaster.* Genome Biol. Evol. 1: 449–450.

Schneider, A., B. Charlesworth, A. Eyre-Walker, and P. D. Keightley, 2011 A method for inferring the rate of occurrence and fitness effects of advantageous mutations. Genetics 189: 1427–1437.

Sella, G., D. A. Petrov, M. Przeworski, and P. Andolfatto, 2009 Pervasive natural selection in the Drosophila genome? PLoS Genet. 6: e1000495.

Singh, N. D., J. M. Macpherson, J. D. Jensen, and D. A. Petrov, 2007 Similar levels of *X*-linked and autosomal nucleotide variation in African and non-African populations of *Drosophila melanogaster.* BMC Evol. Biol. 7: 202.

Stephan, W., 2010 Genetic hitchiking *vs.* background selection: the controversy and its implications. Philos. Trans. R. Soc. B 365: 1245–1253.

Stevison, L. S., and M. A. F. Noor, 2010 Genetic and evolutionary correlates of fine-scale recombination rate variation in *Drosophila persimilis*. J. Mol. Evol. 71: 332–345.

Sturtevant, A. H., and C. C. Tan, 1937 The comparative genetics of *Drosophila pseudoobscura* and *Drosophila melanogaster.* J. Genet. 34: 415–431.

Vicoso, B., and B. Charlesworth, 2009a Effective population size and the Faster-*X* effect: an extended model. Evolution 63: 2413–2426.

Vicoso, B., and B. Charlesworth, 2009b Recombination rates may affect the ratio of *X* to autosomal noncoding polymorphism in African populations of *Drosophila melanogaster.* Genetics 181: 1699–1701.

Wall, J. D., P. Andolfatto, and M. F. Przeworski, 2002 Testing models of selection and demography in Drosophila simulans. Genetics 162: 203–216.

Wilson, D. J., R. D. Hernandez, P. Andolfatto, and M. Przeworski, 2011 A population genetics-phylogenetics approach to inferring natural selection in coding sequences. PLoS Genet. 7: e1002395.

Wright, S., 1931 Evolution in Mendelian populations. Genetics 16: 97–159.

Wright, S., 1934 Physiological and evolutionary theories of dominance. Am. Nat. 68: 25–53.

Zeng, K., 2010 A simple multiallele model and its application to identifying preferred–unpreferred codons using polymorphism data. Mol. Biol. Evol. 27: 1327–1337.

Zeng, K., and B. Charlesworth, 2009 Estimating selection intensity on synonymous codon usage in a non-equilibrium population. Genetics 183: 651–662.

Zeng, K., and B. Charlesworth, 2010 Studying patterns of recent evolution at synonymous sites and intronic sites in *Drosophila melanogaster.* J. Mol. Evol. 70: 116–128.

Zeng, K., and B. Charlesworth, 2011 The joint effects of background selection and genetic recombination on local gene genealogies. Genetics 189: 251–266.

Communicating editor: J. Wakeley

Appendix

Derivation of Equation 2

For a given pair of genes with index j, at the same distance to the right and left of the focal site and with a linear mapping function, the joint contribution to the negative of the exponent in Equation 1 can be written as

$$\frac{U_1 t_s}{0.7\, n_g l_g}\int_{z_{j1}}^{z_{j2}} \frac{\mathrm{d}z}{\left(t_s + r_j(z)[1-t_s]\right)^2} = \frac{U_1 t_s}{n_g l_g}\frac{l_g}{\left(r_{j2}-r_{j1}\right)}\int_{r_{j1}}^{r_{j2}} \frac{dr}{\left(t_s + r[1-t_s]\right)^2}, \tag{A1}$$

where z_{j1} and z_{j2} are the numbers of bases separating the beginning and end of the gene from the focal site, respectively, and r_{j1} and r_{j2} are the corresponding recombination fractions.

The factor of 0.7 in the denominator of the right-hand terms disappears from the left-hand side, because the assumption of a uniform density of selected sites along the coding sequence implies that $dz/dr = 0.7l_g/(r_1 - r_2)$, since the mean recombination frequency between adjacent selected sites is $(r_1 - r_2)/(0.7l_g)$ and not $(r_1 - r_2)/l_g$. Performing the integration, this expression becomes

$$\frac{U_1 t_s}{n_g\left(r_{j2}-r_{j1}\right)}\frac{1}{(1-t_s)}\left\{\frac{1}{\left(t_s + r_{j1}[1-t_s]\right)} - \frac{1}{\left(t_s + r_{j2}[1-t_s]\right)}\right\}, \tag{A2}$$

which yields Equation 2 of the text.

Distances from the Focal Site for Strongly Selected Noncoding Sites and Expressions for Models 4 and 5

The distances d_{1mj} and d_{2mj} from the focal site to the beginning and end of the mth block of strongly selected noncoding sites, in the jth intergenic region to its right, are as follows. Let $m = 1$ for the leftmost block in an intergenic region and $m = n_s$ for the rightmost block (there is a block of weakly selected noncoding sequence between each of these and the adjacent gene).

For $j = 0$, the focal site is located in the center of the weakly selected noncoding block in the middle of the intergenic region, so that

$$d_{1m0} = (m - 0.5)l_{\mathrm{iw}} + (m - 1)l_{\mathrm{is}}, \quad d_{2m0} = d_{1m0} + l_{\mathrm{is}} \quad \left(1 \leq m \leq \frac{n_{\mathrm{s}}}{2}\right). \tag{A3a}$$

For $1 \leq j \leq n_g/2$

$$d_{1mj} = jl_{\mathrm{g}} + (j - 0.5)l_{\mathrm{i}} + ml_{\mathrm{iw}} + (m - 1)l_{\mathrm{is}}, \quad d_{2mj} = d_{1mj} + l_{\mathrm{is}} \quad (1 \leq m \leq n_{\mathrm{s}}). \tag{A3b}$$

Note that $d_{2mj} - d_{1mj} = l_{\mathrm{is}}$, independently of m.

The corresponding population-effective recombination rates can be obtained by multiplying the d's by the appropriate function relating recombination rate to physical distance, yielding values of r_{1mj} and r_{2mj}. For the mth strongly selected noncoding block in the jth intergenic sequence, we can thus obtain expressions similar to Equations 2 and 3, except that U_1 is replaced by U_2 and n_{g} is replaced by $n_{\mathrm{g}}n_{\mathrm{s}}$ in Equation 2, to take into account the fact that there are a total of approximately $n_{\mathrm{g}}n_{\mathrm{s}}$ strongly selected noncoding sequences on the chromosome arm:

$$E_{2mj}(t_s) \approx \frac{U_2 t_s}{n_{\mathrm{g}}n_{\mathrm{s}}\left(t_s + r_{1mj}[1 - t_s]\right)\left(t_s + r_{2mj}[1 - t_s]\right)}. \tag{A4}$$

By taking the exponent of the negative of the sum of this expression over all m and j, we obtain the expression for B_2 in model 4.

The corresponding model 5 approximation to this sum can be obtained as follows. Following the method used for Equations 4 and 5 of the text, m is replaced by a continuous variable x, and Equation A4 is integrated with respect to x from $x = 1$ to $x = n_{\mathrm{s}}$. This yields an expression similar to Equation 5,

$$E_{2j}\,(t_s) \approx \frac{U_2 t_s}{n_{\mathrm{g}}n_{\mathrm{s}}\tilde{\rho}^2\,(l_{\mathrm{iw}} + l_{\mathrm{is}})l_{\mathrm{is}}} \ln\left\{\frac{\left(a_j' + bn_{\mathrm{s}}\right)\left(a_j + b\right)}{\left(a_j + bn_{\mathrm{s}}\right)\left(a_j' + b\right)}\right\}, \tag{A5}$$

where $a_0 = t_s - 0.5\tilde{\rho}l_{\mathrm{iw}}$, $a'_0 = a_0 - \tilde{\rho}l_{\mathrm{is}}$, $a_j = t_s + \tilde{\rho}(jl_{\mathrm{g}} + [j - 0.5]l_{\mathrm{i}})$, $a'_j = a_j - \tilde{\rho}l_{\mathrm{is}}$ (for $1 \leq j \leq n_g/2$), and $b = \tilde{\rho}(l_{\mathrm{iw}} + l_{\mathrm{is}})$.

By taking expectations of the first- and second-order terms in the deviations of the t_s from their mean over the truncated distribution of s, an expression similar to the negative exponent in Equation 6 is obtained for the sum of the contributions from the jth genes to the right and left of the focal site,

$$E_{2j} \approx \frac{U_{2_{\mathrm{T}}}}{n_{\mathrm{g}}n_{\mathrm{s}}\tilde{\rho}\,(l_{\mathrm{iw}} + l_{\mathrm{is}})}\left\{\frac{1}{\left(\bar{t}_2 + \beta_{2j}\right)}\left[\bar{t}_2 - \frac{\beta_{2j}V_2}{\left(\bar{t}_2 + \beta_{2j}\right)^2}\right] - \frac{1}{\left(\bar{t}_2 + \gamma_{2j}\right)}\left[\bar{t}_2 - \frac{\gamma_{2j}V_2}{\left(\bar{t}_2 + \gamma_{2j}\right)^2}\right]\right\}, \tag{A6}$$

where $\beta_{2j} = a_j + b - t_s$, $\gamma_{2j} = a_j + bn_{\mathrm{s}} - t_s$, and $U_{2_{\mathrm{T}}}$ is the deleterious mutation rate for strongly selected noncoding sites after truncation of the distribution of s; t_2 and V_2 are the mean and variance of t_s over this distribution.

The sum of the E_{2j} for all values of j between 0 and $n_g/2$ is needed to obtain the final approximate expression for B_2. The contribution E_{20} from the intergenic sequences immediately surrounding the focal site is given by Equations A5 and A6 with $j = 0$. The remaining part of the sum can be approximated by replacing j with a continuous variable y, so that β_{2j} and γ_{2j} are replaced by β_{2y} and γ_{2y}, and then integrating with respect to y from $y = 1$ to $y = n_g/2$. We have

$$\int_1^{n_g/2} \frac{dy}{\left(\bar{t}_2 + \beta_{2y}\right)} = \frac{1}{\lambda}\ln\left(\frac{\bar{t}_2 + \kappa + (1\,/\,2)\lambda n_{\mathrm{g}}}{\bar{t}_2 + \kappa + \lambda}\right) = I_{1\beta}, \tag{A7}$$

where $\kappa = \tilde{\rho}(l_{\mathrm{iw}} + l_{\mathrm{is}} - 0.5l_{\mathrm{i}})$ and $\lambda = \tilde{\rho}(l_{\mathrm{i}} + l_g)$.

A similar integral $I_{1\gamma}$ can be obtained for γ_{2y}, replacing $l_{\mathrm{iw}} + l_{\mathrm{is}}$ in the expressions for κ and λ with $(l_{\mathrm{iw}} + l_{\mathrm{is}})n_{\mathrm{s}}$.

Similarly, we have

$$\int_1^{n_g/2} \frac{\beta_{2y}dy}{\left(\bar{t}_2 + \beta_{2y}\right)^3} = \frac{1}{2\lambda\bar{t}_2}\left\{\frac{(\kappa + (1\,/\,2)\lambda n_g)}{(\bar{t}_2 + \kappa + (1\,/\,2)\lambda n_g)^2} - \frac{(\kappa + \lambda)^2}{(\bar{t}_2 + \kappa + \lambda)^2}\right\} = I_{2\beta} \tag{A8}$$

with an equivalent expression for $I_{2\gamma}$, again replacing $l_{\mathrm{iw}} + l_{\mathrm{is}}$ in the expressions for κ and λ by $(l_{\mathrm{iw}} + l_{\mathrm{is}})n_{\mathrm{s}}$.

These can be used in place of corresponding components of the sum of the E_{2j} in Equation A5, together with the E_{20} term, yielding the model 5 approximation for B_2.

Distances from the Focal Site for Weakly Selected Noncoding Sites and Expressions for Models 4 and 5

A similar approach can be used for weakly selected sites. The distances d_{1mj} and d_{2mj} from the focal site to the beginning and end of the mth block of weakly selected noncoding sites in the jth intergenic region to its right are given by

$$d_{110} = 0, \quad d_{210} = 0.5 l_{\mathrm{iw}} \quad (j = 0, m = 1) \tag{A9a}$$

$$d_{1m0} = (m-1)(l_{\mathrm{iw}} + l_{\mathrm{is}}) - 0.5 l_{\mathrm{iw}}, \quad d_{2m0} = d_{1m0} + l_{\mathrm{iw}} \quad \left(j = 0,\ 2 \le m \le \frac{n_{\mathrm{s}}}{2}\right). \tag{A9b}$$

For $1 \le j \le n_{\mathrm{g}}/2$

$$d_{1mj} = j l_{\mathrm{g}} + (j - 0.5) l_{\mathrm{i}} + (m-1)(l_{\mathrm{iw}} + l_{\mathrm{is}},), \quad d_{2mj} = d_{1mj} + l_{\mathrm{iw}} \quad (1 \le m \le n_{\mathrm{s}}). \tag{A9c}$$

These expressions yield the corresponding recombination frequencies, r_{1mj} and r_{2mj}, by multiplying by the appropriate function that relates recombination rate to distance.

Following the same procedure as for the strongly selected noncoding sequences, the equivalent of Equation A4 is

$$E_{3mj}(t_s) \approx \frac{U_3 t_s}{n_{\mathrm{g}}(n_{\mathrm{s}} + 1)\left(t_s + r_{1mj}[1 - t_s]\right)\left(t_s + r_{2mj}[1 - t_s]\right)}. \tag{A10}$$

This yields the expression for B_3 in model 4, by taking the exponential function of the negative of its sum over all m and j.

The model 5 approximation to this sum can be obtained in a similar way to that used for the strongly selected noncoding sites. The equivalent to Equation A8 is

$$E_{3j} \approx \frac{U_{3_{\mathrm{T}}}}{n_{\mathrm{g}}(n_{\mathrm{s}} + 1)\tilde{\rho}(l_{\mathrm{iw}} + l_{\mathrm{is}})} \left\{ \frac{1}{\left(\bar{t}_3 + \beta_{3j}\right)} \left[\bar{t}_3 - \frac{\beta_{3j} V_3}{\left(\bar{t}_3 + \beta_{3j}\right)^2} \right] - \frac{1}{\left(\bar{t}_3 + \gamma_{3j}\right)} \left[\bar{t}_3 - \frac{\gamma_{2j} V_3}{\left(\bar{t}_3 + \gamma_{2j}\right)^2} \right] \right\}, \tag{A11}$$

where $\beta_{30} = 0.5\tilde{\rho}\ l_{\mathrm{iw}}$, $\beta_{3j} = \tilde{\rho}\ (j l_{\mathrm{g}} + [j - 0.5] l_{\mathrm{i}} + l_{\mathrm{iw}})$ (for $1 \le j \le n_{\mathrm{g}}/2$), $\gamma_{30} = \tilde{\rho}\ (n_{\mathrm{s}} l_{\mathrm{is}} + [n_{\mathrm{s}} + 0.5] l_{\mathrm{iw}})$, $\gamma_{3\mathrm{j}} = \tilde{\rho}\ (j l_{\mathrm{g}} + [j - 0.5] l_{\mathrm{i}} + n_{\mathrm{s}} l_{\mathrm{is}} + [n_{\mathrm{s}} + 1] l_{iw})$ (for $1 \le j \le n_{\mathrm{g}}/2$), and $\bar{t}_3$ and V_3 are the mean and variance of t_s over the truncated distribution of s for weakly selected noncoding sites. (Here, the discontinuity between $m = 1$ and $m = 2$ for $j = 0$ has been ignored, since it makes only a small contribution to the total; Equation A9b has been used for the case when $j = 0$ and $m = 1$.)

Equivalents to Equations A7 and A8 can then be obtained by the same method as for strongly selected noncoding sites, where now $\bar{t}_2$ is replaced with $\bar{t}_3$, $\kappa = \tilde{\rho}\ (l_{\mathrm{iw}} - 0.5 l_{\mathrm{i}})$, and $\lambda = \tilde{\rho}\ (l_{\mathrm{i}} + l_{\mathrm{g}})$ in the equivalents of Equation A7 and A8; κ is replaced by $\tilde{\rho}\ (n_{\mathrm{s}} l_{\mathrm{is}} + [n_{\mathrm{s}} + 1] l_{\mathrm{iw}} - 0.5 l_{\mathrm{i}})$ in the corresponding expressions involving γ.