The Effects of Deleterious Mutations on Evolution at Linked Sites

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

ABSTRACT The process of evolution at a given site in the genome can be influenced by the action of selection at other sites, especially when these are closely linked to it. Such selection reduces the effective population size experienced by the site in question (the Hill–Robertson effect), reducing the level of variability and the efficacy of selection. In particular, deleterious variants are continually being produced by mutation and then eliminated by selection at sites throughout the genome. The resulting reduction in variability at linked neutral or nearly neutral sites can be predicted from the theory of background selection, which assumes that deleterious mutations have such large effects that their behavior in the population is effectively deterministic. More weakly selected mutations can accumulate by Muller's ratchet after a shutdown of recombination, as in an evolving *Y* chromosome. Many functionally significant sites are probably so weakly selected that Hill–Robertson interference undermines the effective strength of selection upon them, when recombination is rare or absent. This leads to large departures from deterministic equilibrium and smaller effects on linked neutral sites than under background selection or Muller's ratchet. Evidence is discussed that is consistent with the action of these processes in shaping genome-wide patterns of variation and evolution.

MUTATIONS that increase the fitness of their carriers are, of course, the basis of adaptive evolution. For this reason, the literature on evolutionary genetics is full of studies that document "positive" selection on genetic variants in natural populations—with the advent of data on the resequencing of numerous whole genomes from the same species, we can expect a tidal wave of such examples. But, as was pointed out long ago by geneticists such as Timofeeff-Ressovsky (1940) and Muller (1949, 1950), most mutations that affect the phenotype must be deleterious, because biological machinery has been subject to billions of years of selection to improve its performance. Because the nucleotide sites at which deleterious mutations can arise are numerous and distributed throughout the genome, the constant mutational production of deleterious variants and their elimination by "purifying" selection have major effects on the mean fitness of a population, its level of inbreeding depression, and its genetic variability with respect to fitness components.

doi: 10.1534/genetics.111.134288
Manuscript received August 26, 2011; accepted for publication November 4, 2011
[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

These properties of populations have important consequences for the evolution of such fundamental features of organisms as sex and genetic recombination, diploidy *vs.* haploidy, outbreeding *vs.* inbreeding, mate choice, and aging. Jim Crow, a great admirer of Muller (Crow 2005), has devoted much of his career both to improving our theoretical understanding of the population consequences of deleterious mutations and to documenting their rates of occurrence and their effects on fitness, summed up in several masterly reviews (Crow 1970, 1993, 2000; Simmons and Crow 1977; Crow and Simmons 1983). This body of work has been an inspiration to many who, like myself, have never had the opportunity to study or work with him.

In this article, I review work on an important indirect effect on evolution of the presence of deleterious mutations in populations, discovered by R. A. Fisher, another of Jim's heroes (Crow 1990). The following statement appears on pp. 121–122 of Fisher (1930b), referring to the fate of a new mutation in an asexual population:

> If we consider the prospect of a beneficial mutation occurring at any instant, ultimately prevailing throughout the whole group, and so leading to evolutionary progress, it is clear that its prospect of doing so will depend upon its chance of falling, out of the whole population, upon the one individual whose descendants are destined ultimately to survive.

> [. . .] if [. . .] the mutation rates both of beneficial and of deleterious mutations are high enough to maintain any considerable genetic diversity, it will only be the best adapted genotypes which can become the ancestors of future generations. . . (Fisher 1930b, pp. 121–122)

In the context of deleterious mutations, this means that only the portion of an asexual or nonrecombining population that carries the smallest number of mutations will contribute to the ancestry of future generations. A new beneficial mutation will then have a chance of spreading through the population only if it arises in that class—the descendants of all other classes are ultimately destined for elimination. This is equivalent to saying that the effective population size N_e (Wright 1931) is equal to the number of breeding individuals in this "least-loaded" class and is necessarily much smaller than the number of breeding individuals in the whole population (Fisher would not have approved of this way of putting it). This brings out the important point that the effect applies to the fates of neutral and slightly deleterious mutations, as well as beneficial ones. Sexual populations with recombination are far less subject to this reduction in N_e by mutations in the genetic background, since variants at different sites can disentangle themselves from initial chance associations.

This concept seems to have lain dormant for over 50 years; it is not referred to in the standard monograph on the evolution of sex (Maynard Smith 1978), nor does it appear in the reviews of the advantages of sex and recombination by Felsenstein (1988) and Crow (1988). The first step toward quantifying the effect was taken by Manning and Thompson (1984). This was, however, somewhat confusingly placed in the context of Muller's ratchet, which involves an instability in the distribution of the numbers of carriers of deleterious mutations when there is no recombination, brought about by genetic drift in a finite population (Muller 1964; Felsenstein 1974). Rice (1987) similarly made a brief reference to it in the context of the ratchet, while Birky and Walsh (1988) presented simulation results that showed that the presence of deleterious mutations in a nonrecombining genome reduces the chance of fixation of a positively selected new mutation and enhances the chance of fixation of a new deleterious mutation.

The influence of deleterious mutations on evolution at closely linked sites is a form of what Joe Felsenstein (Felsenstein 1974) called the "Hill–Robertson effect", after the seminal article of Hill and Robertson (1966), whereby selection at one site in the genome interferes with the action of selection at other sites (reviewed by Comeron *et al.* 2008). This can usefully be thought of in terms of a reduction in N_e caused by the increased variance in fitness associated with the action of selection, an idea introduced in relation to freely recombining loci by Robertson (1961); the effect is stronger when there is close linkage, because associations between variants at different sites are then maintained for much longer (Santiago and Caballero 1998). As shown below, for a nonrecombining genome this formulation is equivalent to equating N_e to the size of the least-loaded class, provided that selection is sufficiently strong.

The Hill–Robertson effect was discussed by Felsenstein (1974, 1988) in the context of interference among positively selected mutations spreading through a population, a process originally described by Fisher (1930b) and Muller (1932), as well as with reference to Muller's ratchet. For some time, however, these ideas largely remained the subject of theoretical discussion, without many applications to data, but the situation has now changed dramatically.

Interest among population geneticists in the effects of selection on variability and evolution at linked sites increased in the late 1980s and early 1990s, in response to the finding that putatively neutral silent DNA restriction site variability in *Drosophila* populations was lowest in regions of the genome with little or no genetic recombination (Aguadé *et al.* 1989; Stephan and Langley 1989). Across the *Drosophila melanogaster* genome as a whole, there was soon found to be a positive correlation between the local recombination rate experienced by a gene and its level of silent variability (Begun and Aquadro 1992), a finding that has been amply confirmed by later research (Begun *et al.* 2007; Shapiro *et al.* 2007). These findings stimulated the refinement of models of hitchhiking by positively selected mutations ("selective sweeps") (Kaplan *et al.* 1989; Stephan *et al.* 1992; Braverman *et al.* 1995; Simonsen *et al.* 1995), a concept introduced by Maynard Smith and Haigh (1974), and to the use of the data on the relation between variability and recombination in *D. melanogaster* to estimate the frequency of such sweeps (Wiehe and Stephan 1993; Stephan 1995)—other things (such as gene density) being equal, selective sweeps should be more effective in reducing variability when recombination rates are low.

It was, however, soon realized that the elimination of deleterious mutations has a similar effect to selective sweeps on neutral or nearly neutral variability at linked sites, by the process outlined above. A simple analytic formula was derived for the reduction in neutral variability caused by deleterious mutations at mutation–selection equilibrium in the absence of recombination (Charlesworth *et al.* 1993); our work on this problem was helped by discussions with Nick Barton and with Jim, who had been thinking along similar lines. We called this effect "background selection" (a term also used by Birky and Walsh 1988) and showed that it provides an alternative to selective sweeps as a possible explanation of the correlation between recombination rate and diversity in *Drosophila*. It does not, of course, exclude a role for selective sweeps, and both processes are likely to contribute to the observed patterns (Sella *et al.* 2009; Stephan 2010).

In this article, I focus exclusively on the effects of deleterious mutations on evolution at linked sites, both in the context of background selection and when purifying selection is sufficiently weak that mutations deviate substantially from their deterministic equilibria as a result of genetic drift—the latter situation is especially important for

large nonrecombining genomic regions and for asexual or highly self-fertilizing species (Charlesworth *et al.* 2010). Given the pervasive occurrence of deleterious mutations throughout the genome, it may well be necessary to include their effects on variability and evolution at linked sites in any tests for the operation of other forces, such as positive selection. As shown below, there is increasingly good evidence for such effects.

The Properties of Deleterious Mutations

Before discussing the processes introduced above, it is useful to describe briefly what is known about the frequency of occurrence of deleterious mutations and the sizes of their effects on fitness. This is, of course, a field to which Jim has made fundamentally important contributions throughout his career. Methods for studying the genome-wide occurrence of deleterious mutations, and the magnitudes of their effects on fitness, are reviewed by Keightley (2012), so that only a brief summary is given here, with an emphasis on *Drosophila*, Jim's favorite experimental organism.

Mutation accumulation experiments and the deleterious mutation rate

One parameter of major importance is the per genome deleterious mutation rate, U. This is the mean number of mutations that appear per generation in a newly formed zygote, which reduce the fitness of their carriers. It represents the sum of individual mutation rates per nucleotide site over all relevant sites in the genome. The classical method for estimating U has been to measure a fitness-related trait such as viability or fertility, in experiments where lines started from the same initial isogenic stock have fixed mutations independently, an approach pioneered in *Drosophila* by Terumi Mukai (Mukai 1964; Mukai *et al.* 1972). Selection against mutations is rendered largely ineffective by minimizing the N_e of each line and/or maintaining the chromosomes accumulating mutations heterozygous over a balancer chromosome, before they are assayed as homozygotes. Such experiments allow measurement of the rate of decline of the mean value across homozygous lines (DM) of a fitness-related trait such as egg-to-adult viability,and the rate of increase in the variance of line means (DV).

The ratio $(DM)^2/DV$ provides a lower bound to the estimate of U; the mean selection coefficient, $\bar{s}$, for a homozygous mutation is $\leq DV/DM$ (Bateman 1959; Mukai 1964), where the selection coefficient s for a given mutation is defined as the reduction below one in the ratio of the fitness of mutant homozygotes to the fitness of wild-type homozygotes. These relations are equalities only if all mutations have the same s values. In general, there will be a probability distribution of the values of s, $\phi(s)$; a wide distribution of s can lead to U being greatly underestimated and $\bar{s}$ being overestimated. More sophisticated statistical methods have also been used to analyze mutation accumulation experiments, such as fitting maximum-likelihood models to a specified distribution of s to the data, but it has proved hard to obtain accurate parameter estimates (Halligan and Keightley 2009).

Attention is usually restricted to nonlethal or nonsterile mutations ("detrimentals"), as these constitute the majority of spontaneous deleterious mutations (Crow and Simmons 1983). Unfortunately, the results for detrimental mutations have been rather discordant, even for *D. melanogaster* (Halligan and Keightley 2009; Mallet *et al.* 2011, 2012). Overall, there is evidence in this species for a detectable DM for fitness (equal to $U\bar{s}$) of ~1–2%. It is less easy to be sure of the values of U and $\bar{s}$. The most promising way of resolving these uncertainties is to use methods based on DNA sequence information, which are considered next.

Molecular methods for estimating U

It was pointed out by Kondrashov and Crow (1993) that DNA sequence comparisons of related species can provide an estimate of the proportion of mutations that are sufficiently deleterious that they are certain to be eliminated from the population by selection. If sequence divergence is measured for a class of presumptively selectively neutral sequences, then the divergence per neutral nucleotide site, K_{neu}, is proportional to the product of the mutation rate and the time separating the two sets of sequences (Kimura 1983). If K_{sel} is the corresponding divergence value for sequences that are subject to purifying selection, such as nonsynonymous sites or putatively functional noncoding sequences, then $1 - K_{sel}/K_{neu}$ provides an estimate of the "level of selective constraint," c, provided that mutation rates are similar for the two classes of sequences. If this proviso is met, then c is likely to be an underestimate of the true fraction of mutations that are deleterious, since in practice some slightly deleterious mutations get fixed by drift, and some mutations are fixed by positive selection. Various elaborations of this method have since been developed (Keightley *et al.* 2011; Keightley 2012).

An estimate of U can be obtained if estimates of c for different components of the genome are available, by taking the sum over all components of the product of c times the mutation rate per base pair and the number of base pairs for the component in question. It is now possible to determine mutation rates in model organisms by the detection of DNA sequence alterations in mutation accumulation lines; in humans, comparisons of whole genome sequences of parents and offspring are being carried out (Keightley 2012). By using these methods, U has been estimated as 1.2 per diploid zygote per generation in *D. melanogaster*; for humans, the current estimate of U is 2.1 (Keightley 2012). Both of these values should be regarded as provisional.

The distribution of fitness effects of deleterious mutations

An estimate of c tells us little about the magnitude of the selection coefficient associated with a typical deleterious mutation; unless recombination rates are very low, there is a negligible probability of fixation of a deleterious mutation

with a selection coefficient of the order of a small multiple of the reciprocal of the N_e of a species (Fisher 1930a; Kimura 1962). As far as past evolutionary change is concerned, $1/N_e \sim 10^{-4}$ for humans, with their relatively small historical effective population size, and 10^{-6} for many *Drosophila* species (Charlesworth and Charlesworth 2010, Table 5.2). Even very minute selection coefficients are therefore compatible with high levels of selective constraints. Since mutation accumulation experiments have not provided reliable estimates of the distribution of selection coefficients, recent research has used data on the frequencies within a population of variants at nucleotide sites that are putatively subject to purifying selection (Keightley 2012).

To relate data on the frequencies of nucleotide site variants within populations to the underlying distribution of selection coefficients against deleterious amino acid mutations, sequence polymorphism data are fitted to population genetic models that predict measurable features of the population from $\phi(s)$. On the assumption that most autosomal mutations in randomly mating populations are selected against when carried in heterozygotes, the methods estimate the distribution of the reductions in fitness to heterozygous carriers of mutations, which we can denote by t_s for a given homozygous effect s. This assumption is justified by the finding that lethal mutations in *Drosophila* reduce fitness by ~2% when they are heterozygous with wild type (Dobzhansky and Wright 1941; Crow and Temin 1964), while detrimental mutations tend to be much closer to semidominance with respect to their fitness effects (Crow and Simmons 1983; Garcia-Dorado and Caballero 2000), as expected on biochemical grounds (Wright 1934; Kacser and Burns 1981).

Applications of these methods have mainly involved nonsynonymous variants and human, mouse, plants, and *Drosophila* populations (Gossmann *et al.* 2010; Haddrill *et al.* 2010; Halligan *et al.* 2010; Keightley and Eyre-Walker 2010; Slotte *et al.* 2010). In addition, to reduce the dimensionality of the problem, these studies have mostly used distributions that can be described by two parameters (*e.g.*, mean and standard deviation), such as the normal, log-normal, and gamma distributions. The results suggest that the distribution of selection coefficients is very wide, so that gamma or log-normal distributions generally provide a much better fit than normal distributions. In *D. pseudoobscura*, for example, the coefficient of variation of t_s (the ratio of the standard deviation to the mean) exceeds one, and the mean selection coefficient against mutations segregating in the population is very small, such that its product with N_e is of the order of 20 (Haddrill *et al.* 2010); this, of course, indicates a very small value of the relevant mean of t_s, $\sim 10^{-5}$. It has proved more difficult, however, to obtain accurate estimates of the arithmetic mean of t_s for newly arising mutations—the point estimates of this parameter have very large confidence intervals and are sensitive to the model that is used. In *Drosophila*, it seems clear, however, that there is only a small proportion (≤5%) of new nonsynonymous mutations whose behavior as polymorphic variants within the population is effectively selectively neutral ($N_e t_s \leq 0.5$). Around 90% are so strongly selected ($N_e t_s \geq 5$) that they have a negligible chance of fixation. In humans with their much smaller effective population size, the proportion of effectively neutral mutations is closer to 25% (Eyre-Walker and Keightley 2009).

These results can be shown to imply that the mean number of deleterious nonsynonymous mutations that behave approximately deterministically and that are carried by a typical individual is ~5000 for *Drosophila* (Haddrill *et al.* 2010) and 800 for humans (Charlesworth and Charlesworth 2010, p. 295). The estimated mean number of deleterious nonsynonymous mutations present in a randomly chosen gene sampled from a population is surprisingly large for organisms like *Drosophila* (which has 14,000 genes)—$0.5 \times (5000/14{,}000) = 0.18$ per gene in *D. pseudoobscura* (Haddrill *et al.* 2010).

These findings suggest that (1) we cannot ignore the possibility that the continual input and selective elimination of deleterious mutations will affect evolution at nearby sites in the genome and (2) much of the behavior of these mutations, at least in freely recombining parts of the genome, can be captured by deterministic models of mutation and selection.

Mutation and selection

Before describing how selection against deleterious mutations may influence evolution at linked sites, the basic population genetics theory of mutation and selection in a diploid, infinitely large, randomly mating population is presented. Any complications due to genetic drift and nonrandom associations between linked sites are ignored at this point. It is assumed that the deleterious mutations of interest are so strongly selected against that they are kept at low frequencies, so that second-order terms in their frequencies can be neglected. Provided that mutations affect fitness independently of each other (*i.e.*, fitnesses can be combined multiplicatively across sites), the following results apply equally well to sexual and asexual populations (Kimura and Maruyama 1966; Crow 1970).

For nonrecessive, autosomal mutations affecting a nonsynonymous site or noncoding site of functional importance, these assumptions mean that the equilibrium frequency q_i of the mutant variant at the ith site in question is determined by the ratio of the mutation rate to the deleterious variant, u_i, and the heterozygous selection coefficient against the variant, t_i (back mutations can be neglected, since the mutant variants are all at low frequencies). With a dominance coefficient h_i, the effective selection coefficient against the mutation is $t_i = h_i s_i$ (provided that $h_i > 0$; as noted above, this is usually likely to be the case), where s_i is the selection coefficient against homozygotes. For *X*-linked mutations with equal selection on the two sexes, the corresponding effective selection coefficient against a mutation is $t_i = (2h_i + 1)s_i/3$. Haploid populations can be treated similarly,

with t_i representing the haploid selection coefficient. In all three cases, we have $q_i = u_i/t_i$ (Charlesworth and Charlesworth 2010, pp. 160–161).

Summing $2q_i$ across sites, the mean number of mutations carried by a newly born diploid individual in a given genomic region is

$$\bar{n} = 2\sum_i \frac{u_i}{t_i}, \tag{1a}$$

where the sum is taken over all relevant sites in the region in question (this is the value for females in the case of *X* linkage). The corresponding mean number of mutations in a haploid genome is one-half of this.

If the mutation rates and selection coefficients vary independently across sites, this simplifies to

$$\bar{n} = \frac{U}{t_h}, \tag{1b}$$

where t_h is the harmonic mean value of the t_i across sites (*i.e.*, the reciprocal of the mean of the reciprocals of t_i).

It follows that, in an equilibrium population, t_h measures the mean selection coefficient of a segregating deleterious mutation discussed above; this mean is, of course, biased toward weakly selected mutations compared with the arithmetic mean of the *t* values for new mutations.

Finally, it is useful to note that the equilibrium distribution of the numbers of mutations carried by different individuals in the population is a Poisson distribution with mean $\bar{n}$ (Kimura and Maruyama 1966). In particular, this implies that the equilibrium frequency of mutation-free individuals in an infinitely large population, f_0, is equal to $\exp(-\bar{n})$. If we average across this distribution and assume multiplicative fitness interactions between variants at different sites, the mean fitness of the population (relative to a value of one for a mutant-free individual) is

$$\bar{w} \approx \prod_i (1-2q_i t_i)^i \approx \exp(-U). \tag{2}$$

This result holds for any mode of selection against deleterious mutations in an asexual population (Kimura and Maruyama 1966); in sexual populations, the mean fitness is higher than this when there is synergistic (negative) epistasis among mutations, such that log fitness declines faster than linearly with the number of mutations and is lower with diminishing-returns (positive) epistasis (log fitness declines more slowly than linearly with the number of mutations) (Kimura and Maruyama 1966; King 1966; Crow 1970; Kondrashov 1982; Charlesworth 1990; Shnol and Kondrashov 1993). This finding has important implications for the evolutionary significance of sex and recombination (Kondrashov 1982, 1988), which continues to be debated but is not considered here; the consequences of Hill–Robertson interference for this question are reviewed briefly in the *Discussion*.

Background Selection

The above empirical and theoretical findings set the stage for examining the consequences of the continual input and removal of deleterious mutations for evolution in nearby regions of the genome. It is helpful to think about this problem in terms of three divisions of a continuum, according to the effectiveness of selection against deleterious mutations relative to drift. At one extreme, sites are subject to purifying selection that is so strong that $N_e t >> 1$, and their average frequencies are close to their equilibrium frequencies under mutation and selection, as analyzed in the previous section [the *background selection* (BGS) regime]. (It is important to note that N_e is here taken to be the value in the absence of any interference effects.)

For somewhat smaller $N_e t$ values, most sites remain close to their equilibria, so that any reverse mutations from mutant to wild type can be neglected; however, with no or very little recombination there can be repeated stochastic losses of the genotypes with the smallest number of deleterious mutations, leading to a gradual buildup of deleterious mutations [*Muller's ratchet* (MR)]. With even weaker selection, the efficacy of selection is so greatly reduced by interference among the selected sites that deleterious mutations can drift to intermediate frequencies [the *weak selection Hill–Robertson interference* (WSHRI) regime]. This is, of course, a somewhat arbitrary division of a continuum of effects into discrete categories, with gray zones at the boundaries; in addition, given that there is probably a wide range of selection coefficients and recombination rates across the genome, different sites in the genome of the same species may fall into different categories.

In the remainder of this section, I discuss background selection. The other two processes are considered, rather more briefly, in the next section.

A general result for the effect of BGS on the level of variability

The simplest question that can be asked about this process is, What is the effect of BGS on the mean coalescence time, T_2, for a pair of alleles sampled from the population, at a given neutral site? Since the mean pairwise nucleotide site diversity (π) for neutral mutations under the infinite sites model is directly proportional to T_2 (Hudson 1990), the answer to this question allows predictions to be made about the effect of BGS on DNA sequence variability. Under the standard assumptions of coalescent theory for a panmictic population with constant size, $T_2 = 2N_e$ (Hudson 1990; Charlesworth and Charlesworth 2010, p. 217), so that T_2 can be used to provide a measure of the effect of BGS on the effective population size or level of neutral variability at a given location in the genome.

In a large randomly mating population, for sites with $N_e t_i >> 1$ that are distributed along a single chromosome or chromosomal region, the following formula (Hudson and Kaplan 1995; Nordborg *et al.* 1996a; Nordborg 1997)

provides a good approximation for B for a given neutral site, which is defined as the ratio of T_2 under BGS to its value in the absence of any interference effects of selection,

$$B \approx \exp\left(-\sum_i \frac{u_i t_i}{(t_i + r_i[1-t_i])^2} \right), \quad (3)$$

where r_i is the frequency of recombination between the focal neutral site under consideration and the ith site subject to purifying selection; the other variables are defined in the previous section.

A heuristic derivation, using the idea that the variance in fitness associated with linked deleterious mutation reduces N_e, is given in the *Appendix*. This shows that this expression underestimates the effect of BGS when there is loose linkage between the focal neutral site and the sites under selection, in the case of diploidy but not haploidy (Santiago and Caballero 1998). Alternative derivations that use the coalescent process with subdivision of the population according to genotype, which are close in spirit to the least-loaded approach for the zero recombination case, are given by Hudson and Kaplan (1995), Nordborg (1997), and Charlesworth and Charlesworth (2010, pp. 401–402).

With free recombination ($r_i = 0.5$), the following result applies to diploids:

$$B \approx \exp\left(-8 \sum_i u_i t_i \right). \quad (4)$$

The term $8\sum_i u_i t_i$ is equal to four times the additive variance in fitness under mutation–selection balance contributed by the sites in question (Mukai *et al.* 1972). Data on genetic variability in fitness suggest that this variance is unlikely to be much more than a few percent for the genome as a whole (Charlesworth and Hughes 2000), so that unlinked sites make only a relatively small contribution to BGS in a randomly mating population, as would be expected intuitively. It should, however, be borne in mind that a typical site on a given *Drosophila* or mammalian chromosome has an N_e or T_2 that is probably reduced to 95% or so of the strictly neutral value by mutation and selection acting on other chromosomes. The comparisons made below should therefore be interpreted as relating N_e or T_2 values for a focal site to a value that includes the effects of unlinked sites, not to the ideal situation without any BGS effects.

Highly inbreeding species, such as *Caenorhabditis elegans* and *Arabidopsis thaliana* (currently the subject of intensive population genetic studies), need a different treatment for loose linkage. In a population with inbreeding coefficient F, the population-effective recombination rate is reduced from r to $r(1 - F)$ (Dye and Williams 1997; Nordborg 1997). When F is close to one, even unlinked sites behave as though they are closely linked, and the combined effects of selection at sites across the whole genome must be considered (Nordborg *et al.* 1996b), in a similar way to the case of low recombination in randomly mating populations discussed below.

Overall effects of BGS on levels of variability

Equation 3 can be used to provide a broad-brush picture of the effect of BGS on variability for a chromosome or part of a chromosome (Nordborg *et al.* 1996a). The simplest approach is to assume that sites subject to mutation and selection are distributed uniformly along a chromosome, with the same effective selection coefficient t at each site and a diploid mutation rate U for the chromosome. The map length of the chromosome is M, and double crossovers are assumed to be negligible in frequency, so that recombination is linearly related to map distance, with a constant frequency of recombination between adjacent sites along the chromosome. A proportion x of the chromosome is assumed to be located to the left of the focal site and $1 - x$ to the right. Summation over sites can be approximated by integration, yielding the following expression:

$$B \approx \exp\left(-\frac{U(t+2x[1-x]M[1-t])}{2(t+xM[1-t])(t+[1-x]M[1-t])} \right) \quad (5a)$$

(Nordborg *et al.* 1996a). For $t << M$ and when the focal site is not too close to the end of the chromosome (see Nordborg *et al.* 1996a), we have $x(1 - x)M >> t$, and this expression reduces to

$$B \approx \exp\left(-\frac{U}{M} \right); \quad (5b)$$

i.e., the effect of BGS depends only on the density of deleterious mutations per unit map length, as first pointed out by Hudson and Kaplan (1994) and Barton (1995). (The lack of dependence on t means that this result also holds with an arbitrary distribution of t values.)

For a site at either end of the chromosome, such that either x or $(1 - x) = 0$, a similar approximation yields

$$B \approx \exp\left(-\frac{U}{2M} \right). \quad (5c)$$

The effect of background selection in this case is thus smaller than for a focal site in most of the rest of the chromosome, by a factor of $\exp(-U/2M)$; *i.e.*, B is the square root of the previous value (Nordborg *et al.* 1996a).

BGS in a recombining region is thus most effective in the middle of a chromosome or chromosomal region and least effective at either end of it, as one would intuitively expect from the fact that the density of sites under selection that are close to the focal site is greatest in the center of the region. Equations 5b and 5c can thus be used to predict the extent to which BGS is effective in reducing the overall level of neutral or nearly neutral variability in a defined genomic region. For example, the map length in females is ~0.5 M for an arm of one of the two major autosomes of *D. melanogaster* (Ashburner *et al.* 2005, Chap. 10). Given that there is no crossing over in males and that an autosomal gene spends half of its time in each sex, the population-effective

map length (M_e) to be used in the equations is 0.25, since recombination rates in males and females should each be given a weight of one-half in this case (Charlesworth and Charlesworth 2010, p. 381). Using a slightly conservative estimate of 1.0 for the deleterious mutation rate for *D. melanogaster* and the fact that a chromosome arm is on average ~20% of the genome, the relevant $U = 0.20$. Equation 5b then gives $B = 0.45$, a large reduction below 1. Even with $U = 0.5$, $B = 0.67$.

In contrast, humans have 23 chromosomes, with a sex-averaged mean map length per autosome of ~1.57 M (Jensen-Seaman *et al.* 2004); U for the whole genome is estimated to be ~2.1 (see above), giving a value of 1.99 for the autosomes, from their contribution to the assembled genome sequence relative to the whole genome (Jensen-Seaman *et al.* 2004). This yields an estimate of $B \approx 0.94$ for a centrally located site on a typical autosome, a fairly trivial effect [but note that the considerable variation in physical sizes and map lengths among chromosomes (Jensen-Seaman *et al.* 2004) has been ignored here]. This corresponds well to the estimate that hitchhiking effects have reduced the average neutral diversity on a human chromosome by ~6%, on the basis of the relation between variability and proximity to functionally important sequences (Cai *et al.* 2009).

X chromosomes have a different exposure to recombination from that of autosomes, since they spend two-thirds of their time in females and one-third in males, where they do not recombine with most of the *Y* (the reverse is true for *Z* chromosomes in species with female heterogamety). In mammals and other groups with crossing over on the autosomes in both sexes, this implies a stronger overall effect of BGS on the *X* chromosome. Values for physical size and map lengths (Jensen-Seaman *et al.* 2004) allow estimation of the relevant U and population-effective map length for the *X*, yielding $B \approx 0.91$ for humans, ~97% of the autosomal value.

This does not, however, directly predict the ratio of pairwise coalescent times or neutral diversities for *X vs.* autosomes, since these also depend on the effective population sizes in the absence of BGS. If the sex ratio is 1:1 and the variances in offspring numbers are similar for males and females (*e.g.*, when there is little competition among males for mates), the neutral value of N_e for the *X* is three-quarters of that for the autosomes (Wright 1931; Charlesworth and Charlesworth 2010, p. 224), so that the predicted ratio of T_2 values for an equilibrium human population is $(3 \times 0.97)/4 = 0.73$. In addition, the difference between male and female mutation rates in mammals (Crow 2000) means that putatively neutral diversity estimates for *X* and autosomes should be divided by the respective estimates of neutral divergence from an outgroup species for this prediction to be tested.

Overall *X* to autosome within-population diversity ratios of 0.73 and 0.61 (after such divergence adjustments) were obtained from African and European whole genome resequencing data by Gottipatti *et al.* (2011); values of >0.75 and 0.71 for Africans and Europeans were obtained by Hammer *et al.* (2010) from smaller data sets. The ratios are significantly higher for sites close to genes than for sites remote from genes, strongly suggesting the action of hitchhiking effects caused by selection on coding or regulatory sequences, as was also found in another study of four human populations (Hernandez *et al.* 2011). In Africans, but not Europeans, the *X* to autosome diversity ratio is >0.85 far from genes (Gottipatti *et al.* 2011). These differences between populations may reflect nonequilibrium demography in European populations and the effect of an excess variance in male mating success due to sexual competition, which inflates the *X* to autosome diversity ratio (Charlesworth and Charlesworth 2010, p. 224).

In *Drosophila*, the lack of crossing over in males means that the *X* has $\sim\frac{4}{3}$ times the autosomal population-effective recombination rate, for pairs of sites with comparable frequencies of recombination in females on the *X* and the autosomes. In *D. melanogaster*, it also has a slightly longer map than an autosomal arm (0.6 M) (Ashburner *et al.* 2005, Chap. 10). Equation 5b with $U = 0.20$ then yields $B \approx 0.61$ for the *X* chromosome. The *X*/autosome ratio of *B* values is thus 1.36, in sharp contrast to the mammalian values, and the predicted ratio of diversity values for an equilibrium population is $(3 \times 1.36)/4 = 1.02$, in the absence of sexual competition effects. There is little evidence for mutation rate differences between *X* and autosomes in *Drosophila* (Bauer and Aquadro 1997; Keightley *et al.* 2009; Zeng and Charlesworth 2010b), so that this implies that there should be similar or even slightly higher levels of silent site diversity for the *X* chromosome than the autosomes, as is indeed observed for the presumptively ancestral East African populations of *D. melanogaster* (Andolfatto 2001; Hutter *et al.* 2007; Singh *et al.* 2007). In contrast, a similar calculation for *D. pseudoobscura*, which has map lengths that are more than twice those in *D. melanogaster* (Kulathinal *et al.* 2008; Stevison and Noor 2010), yields a predicted *X* to autosome diversity ratio of 0.83. This is close to the values obtained from estimates of silent site variability at *X*-linked and autosomal loci in *D. pseudoobscura* and its relative *D. miranda* (Haddrill *et al.* 2010).

These calculations show that even species in the same taxonomic group may differ quite widely in the overall predicted effect of BGS on a chromosome and in the relative values of its effect on the *X vs.* the autosomes. Opposite patterns of *X* to autosome diversity ratios are expected for mammals and *Drosophila*. However, the assumption of a uniform density of sites subject to selection along a chromosome is obviously unrealistic, since these are always clustered into coding sequences and groups of functional noncoding sequences, separated by blocks of sequence (intronic and intergenic), much of which is likely to be under weak or no selection, especially in organisms like mammals where only a small fraction of the genome is apparently subject to purifying selection (Keightley 2012). More realistic models

of a chromosome arm of *Drosophila*, with alternating weakly selected and strongly selected noncoding and coding sequences, have recently been examined (B. Charlesworth, unpublished results). These show that the ratio of *X* chromosome to autosomal diversities seems to be fairly insensitive to the presence of intergenic weakly selected sequences.

Recombination rate and BGS

We now consider the question of how variation in recombination rates across the genome affects *B* for local regions of the genome. The most extreme situation is a genome region with no recombination at all (or at least no crossing over), as is the case for *Y* and *W* chromosomes in many species and the "dot" chromosome of *Drosophila*. In the complete absence of recombination, and assuming that the t_i follow a distribution with harmonic mean t_h, Equations 1b and 3 together yield the expression

$$B \approx \exp\left(-\frac{U}{2t_h}\right), \tag{6}$$

where *B* is the first term P_0 of a Poisson distribution whose mean is equal to the mean number of mutations per haploid genome, $U/(2t_h)$; *i.e.*, it is the frequency in the population of haplotypes that lack any deleterious mutations. (This expression was originally derived by Charlesworth *et al.* 1993, using the argument outlined in the Introduction.)

This result suggests that, to predict the reduction in nucleotide site variability caused by BGS in a nonrecombining region of the genome, we need to know the deleterious mutation rate *U* for this region and the harmonic mean selection coefficient for deleterious mutations, t_h (see *The Properties of Deleterious Mutations*). The small dot chromosome of *Drosophila* is classic material for asking this question, since it represents an isolated region of the genome with an almost complete lack of crossing over in all species that have been studied (Ashburner *et al.* 2005). With a gene content of ~80 genes, it represents ~0.57% of the coding sequence in the genome. Using a conservative estimate of the deleterious mutation rate of 0.5 per diploid genome and a t_h of 10^{-3} (substantially higher than the estimate mentioned earlier), $P_0 = \exp(-0.25 \times 0.057 \times 10^3) = \exp(-14.2) = 6.8 \times 10^{-7}$.

This implies that no variability should be found on the dot chromosome in a sequence variability study of reasonable size, even allowing for the considerable stochastic variation of diversity around its expected value that is caused by the vagaries of the coalescent process (Hudson 1990). In fact, the dot is generally found to have ~10% as much variability as a typical region of the genome (Betancourt *et al.* 2009; Arguello *et al.* 2010). Equation 6 thus wildly *overpredicts* the effect of BGS for the dot; as shown below, this is because the close linkage between the sites under selection undermines the assumption that these are all held close to their equilibria under selection.

When there is a moderate amount of recombination, however, simulations show that Equation 3 should work well as a description of the effect of BGS for a single chromosome, for sites at which $t_i N_e > \sim 3$ (Nordborg *et al.* 1996a; Barton and Etheridge 2004). Equation 3, or the related equation of Hudson and Kaplan (1995), has therefore been used in several attempts to test whether BGS can explain the observed relations between nucleotide site diversity and the chromosomal location of a sequence or the local rate of genetic recombination that it experiences.

The earliest of these attempts successfully used the then available data on DNA sequence variation in *D. melanogaster*, combined with estimates of local recombination rates from the *Drosophila* genetics literature, and per genome mutation rates and selection coefficients against deleterious mutations obtained from the Mukai experiments (Hudson and Kaplan 1995; Charlesworth 1996); a parallel study was carried for the second chromosome of *D. pseudoobscura* (Hamblin and Aquadro 1999). However, the *D. melanogaster* material available at the time came largely from European and North American samples, which are known to be depauperate in variability compared with the ancestral East African populations, probably as a result of population bottlenecks associated with movement to new continents (Begun and Aquadro 1993; Andolfatto 2001; Haddrill *et al.* 2005b; Hutter *et al.* 2007). It remains to be seen whether the use of the more recent estimates of selection and mutation parameters discussed earlier, together with genome-wide data from East African populations, produces similarly good fits and to what extent selective sweeps also contribute to these patterns.

Attempts have also been made to investigate the fit to BGS models of data on the level of human genetic diversity in the human genome (Payseur and Nachman 2002; Reed *et al.* 2005; Hellmann *et al.* 2008; Cai *et al.* 2009; McVicker *et al.* 2009; Lohmuller *et al.* 2011). As mentioned above, these studies, and those of Hammer *et al.* (2010), Hernandez *et al.* (2011), and Gottipatti *et al.* (2011), provide evidence that diversity at putatively neutral or nearly neutral sites is reduced when these are close to selectively constrained coding and noncoding sequences and that diversity is significantly correlated with local recombination rates. While BGS models generally provide a good fit to the data, it remains unclear to what extent they provide a unique explanation or whether the effects of selective sweeps also contribute (Hellmann *et al.* 2008; Cai *et al.* 2009; Lohmuller *et al.* 2011).

Some attempts have been made to distinguish between the two explanations using a method proposed by Innan and Stephan (2003), which relies on differences between the expected curves relating diversity and recombination under sweeps *vs.* BGS when recombination rates are moderately low, *e.g.*, Hellmann *et al.* (2008). These should, however, be regarded with caution as a way of rejecting BGS, because this approach uses Equation 5b. This equation is inaccurate when the recombination rate between the focal site and the selected sites is of similar magnitude to, or smaller than, the selection coefficient, as is likely to be the case with low

recombination rates, and it also assumes that selected sites are evenly spread over the region in question. A recent analysis of whole human genome resequencing data has provided little evidence for recent selective sweeps (Hernandez *et al.* 2011), suggesting that BGS may be the primary factor in humans.

Apart from *Drosophila* and humans, there have been few genome-wide studies of the relation between recombination rate and local recombination rates and gene densities. In the highly selfing nematode species *C. briggsae*, there is a significant relation between silent nucleotide site diversity and recombination rate, which fits poorly to the expectation under analytical and simulation models of the effects of recurrent selective sweeps (Cutter and Choi 2010). Rockman *et al.* (2010) showed that patterns of quantitative variation in gene expression levels in *C. elegans* in relation to recombination rate fit a BGS model well, although a contribution from sweeps could not be ruled out. Roselius *et al.* (2005) found a significant relation between diversity levels and recombination rate in wild tomato species. Nordborg *et al.* (2005) and Kawabe *et al.* (2008) found a negative association between gene density and diversity levels in *A. thaliana* and *A. lyrata*, respectively, as did Flowers *et al.* (2012) in rice; given the positive correlation between recombination rate and gene density, this tends to obscure any effect of recombination rate. Positive correlations between recombination rate and putatively neutral diversity have also been reported in budding yeast (Cutter and Moses 2011) and chickens (Rao *et al.* 2011).

Effects of BGS on the shapes of gene genealogies

The reduction in pairwise coalescent time caused by BGS, as described by Equation 3, can heuristically be regarded as equivalent to a reduction in N_e. However, this is only part of the picture. It was recognized from the start of theoretical studies of BGS that it also distorts the shape of gene genealogies, causing an excess of terminal branches relative to the equilibrium neutral expectation and a corresponding excess of rare variants, especially when selection is weak (Charlesworth *et al.* 1993, 1995). The number of neutral sites in a sample that are segregating for polymorphic variants (S) is, therefore, not reduced by BGS to the same extent as the pairwise diversity, π, resulting in negative expected values of test statistics such as Tajima's D_T (Charlesworth *et al.* 1993, 1995; Tachida 2000; Gordo *et al.* 2002; Williamson and Orive 2002; O'Fallon *et al.* 2010; Seger *et al.* 2010). One way of looking at this effect is to note that, unless $N_e t >> 1$, the time to loss of a deleterious mutation is not very different from the time to loss of a neutral variant that is destined to be lost (Kimura and Ohta 1969). Low-frequency mutations that are likely to be lost contribute more to S than to π, so that S is less affected than π by the elimination of deleterious variants at linked sites, unless $N_e t$ is very large.

Apart from the heuristic expression for the expected value of S derived by Santiago and Caballero (1998), the magnitude of this effect has mainly been studied by structured coalescent calculations or simulations that assume a fixed selection coefficient and no recombination. This limits their applicability to data on natural variability, but the results provide a general guide to what might be happening. Distortions of the genealogy in a nonrecombining genome can occur, such that the mean value of Tajima's D_T statistic $\approx$ –0.6, even with $N_e t$ as large as 30, when U/t is such that $B \approx$ 0.2 (Gordo *et al.* 2002). Larger values of U/t or smaller values of N_e lead to violation of the assumptions of the BGS model (see below) and to much larger distortions of the genealogy and variant frequency spectra (see below).

Recombination should greatly diminish these distortions of the genealogy, but its effect has been little studied, except by Santiago and Caballero (1998). A structured coalescent procedure for modeling BGS that incorporates recombination has recently been developed, which shows that some degree of distortion occurs even with recombination (Zeng and Charlesworth 2011); this approach should enable tests of significance for agreement between the data and the predictions of BGS models to be developed. For parameters that correspond reasonably well to those for a single gene of *D. melanogaster* in a region of normal recombination, $B \sim$ 0.90 and 0.85 for sites at one end or in the middle of a gene, respectively; the ratio of the expected length of the gene genealogy represented by the terminal branches to the size of the rest of the genealogy is increased to $\sim$1.02 at the end of the gene and 1.05 in the middle (Zeng and Charlesworth 2011) and of course declines away from genes. While these effects are small, they should be detectable in genome-wide resequencing screens of variability.

The first attempt to test for a role of BGS in producing distortions of neutral genealogies was made by Charlesworth *et al.* (1995), using a rather limited *D. melanogaster* data set. Larger than average negative D_T values have subsequently been found in low recombination regions in African populations (Andolfatto and Przeworski 2001; Sella *et al.* 2009). Reed *et al.* (2005) and Stajich and Hahn (2005) found a similar pattern in humans, as did Cutter and Choi (2010) in *C. briggsae*. Given that selective sweeps can also generate this type of effect (Braverman *et al.* 1995), its interpretation remains ambiguous, although Lohmuller *et al.* (2011) concluded that most of the excess of rare silent variants in human populations observed near genes is likely to be caused by BGS. In addition, a causal role for selective sweeps in the reduced variability on the dot chromosome of *D. americana* seems to be ruled out (Betancourt *et al.* 2009).

Effects of BGS on selection at linked sites

The question of the effect of BGS on the outcome of selection at a focal site has mostly been studied by determining the probability of ultimate fixation in the population of a deleterious or an advantageous mutation at the site in question, relative to its value in the absence of BGS. Since these fixation probabilities are widely used for interpreting rates of molecular evolution (Kimura 1983) and features of the genome such as GC content and codon usage bias (Bulmer 1991; Sharp *et al.* 2010), knowledge of

the effects of BGS on fixation probabilities is important for understanding the effects of different levels of genetic recombination on the efficacy of selection. The fixation probability of a single copy of a semidominant mutation with selection coefficient s when homozygous (s is positive if the mutation is advantageous and negative when deleterious), relative to the value for a neutral mutation, is given by

$$\lambda \approx \frac{\gamma}{\{1 - \exp(-\gamma)\}}, \quad (7)$$

where $\gamma = 2N_e s$ (Fisher 1930a; Kimura 1983).

This is also the ratio of the rate of evolutionary substitution of mutations with this value of s, relative to the neutral rate, when new mutations are the source of evolutionary change (Kimura 1983). The effectiveness of selection relative to genetic drift can thus conveniently be measured by γ. This suggests that we could predict the effect of BGS on the fixation probabilities of mutations through its effect on N_e for neutral variants at a focal site, simply by replacing γ in Equation 7 by $B\gamma$, where B is determined by the equations described above. Numerical and analytical studies show that a sufficient condition for this to be true is for the selection coefficient at the focal site to be smaller than the selection coefficients at the sites causing BGS (Barton 1994, 1995; Charlesworth 1994; Peck 1994; Stephan *et al.* 1999; Johnson and Barton 2002). This is likely to apply, for example, when selection against nonsynonymous mutations in coding sequences influences the much weaker selection on codon usage at synonymous sites in the same gene, although the picture here is complicated by interference among the synonymous sites themselves (see below).

However, it is less clear what happens when selection at the focal site is greater than or of similar strength to those causing BGS. Johnson and Barton (2005) examined the fixation probability of a beneficial mutation in a nonrecombining genome subject to BGS and found a complex pattern when this has a selection coefficient as large as those causing BGS. If s for a beneficial mutation is $>U$, but t for the deleterious mutations is $<U$, then there is essentially no reduction in fixation probability as a result of BGS. Unfortunately, we currently lack firm estimates of the distribution of the selection coefficients on beneficial mutations, although some recent work suggests that a substantial proportion of favorable nonsynonymous mutations in *Drosophila* could have mean values of s as low as 10^{-5} (Sella *et al.* 2009; Sattath *et al.* 2011; Schneider *et al.* 2011). For a fairly large genomic region, such as a newly evolving Y chromosome with $\geq$1000 genes, U will probably be $\geq$0.05, and there will then be a substantial reduction in the fixation probabilities of beneficial nonsynonymous or regulatory noncoding mutations and an increase in the fixation probabilities of very weakly deleterious synonymous or noncoding mutations, even for s values as high as 0.01.

The expectation of a relaxed efficiency of selection in genomic regions with low frequencies of recombination has been tested in a number of *Drosophila* systems, reviewed by Charlesworth *et al.* (2010). A virtual absence of recombination seems to be associated with a higher rate of substitution of nonsynonymous mutations, a higher ratio of nonsynonymous to silent diversity levels, a reduced rate of substitution of adaptive nonsynonymous mutations, and reduced selection for codon usage. These patterns are consistent with the action of BGS and the other Hill–Robertson interference effects discussed in the next section (Charlesworth *et al.* 2010), although there may also be contributions from selective sweeps.

Loewe and Charlesworth (2007) used Equation 3 to examine the effect of BGS due to purifying selection acting on nonsynonymous mutations within genes of *D. melanogaster*, considered in isolation from other genes. They found that B is reduced in the middle of genes compared with their ends and is lower for genes that lack introns compared with genes with introns [these were assumed to evolve neutrally, but this is not necessarily realistic (Haddrill *et al.* 2005a)]; it is higher, the greater the contribution of introns to the total size of a gene and for short compared with long coding sequences. These conclusions parallel observations on overall patterns of codon usage in the *D. melanogaster* genome (Comeron and Kreitman 2002; Qin *et al.* 2004), suggesting that BGS may play an important role in generating these very local patterns by reducing the efficiency of selection on synonymous mutations affecting codon usage. Population genetic estimates of such selection (Comeron and Guthrie 2005) are consistent with this interpretation, although Hill–Robertson interference among the synonymous sites themselves (see below) may also contribute to these patterns (Comeron and Kreitman 2002; Comeron and Guthrie 2005; Comeron *et al.* 2008). Recent simulation results suggest, however, that the effect of selection on nonsynonymous sites dominates, except for near-zero recombination rates (Zeng and Charlesworth 2010a).

Muller's Ratchet and Weak Selection Hill–Robertson Interference

Muller's ratchet

Muller's ratchet operates when a nonrecombining genome or genome segment fixes deleterious mutations at a much higher rate than with recombination, but most sites that are subject to purifying selection remain close to their equilibria under mutation and selection. This means that back mutations from mutant to wild type can be ignored, at least in the initial stages of the process. The effects of MR on neutral or weakly selected sites embedded in the region in question are thus very similar to those of BGS (Gordo *et al.* 2002). Most modeling work on MR has considered a population of haploid organisms, with the same selection coefficient at all sites. Since the Y chromosome is permanently heterozygous with the X chromosome, this also describes the evolution of a nonrecombining Y, except that the selection coefficient against a deleterious mutations in a haploid is replaced by

that for heterozygotes with wild type, to a good approximation (Kaiser and Charlesworth 2010). The same general principles apply to diploids, with some differences in details (Charlesworth and Charlesworth 1997). We can imagine the ratchet starting to move when a recombining genome or genome region becomes nonrecombining quite suddenly, as a result of the formation of a new *Y* chromosome or the loss of sexual reproduction or outcrossing.

The classical approach to understanding MR is to recall that there is a Poisson distribution of the equilibrium number of mutations per individual, if fitness effects are multiplicative. The frequency of the zero mutation class of haplotype, P_0, is then given by Equation 6 or its haploid equivalent. As discussed earlier, P_0 is likely to be very small for a large region of the genome, and at most only a few individuals will be mutation-free. Assume for the moment that the genome region in question is sufficiently small that some mutation-free individuals are indeed present. If recombination is shut down in this region, the "least-loaded" zero-mutation class is then easily lost from the population by genetic drift. With no recombination and no back mutation, it can never be regenerated (a "click of the ratchet"). Once the zero class has been lost, the distribution of the numbers of mutations reshapes itself and eventually approaches the deterministic equilibrium distribution, with the least-loaded class carrying one mutation (Haigh 1978). This process is repeated indefinitely in a ratchet-like manner, at an approximately constant click rate, which depends in a complex way on the effective population size, the value of P_0, and the strength of selection (Haigh and Maynard Smith 1974; Gessler 1995; Stephan and Kim 2002; Loewe 2006; Söderberg and Berg 2007; Brunet *et al.* 2008; Jain 2008). With a fixed selection coefficient s in a haploid population, the ratchet will move at an evolutionarily significant rate when $sN_0 \leq 15$, where $N_0 = NP_0$ is the size of the zero mutation class in a population of N breeding individuals (Gordo and Charlesworth 2000).

In the haploid case, each click of the ratchet is followed by the fixation of a single mutation somewhere in the genome region involved (Rice 1987; Higgs and Woodcock 1995; Charlesworth and Charlesworth 1997). This is because, after the loss of a least-loaded class, the new least-loaded class contains a mix of haplotypes, each carrying a mutation that was absent from the previous least-loaded class. Because this class is small in number, one of the haplotypes will rapidly become fixed within it by drift. The absence of back mutations means that, as with BGS and no recombination (see the Introduction), the least-loaded class eventually becomes the ancestor of all of the other classes in the population, so that the mutation that was fixed in the least-loaded class is present throughout the entire population.

This assumes that $N_0 > 1$, so that some mutation-free individuals are present in the initial population. But even with $\bar{n}$ as low as 50, the frequency of the zero class is 1.9×10^{-22}. No mutation-free individuals can therefore be present in the population unless it is improbably large. In such a case, a more complex analysis is required (Gessler 1995; Brunet *et al.* 2008). This shows that the ratchet moves very fast, accompanied by fixations of deleterious mutations. This situation is likely to be close to reality for an asexual population derived from a sexual ancestor or even for the small dot chromosome of Drosophila. The mean fitness of an asexual population of multicellular eukaryotes, with a deleterious mutation rate ≥ 1, must decline rapidly, potentially dooming the population to extinction (Lynch *et al.* 1993). Similarly, an evolving *Y* chromosome may accumulate large numbers of mutations that reduce its functionality, leading to its degeneration over time (Charlesworth 1978; Gordo and Charlesworth 2001; Gordo *et al.* 2002; Kaiser and Charlesworth 2010).

The effects of the ratchet on neutral sites within a nonrecombining region affected by MR can also be studied; it turns out that structured coalescent simulations that assume a stable frequency distribution of genotypic classes with respect to the selected sites give a remarkably good fit to full simulations of MR, even when the condition that N_0 exceeds one is violated (Gordo *et al.* 2002). Provided that sN_0 is so large that MR does not operate and the population stays in the BGS regime, B decreases with smaller s, for a fixed population size, corresponding to the decline in P_0 in the haploid equivalent of Equation 6. This is accompanied by an increased distortion of the gene genealogy, as mentioned in the *Background Selection* section. Once s is small enough for the ratchet to start operating, B reaches a minimal value and then starts to increase again, reaching unity when $s = 0$; this minimum coincides approximately with a maximal distortion of neutral genealogies, but the distortion stays fairly flat until s approaches zero (Gordo *et al.* 2002).

An unrealistic assumption of the ratchet model is, however, one-way mutation from good to bad, except for irreversible types of mutation such as deletions or insertions. This assumption works reasonably well for the initial process of departure from the deterministic equilibrium under mutation and selection, when the wild-type variant is much more frequent than the mutant at each site. In such situations, reverse mutations can be neglected. However, for nucleotide substitutions, back mutations will become increasingly frequent as deleterious mutations become fixed at many sites. This slows the ratchet down, until the rate of fixation of deleterious mutations is balanced by fixations of reverse mutations. Once this state has been reached, there will be a flux of fixations of both deleterious mutations and beneficial reverse mutations and no ongoing decline in mean fitness. Unless the population has gone extinct, a new equilibrium, associated with a reduced mean fitness, will be established (Gordo and Charlesworth 2001; Kaiser and Charlesworth 2009).

Thus, although MR is formally an irreversible process of accumulation of deleterious mutations, this does not accurately describe what happens after a long period of time has elapsed, as far as the nucleotide substitutions that form the

majority of mutations are concerned. Irreversible mutations can, however, accumulate by MR, and this may explain the fixation of deletions and insertions causing loss of function of genes carried on newly evolving *Y* or neo-*Y* chromosomes (Kaiser and Charlesworth 2010). These arise at a much lower rate than single-nucleotide mutations and have much larger fitness effects, so that the assumption that $N_0 > 1$ is quite reasonable, even for a whole *Drosophila* chromosome arm (Kaiser and Charlesworth 2010). Interestingly, MR for irreversible mutations is speeded up significantly by the presence of more weakly selected, reversible mutations in the background, presumably because they reduce N_e (Kaiser and Charlesworth 2010).

In very small, diploid nonrecombining populations, a buildup of partially recessive mutations can cause a "crystallization" of the genome into two complementary segregating haplotypes, within each of which deleterious alleles become fixed (Charlesworth and Charlesworth 1997; Pálson and Pamilo 1999; Pálson 2001). If maintained for a long time, this situation leads to the accumulation of differences at neutral sites and hence an enhancement of silent site variability in regions with no recombination. Currently, there seems to be little evidence that this occurs in nature.

Weak selection Hill–Robertson interference

These considerations imply that MR is only a partial description of the behavior of a nonrecombining genome or genomic region that is fixing reversible deleterious mutations more rapidly than with free recombination. In fact, it may be regarded as a subclass of what has rather clumsily been called WSHRI (McVean and Charlesworth 2000), which is now described. If many sites under purifying selection (*e.g.*, nonsynonymous sites) are packed into a nonrecombining genome or genomic region, Hill–Robertson interference (see the Introduction) among them can weaken the effective strength of selection acting on each of them. Thus, deleterious nonsynonymous or noncoding variants can reach high frequencies or fixation. For a large nonrecombining genome or genomic region with several tens of thousands of sites under selection, this effect can be very large (for example, for an incipient *Y* chromosome or even the dot chromosome of *Drosophila*).

Simulations with realistic distributions of selection coefficients show that the decline in the overall level of adaptation in this situation is accompanied by an increased rate of substitution of deleterious variants and a reduction in DNA sequence variability at both the selected sites and neutral sites located in the same genomic region, reflecting a reduction in N_e; the reduction in variability can be on the order of 100-fold in regions with $\geq$1 Mb of coding sequence and is ~10-fold for a region the size of the dot chromosome (~80 kb of coding sequence), in agreement with data on sequence variability (Kaiser and Charlesworth 2009). The distortion in gene genealogies at neutral sites embedded in the region, with an accompanying excess of rare variants, can be very large, much greater than with BGS or MR, and indeed approaches its maximum possible value (Kaiser and Charlesworth 2009; Zeng and Charlesworth 2010a).

The ratio of diversity at selected sites to that at neutral sites is greatly increased compared with the case of free recombination, indicating a decline in the efficacy of selection *vs.* drift at the selected sites. Interestingly, the effects of adding more sites under selection into a nonrecombining region seem to become weaker as the number of sites increases, and the level of neutral diversity approaches an asymptotic value, presumably because the interference among the selected sites increasingly undermines their effective strength of selection (Kaiser and Charlesworth 2009). This is probably why a large nonrecombining section of the genome, such as the nonrecombining neo-*Y* chromosome of *D. miranda*, which is estimated to have ~1500 genes that are still functional (Bachtrog *et al.* 2008), shows a silent site diversity value that is still measurable [~1% of the value for homologous sites on the neo-*X* chromosome (Bartolomé and Charlesworth 2006)]. As noted by McVean and Charlesworth (2000), these effects of WSHRI imply that organisms such as bacteria with low genome-wide levels of recombination will have much lower effective population sizes and hence levels of silent site variability and codon usage bias than would be expected from their very large numbers of individuals.

Discussion

There is now a solid body of theory that makes testable predictions about the effects of deleterious mutations on evolution at linked sites. There are also clear signals from genome analyses, population genetic data, and between-species comparisons that are consistent with these predictions, but it is often unclear to what extent selective sweeps also contribute to the observed patterns. There is reason to be optimistic that the impending explosion of data from genome-wide resequencing projects will provide better estimates of critical parameters such as the deleterious mutation rate, the distribution of fitness effects of new mutations, and the frequency and strength of selective sweeps, enabling more conclusive tests of alternative hypotheses.

An important aspect of local variation in N_e across the genome, driven either by BGS or selective sweeps, is that it influences the level of neutral divergence between closely related species, as noted by Reed *et al.* (2005). This is because the fixation of ancestral polymorphisms subsequent to the split between two species can contribute a substantial fraction of the total number of neutral fixations along a lineage, for $>10N_e$ generations, where N_e is the long-term effective population size for the lineage in question (Charlesworth *et al.* 2005)—this represents ~2.5 million years in the case of humans. This boosts the rate of divergence over the value expected from the fixation of new mutations, which is unaffected by selection at linked sites (Birky and Walsh 1988). Where local N_e is reduced by the effects of selection, there will be a smaller amount of

polymorphism and hence lower divergence. In humans, it seems likely that at least part of the observed relations between recombination rate and divergence can be explained by this effect (Reed *et al.* 2005; McVicker *et al.* 2009; Lohmuller *et al.* 2011), suggesting that the proposed mutagenic effect of recombination (Hellmann *et al.* 2003) is probably an artifact. For between-population comparisons within a species, F_{ST} values are expected to be higher for genes or genomic regions for which N_e is low (Charlesworth *et al.* 1997; Charlesworth 1998); there is evidence for this in human populations (Keinan and Reich 2010) and *C. briggsae* (Cutter and Choi 2010), for example. Again, this effect needs to be carefully considered when interpreting data, since ignoring it can lead to false inferences of local selective differences among populations (Charlesworth 1998; Noor and Bennett 2009).

More comparative studies involving related species that differ with respect to their breeding systems (outcrossing sexuality *vs.* asexuality or predominance of self-fertilization) would also be valuable in seeking evidence for the effects of restricted recombination on variation and the efficacy of selection. For example, while it is well established that self-fertilizing plants generally have greatly reduced levels of within-population variability (Charlesworth 2003), it is less clear whether they show evidence of a reduced efficiency of selection, partly because they are often of comparatively recent origin. A recent study using population genetic data shows, however, that the effectiveness of selection on codon usage is greatly reduced in two different selfing species of plants compared with their outcrossing relatives (Qiu *et al.* 2011). Similarly, the efficacy of selection on nonsynonymous mutations is higher in the outcrossing plant *Capsella grandiflora* than in the related selfer *A. thaliana* (Slotte *et al.* 2010); in outcrossing *vs.* selfing species of *Neurospora* (Nygren *et al.* 2011); and in sexual *vs.* asexual lineages of *Daphnia* (Paland and Lynch 2006), freshwater snails (Johnson and Howard 2007), and rotifers (Barraclough *et al.* 2007). We are, however, far from having a comprehensive picture of the consequences of evolving low or zero rates of recombination throughout the genome.

In addition, it may be worth pointing out that the possible effects of Hill–Robertson interference due to purifying selection in microbial populations seem to be largely unstudied, other than for endosymbiont bacteria (Moran *et al.* 2008). Given the high mutation rates of RNA viruses (Drake *et al.* 1998), these organisms may well be vulnerable to the processes discussed here, as well as bacterial species that have only sporadic and patchy recombinational exchange among individuals (Didelot and Maiden 2010).

Finally, the theoretical and empirical evidence that restricted recombination among a large number of sites subject to purifying selection reduces the ability of selection to remove deleterious variants and fix favorable ones suggests that modifier alleles that promote nonzero levels of recombination will be selectively favored. This conjecture has been confirmed by computer simulations (*e.g.*, Keightley and Otto 2006; Gordo and Campos 2008; Charlesworth *et al.* 2010). The processes discussed here may therefore play a significant role in the evolution of sex and recombination.

Acknowledgments

I thank Nick Barton, Deborah Charlesworth, Isabel Gordo, Penny Haddrill, Matty Hartfield, Peter Keightley, Michael Lynch, Michael Turelli, Kai Zeng, and an anonymous reviewer for their comments on the manuscript and Michael Turelli for inviting me to write it.

Literature Cited

Aguadé, M., N. Miyashita, and C. H. Langley, 1989 Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics 122: 607–615.

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.

Andolfatto, P., and M. Przeworski, 2001 Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. Genetics 158: 657–665.

Arguello, J. R., Y. Zhang, T. Kado, C. Z. Fan, R. P. Zhao *et al.*, 2010 Recombination yet inefficient selection along the *Drosophila melanogaster* subgroup's fourth chromosome. Mol. Biol. Evol. 27: 848–861.

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., E. Hom, K. M. Wong, X. Maside, and P. De Jong, 2008 Genomic degradation of a young *Y* chromosome in *Drosophila miranda*. Genome Biol. 9: R30.

Barraclough, T. G., D. Fontaneto, C. Ricci, and E. A. Herniou, 2007 Evidence for inefficient selection against deleterious mutations in cytochrome oxidase I of asexual bdelloid rotifers. Mol. Biol. Evol. 24: 1952–1962.

Bartolomé, C., and B. Charlesworth, 2006 Evolution of amino-acid sequences and codon usage on the *Drosophila miranda* neo-sex chromosomes. Genetics 174: 2033–2044.

Barton, N. H., 1994 The reduction in fixation probability caused by substitutions at linked loci. Genet. Res. 64: 199–208.

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

Barton, N. H., and A. M. Etheridge, 2004 The effect of selection on genealogies. Genetics 166: 1115–1131.

Bateman, A. J., 1959 The viability of near-normal irradiated chromosomes. Int. J. Radiat. Biol. 1: 170–180.

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, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rate in *Drosophila melanogaster*. Nature 356: 519–520.

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., A. K. Holloway, K. Stevens, L. W. Hillier, Y.-P. Poh *et al.*, 2007 Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 5: e310.

Betancourt, A. J., J. J. Welch, and B. Charlesworth, 2009 Reduced effectiveness of selection caused by lack of recombination. Curr. Biol. 19: 655–660.

Birky, C. W., and J. B. Walsh, 1988 Effects of linkage on rates of molecular evolution. Proc. Natl. Acad. Sci. USA 85: 6414–6418.
Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan, 1995 The hitchiking effect on the site frequency spectrum of DNA polymorphism. Genetics 140: 783–796.
Brunet, E., I. M. Rouzine, and C. O. Wilke, 2008 The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. Theor. Popul. Biol. 73: 24–46.
Bulmer, M. G., 1991 The selection-mutation-drift theory of synonymous codon usage. Genetics 129: 897–907.
Cai, J. J., J. M. Macpherson, G. Sella, and D. A. Petrov, 2009 Pervasive hitchiking at coding and regulatory sites in humans. PLoS Genet. 5: e1000336.
Charlesworth, B., 1978 Model for evolution of *Y* chromosomes and dosage compensation. Proc. Natl. Acad. Sci. USA 75: 5618–5622.
Charlesworth, B., 1990 Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet. Res. 55: 199–221.
Charlesworth, B., 1994 The effect of background selection against deleterious alleles on weakly selected, linked variants. Genet. Res. 63: 213–228.
Charlesworth, B., 1996 Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. 68: 131–150.
Charlesworth, B., 1998 Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. 15: 538–543.
Charlesworth, B., and D. Charlesworth, 1997 Rapid fixation of deleterious alleles by Muller's ratchet. Genet. Res. 70: 63–73.
Charlesworth, B., and D. Charlesworth, 2010 *Elements of Evolutionary Genetics*. Roberts & Co., Greenwood Village, CO.
Charlesworth, B., and K. A. Hughes, 2000 The maintenance of genetic variation in life-history traits, pp. 369–392 in *Evolutionary Genetics from Molecules to Morphology*, edited by R. S. Singh and C. B. Krimbas. Cambridge University Press, Cambridge, UK.
Charlesworth, B., M. T. Morgan, and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289–1303.
Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. 70: 155–174.
Charlesworth, B., C. Bartolomé, and V. Noël, 2005 Estimating the incidence of ancestral polymorphisms. Genet. Res. 86: 149–157.
Charlesworth, B., A. J. Betancourt, V. B. Kaiser, and I. Gordo, 2010 Genetic recombination and molecular evolution. Cold Spring Harbor Symp. Quant. Biol. 74: 177–186.
Charlesworth, D., 2003 Effects of inbreeding on the genetic diversity of plant populations. Philos. Trans. R. Soc. B 358: 1051–1070.
Charlesworth, D., B. Charlesworth, and M. T. Morgan, 1995 The pattern of neutral molecular variation under the background selection model. Genetics 141: 1619–1632.
Comeron, J. M., and T. B. Guthrie, 2005 Intragenic Hill-Robertson interference influences selection on synonymous mutations in *Drosophila*. Mol. Biol. Evol. 22: 2519–2530.
Comeron, J. M., and M. Kreitman, 2002 Population, evolutionary and genomic consequences of interference selection. Genetics 161: 389–410.
Comeron, J. M., A. Williford, and R. M. Kliman, 2008 The Hill-Robertson effect: evolutionary consequences of weak selection in finite populations. Heredity 100: 19–31.
Crow, J. F., 1970 Genetic loads and the cost of natural selection, pp. 128–177 in *Mathematical Topics in Population Genetics*, edited by K. Kojima. Springer-Verlag, Berlin.
Crow, J. F., 1988 The importance of recombination, pp. 56–73 in *The Evolution of Sex*, edited by R. E. Michod and B. R. Levin. Sinauer Associates, Sunderland, MA.
Crow, J. F., 1990 R. A. Fisher, a centennial view. Genetics 124: 207–211.
Crow, J. F., 1993 Mutation, mean fitness, and genetic load. Oxf. Surv. Evol. Biol. 9: 3–42.
Crow, J. F., 2000 The origins, patterns and implications of human spontaneous mutation. Nat. Rev. Genet. 1: 40–47.
Crow, J. F., 2005 Timeline—Hermann Joseph Muller, evolutionist. Nat. Rev. Genet. 6: 941–945.
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.
Crow, J. F., and R. G. Temin, 1964 Evidence for the partial dominance of recessive lethal genes in natural populations of Drosophila. Am. Nat. 98: 21–33.
Cutter, A., and J. Y. Choi, 2010 Natural selection shapes nucleotide polymorphism across the genome of the nematode *Caenorhabditis briggsae*. Genome Res. 20: 1103–1111.
Cutter, A. D., and A. M. Moses, 2011 Polymorphism, divergence and the role of recombination in *Saccharomyces cerevisiae* genome evolution. Mol. Biol. Evol. 28: 1745–1754.
Didelot, X., and M. C. J. Maiden, 2010 Impact of recombination on bacterial evolution. Trends Microbiol. 18: 315–322.
Dobzhansky, T., and S. Wright, 1941 Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscura*. Genetics 26: 23–51.
Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow, 1998 Rates of spontaneous mutation. Genetics 148: 1667–1686.
Dye, C., and B. G. Williams, 1997 Multigenic drug resistance among inbred malaria parasites. Proc. Biol. Sci. 264: 61–67.
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.
Felsenstein, J., 1974 The evolutionary advantage of recombination. Genetics 78: 737–756.
Felsenstein, J., 1988 Sex and the evolution of recombination, pp. 74–86 in *The Evolution of Sex*, edited by R. E. Michod and B. R. Levin. Sinauer Associates, Sunderland, MA.
Fisher, R. A., 1930a The distribution of gene ratios for rare mutations. Proc. R. Soc. Edinb. 50: 205–220.
Fisher, R. A., 1930b *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.
Flowers, J. M., J. Molina, S. Rubinstein, P. Huang, B. A. Schaal *et al.*, 2012 Natural selection in gene dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. Mol. Biol. Evol. (in press).
Garcia-Dorado, A., and A. Caballero, 2000 On the average coefficient of dominance of deleterious spontaneous mutations. Genetics 155: 1991–2001.
Gessler, D. D. G., 1995 The constraints of finite size in asexual populations and the rate of the ratchet. Genet. Res. 66: 241–253.
Gordo, I., and R. A. Campos, 2008 Sex and deleterious mutations. Genetics 179: 621–626.
Gordo, I., and B. Charlesworth, 2000 On the speed of Muller's ratchet. Genetics 156: 2137–2140.
Gordo, I., and B. Charlesworth, 2001 The speed of Muller's ratchet with background selection, and the degeneration of *Y* chromosomes. Genet. Res. 78: 149–162.
Gordo, I., A. Navarro, and B. Charlesworth, 2002 Muller's ratchet and the pattern of variation at a neutral locus. Genetics 161: 835–848.
Gossmann, T. I., B.-H. Song, A. J. Windsor, T. Mitchell-Olds, C. J. Dixon *et al.*, 2010 Genome wide analyses reveal little evidence

for adaptive evolution in many plant species. Mol. Biol. Evol. 27: 1822–1832.

Gottipatti, S., L. Arbiza, A. Siepel, and A. G. Clark, 2011 Analyses of *X*-linked and autosomal genetic variation in population-scale whole genome sequencing. Nat. Genet. 43: 741–743.

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

Haddrill, P. R., K. R. Thornton, B. Charlesworth, and P. Andolfatto, 2005b Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Genome Res. 15: 790–799.

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.

Haigh, J., 1978 The accumulation of deleterious genes in a population. Theor. Popul. Biol. 14: 251–267.

Halligan, D. L., and P. D. Keightley, 2009 Spontaneous mutation accumulation studies in evolutionary genetics. Annu. Rev. Ecol. Evol. Syst. 40: 151–172.

Halligan, D. L., F. Oliver, A. Eyre-Walker, B. Harr, and P. D. Keightley, 2010 Evidence for pervasive adaptive protein evolution in wild mice. PLoS Genet. 6: e1000825.

Hamblin, M. T., and C. F. Aquadro, 1999 DNA sequence variation and the recombinational landscape in *Drosophila pseudoobscura*: a study of the second chromosome. Genetics 153: 859–869.

Hammer, M. F., A. E. Woerner, F. L. Mendez, J. C. Watkins, M. P. Cox *et al.*, 2010 The ratio of human *X* chromosome to autosome diversity is positively correlated with genetic distance from genes. Nat. Genet. 42: 830–831.

Hellmann, I., I. Ebersberger, S. E. Ptak, S. Paabo, and M. Przeworski, 2003 A neutral explanation for the correlation of diversity with recombination rates in humans. Am. J. Hum. Genet. 72: 1527–1535.

Hellmann, I., Y. Mang, G. Zhiping, P. Li, F. M. De La Vega *et al.*, 2008 Population genetic analysis of shotgun assemblies of genomic sequences from multiple individuals. Genome Res. 18: 1020–1029.

Hernandez, R., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton *et al.*, 2011 Classic selective sweeps were rare in recent human evolution. Science 331: 920–924.

Higgs, P. G., and G. Woodcock, 1995 The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection. J. Math. Biol. 33: 677–702.

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

Hudson, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. 7: 1–45.

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.

Innan, H., and W. Stephan, 2003 Distinguishing the hitchiking and background selection models. Genetics 165: 2307–2312.

Jain, K., 2008 Loss of least-loaded class in asexual populations due to drift and epistasis. Genetics 179: 2125–2134.

Jensen-Seaman, M. I., T. S. Furey, B. A. Payseur, Y. Lu, K. M. Roskin *et al.*, 2004 Comparative recombination rates in the rat, mouse and human genomes. Genome Res. 14: 528–538.

Johnson, S. G., and R. S. Howard, 2007 Contrasting patterns of synonymous and nonsynonymous sequence evolution in asexual and sexual freshwater snail lineages. Evolution 61: 2728–2735.

Johnson, T., and N. H. Barton, 2002 The effect of deleterious alleles on adaptation in asexual populations. Genetics 162: 395–411.

Johnson, T., and N. H. Barton, 2005 Theoretical models of selection and mutation on quantitative traits. Philos. Trans. R. Soc. B. 360: 1411–1425.

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

Kaiser, V. B., and B. Charlesworth, 2009 The effects of deleterious mutations on evolution in non-recombining genomes. Trends Genet. 25: 9–12.

Kaiser, V. B., and B. Charlesworth, 2010 Muller's ratchet and the degeneration of the *Drosophila miranda* neo-*Y* chromosome. Genetics 185: 339–348.

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

Kawabe, A., A. Forrest, S. I. Wright, and D. Charlesworth, 2008 High DNA sequence diversity in pericentric genes of the plant *Arabidopsis lyrata*. Genetics 179: 985–995.

Keightley, P. D., 2012 Rates and fitness consequences of new mutations in humans. Genetics (in press).

Keightley, P. D., and A. Eyre-Walker, 2010 What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? Philos. Trans. R. Soc. B 365: 1187–1193.

Keightley, P. D., and S. P. Otto, 2006 Interference among deleterious mutations favours sex and recombination in finite populations. Nature 443: 89–92.

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.

Keightley, P. D., L. Eory, D. L. Halligan, and M. Kirkpatrick, 2011 Inference of mutation parameters and selective constraint in mammalian coding sequences by approximate Bayesian computation. Genetics 187: 1153–1161.

Keinan, A., and D. Reich, 2010 Human population differentiation is strongly correlated with local recombination rate. PLoS Genet. 6: e1000886.

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

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kimura, M., and T. Maruyama, 1966 The mutational load with epistatic gene interactions in fitness. Genetics 54: 1303–1312.

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

King, J. L., 1966 The gene interaction component of the genetic load. Genetics 53: 403–413.

Kondrashov, A. S., 1982 Selection against harmful mutations in large sexual and asexual populations. Genet. Res. 44: 325–332.

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

Kondrashov, A. S., and J. F. Crow, 1993 A molecular approach to estimating the human deleterious mutation rate. Hum. Mutat. 2: 229–234.

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., 2006 Quantifying the genomic decay paradox due to Muller's ratchet in human mitochondrial DNA. Genet. Res. 87: 133–159.

Loewe, L., and B. Charlesworth, 2007 Background selection in single genes may explain patterns of codon bias. Genetics 175: 1381–1393.
Lohmuller, K. E., A. Albrechtsen, Y. Li, S. Y. Kim, T. Korneliussen *et al.*, 2011 Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. PLoS Genet. 7: e1002326.
Lynch, M., R. Bürger, D. Butcher, and W. Gabriel, 1993 The mutational meltdown in asexual populations. J. Hered. 84: 339–344.
Mallet, M. A., J. M. Bourchard, C. M. Kimber, and A. K. Chippindale, 2011 Experimental mutation-accumulation on the *X* chromosome of *Drosophila melanogaster* reveals stronger selection on males than females. BMC Evol. Biol. 11: 156.
Mallet, M. A., C. M. Kimber, and A. K. Chippindale, 2012 Susceptibility of the male fitness phenotype to spontaneous mutation. Biol. Lett. 8: (in press).
Manning, J. T., and D. J. Thompson, 1984 Muller's ratchet and the accumulation of favourable mutations. Acta Biotheor. 33: 219–225.
Maynard Smith, J., 1978 *The Evolution of Sex*. Cambridge University Press, Cambridge, UK.
Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23–35.
McVean, G. A. T., and B. Charlesworth, 2000 The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics 155: 929–944.
McVicker, G., D. Gordon, C. Davis, and P. Green, 2009 Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 5: e1000471.
Moran, N. A., J. P. Mccutcheon, and A. Nakabatchi, 2008 Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. 42: 165–190.
Mukai, T., 1964 The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. Genetics 50: 1–19.
Mukai, T., S. I. Chigusa, L. E. Mettler, and J. F. Crow, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. Genetics 72: 335–355.
Muller, H. J., 1932 Some genetic aspects of sex. Am. Nat. 66: 118–138.
Muller, H. J., 1949 Redintegration of the symposium on genetics, paleontology, and evolution, pp. 421–445 in *Genetics*, *Paleontology and Evolution*, edited by G. L. Jepsen, G. G. Simpson, and E. Mayr. Princeton University Press, Princeton, NJ.
Muller, H. J., 1950 Our load of mutations. Am. J. Hum. Genet. 2: 111–176.
Muller, H. J., 1964 The relation of recombination to mutational advance. Mutat. Res. 1: 2–9.
Noor, M. A. F., and S. M. Bennett, 2009 Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103: 439–444.
Nordborg, M., 1997 Structured coalescent processes on different time scales. Genetics 146: 1501–1514.
Nordborg, M., B. Charlesworth, and D. Charlesworth, 1996a The effect of recombination on background selection. Genet. Res. 67: 159–174.
Nordborg, M., B. Charlesworth, and D. Charlesworth, 1996b Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. Proc. Biol. Sci. 163: 1033–1039.
Nordborg, M., T. T. Hu, Y. Ishino, J. Jhaveri, C. Toomajian *et al.*, 2005 The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. 3: 1289–1299.
Nygren, K., R. Strandberg, A. Wallberg, B. Nabholz, T. Gustafsson *et al.*, 2011 A comprehensive phylogeny of *Neurospora* reveals a link between reproductive mode and molecular evolution in fungi. Mol. Phylogenet. Evol. 59: 649–663.
O'Fallon, B. D., J. Seger, and F. R. Adler, 2010 A continuous-state coalescent and the impact of weak selection on the structure of gene genealogies. Mol. Biol. Evol. 27: 1162–1172.
Paland, S., and M. Lynch, 2006 Transitions to asexuality result in excess amino-acid substitutions. Science 311: 990–992.
Pálson, S., 2001 The effects of deleterious mutations in cyclically parthenogenetic organisms. J. Theor. Biol. 208: 201–214.
Pálson, S., and P. Pamilo, 1999 The effect of deleterious mutations on linked, neutral variation in small populations. Genetics 153: 475–483.
Payseur, B. A., and M. Nachman, 2002 Natural selection at linked sites in humans. Gene 300: 31–42.
Peck, J., 1994 A ruby in the rubbish: beneficial mutations, deleterious mutations, and the evolution of sex. Genetics 137: 597–606.
Qin, H., W. B. Wu, J. M. Comeron, M. Kreitman, and W.-H. Li, 2004 Intragenic spatial patterns of codon usage in prokaryotic and eukaryotic genomes. Genetics 168: 2245–2260.
Qiu, S., K. Zeng, T. Slotte, S. I. Wright, and D. Charlesworth, 2011 Reduced efficacy of natural selection on codon usage bias in selfing Arabidopis and Capsella species. Genome Biol. Evol. 3: 838–880.
Rao, Y., L. Sun, Q. Nie, and X. Zhang, 2011 The influence of recombination on SNP diversity in chickens. Hereditas 148: 63–69.
Reed, F. A., J. M. Akey, and C. F. Aquadro, 2005 Fitting background selection models to levels of nucleotide variation and divergence along the human autosomes. Genome Res. 15: 1211–1221.
Rice, W. R., 1987 Genetic hitch-hiking and the evolution of reduced genetic activity of the *Y* sex chromosome. Genetics 116: 161–167.
Robertson, A., 1961 Inbreeding in artificial selection programmes. Genet. Res. 2: 189–194.
Rockman, M. V., S. S. Skrovanek, and L. Kruglyak, 2010 Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. Science 330: 372–376.
Roselius, K., W. Stephan, and T. Städler, 2005 The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. Genetics 171: 753–763.
Santiago, E., and A. Caballero, 1998 Effective size and polymorphism of linked neutral loci in populations under selection. Genetics 149: 2105–2117.
Sattath, S., E. Elyashiv, O. Kolodny, Y. Rinott, and G. Sella, 2011 Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in Drosophila simulans. PLoS Genet. 7: e1001302.
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.
Seger, J., W. A. Smith, J. J. Perry, K. Hunn, Z. A. Kaliszewska *et al.*, 2010 Gene genealogies are distorted by weakly interfering mutations in constant environments. Genetics 184: 529–545.
Sella, G., D. A. Petrov, M. Przeworski, and P. Andolfatto, 2009 Pervasive natural selection in the Drosophila genome? PLoS Genet. 6: e1000495.
Shapiro, J. A., W. Huang, C. Zhang, M. Hubisz, J. Lu *et al.*, 2007 Adaptive genic evolution in the Drosophila genome. Proc. Natl. Acad. Sci. USA 104: 2271–2276.
Sharp, P. M., L. B. Emery, and K. Zeng, 2010 Forces that influence the evolution of codon bias. Philos. Trans. R. Soc. B 365: 1203–1212.
Shnol, E. E., and A. S. Kondrashov, 1993 Effect of selection on the phenotypic variance. Genetics 134: 995–996.
Simmons, M. J., and J. F. Crow, 1977 Mutations affecting fitness in Drosophila populations. Annu. Rev. Genet. 11: 49–78.

Simonsen, K. L., G. A. Churchill, and C. F. Aquadro, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141: 413–429.
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.
Slotte, T., J. P. Foxe, K. M. Hazzouri, and S. I. Wright, 2010 Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. Mol. Biol. Evol. 27: 1813–1821.
Söderberg, R. J., and O. G. Berg, 2007 Mutational interference and the progression of Muller's ratchet when mutations have a broad range of deleterious effects. Genetics 177: 971–986.
Stajich, J. E., and M. W. Hahn, 2005 Disentangling the effects of demography and selection in human history. Mol. Biol. Evol. 22: 63–73.
Stephan, W., 1995 An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorphism data. Mol. Biol. Evol. 12: 959–962.
Stephan, W., 2010 Genetic hitchiking *vs.* background selection: the controversy and its implications. Philos. Trans. R. Soc. B 365: 1245–1253.
Stephan, W., and Y. Kim, 2002 Recent applications of diffusion theory to population genetics, pp. 72–93 in *Modern Developments in Theoretical Population Genetics*, edited by M. Slatkin and M. Veuille. Oxford University Press, Oxford.
Stephan, W., and C. H. Langley, 1989 Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. Genetics 121: 89–99.
Stephan, W., T. H. E. Wiehe, and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. 41: 237–254.
Stephan, W., B. Charlesworth, and G. A. T. McVean, 1999 The effect of background selection at a single locus on weakly selected, partially linked variants. Genet. Res. 73: 133–146.
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.
Tachida, H., 2000 DNA evolution under weak selection. Gene 261: 3–9.
Timofeeff-Ressovsky, N. W., 1940 Mutations and geographical variation, pp. 73–136 in *The New Systematics*, edited by J. S. Huxley. The Systematics Association, London.
Wiehe, T. H. E., and W. Stephan, 1993 Analysis of a genetic hitchiking model and its application to DNA polymorphism data. Mol. Biol. Evol. 10: 842–854.
Williamson, S. M., and M. E. Orive, 2002 The genealogy of a sequence subject to purifying selection at multiple sites. Mol. Biol. Evol. 19: 1376–1384.
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.
Wright, S., 1938 Size of population and breeding structure in relation to evolution. Science 87: 430–431.
Zeng, K., and B. Charlesworth, 2010a The effects of demography and linkage on the estimation of selection and mutation parameters. Genetics 186: 1411–1424.
Zeng, K., and B. Charlesworth, 2010b 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: M. Turelli

Appendix

Consider a population with N diploid breeding adults, a 1:1 sex ratio, and purely random variation in offspring number other than that contributed by the presence of deleterious mutations. At a single site i subject to mutation–selection balance, mutations from wild-type to deleterious alleles arise at rate u_i per generation and are subject to a selection coefficient t_i. Recombination between the neutral site and the selected site takes place at rate r_i. This is effective only in heterozygous carriers of deleterious mutations, which have fitness $1 - t_i$; since the population mean fitness is equal to $1 - 2u_i$ (*cf.* Equation 2), only a fraction $(1 - t_i)/(1 - 2u_i)$ of these individuals contribute to the next generation, so that the effective rate of recombination is $r_i' = r_i(1 - t_i)/(1 - 2u_i) \approx r_i(1 - t_i)$.

Consider the effect of selection and recombination on a neutral variant that is initially present in the same individual as a deleterious mutation. Two possible initial states need to be considered in the case of a diploid population (Santiago and Caballero 1998). First, the neutral variant is in coupling with the deleterious variant. The rate at which it either is eliminated from the population by selection or loses its association with the mutant allele by recombination is $t_i + r_i'$, so that the fraction of descendant gametes that still carry the deleterious allele n generations later is $(1 - t_i - r_i')^n$. In the initial generation ($n = 0$), the association of a neutral variant with the deleterious allele leads to an expected change of $-t_i$ in its representation in the population; there will be a further expected change of $-t_i(1 - t_i - r_i')$ after one generation, and so on, so that the net change after n generations is the sum of $-t_i(1 - t_i - r_i')^n$; summing this geometric progression from $n = 0$ to ∞, the expected net reduction in frequency over all generations is $\delta_{c_i} = t_i/(t_i + r_i')$.

The second case is when the neutral variant is in repulsion with the deleterious variant. In the initial generation, there is again an expected reduction in frequency of $-t_i$. In the next generation, the chance that the deleterious mutation becomes associated with the neutral variant is r_i'; there is then a further change in expected frequency of $-t_i r_i'$. For all subsequent generations, the argument applied to the coupling case can be applied, but all terms are multiplied by r_i'. The net frequency reduction in this case is $\delta_{r_i} = t_i(1 + r_i'/[t_i + r_i'])$.

Following Robertson (1961), the δ's can be regarded as the asymptotic reductions in fitness associated with the original neutral variant. Given that a fraction q_i of the initial population carries the mutant allele, these associations therefore contribute asympotic additive genetic variances at the neutral site of $2q_i\delta_{c_i}^2$ and $2q_i\delta_{r_i}^2$, respectively,

assuming that q_i is so low that ${q_i}^2$ is negligible (Mukai *et al.* 1972). The effect of these associations on N_e can be found from the standard formula for the ratio N/N_e when there is an excess variance over random expectation in offspring number per individual, ΔV_i (measured relative to the square of the population mean): $N/N_e \approx (1 + \Delta V_i)$ (Wright 1938; Charlesworth and Charlesworth 2010, p. 222). Only one-half of the progeny of carriers of the mutant allele suffer a loss in fitness, so that $\Delta V_i \approx q_i({\delta_{c_i}}^2 + {\delta_{r_i}}^2)$.

This can be extended to multiple sites by using the fact that $1 + \Delta V_i$ is the expectation of the square of the effect of variation at site i on per capita offspring number, scaled relative to the population mean (Nordborg *et al.* 1996a; Santiago and Caballero 1998). With independence among sites and multiplicative fitness effects, the expectation of the squared effects over all sites is given by the product of the expectations for each site (Nordborg *et al.* 1996a; Santiago and Caballero 1998), so that $N/N_e = \prod_i\{1 + q_i({\delta_{c_i}}^2 + {\delta_{r_i}}^2)\}$, giving $N_e/N \approx \prod_i\{1 - q_i({\delta_{c_i}}^2 + {\delta_{r_i}}^2)\}$. When r_i and t_i are both small, $N_e/N \approx \prod_i(1 - q_i{t_i}^2/[t_i + {r_i}']^2) \approx \exp(-\sum_i q_i{t_i}^2/[t_i + {r_i}']^2)$. By using the expression $q_i = u_i/t_i$ for the equilibrium frequencies of mutations, and setting $T_2 = 2N_e$, we obtain Equation 3. With free recombination, $q_i({\delta_{c_i}}^2 + {\delta_{r_i}}^2) \sim 8q_i{t_i}^2$, which yields Equation 4.