Genetic Recombination and Molecular Evolution

B. Charlesworth,[1] A.J. Betancourt,[1] V.B. Kaiser,[1] and I. Gordo[2]
[1]*Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JT, United Kingdom;* [2]*Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal*
Correspondence: Brian.Charlesworth@ed.ac.uk

Reduced rates of genetic recombination are often associated with reduced genetic variability and levels of adaptation. Several different evolutionary processes, collectively known as Hill–Robertson (HR) effects, have been proposed as causes of these correlates of recombination. Here, we use DNA sequence polymorphism and divergence data from the non-crossing-over dot chromosome of *Drosophila* to discriminate between two of the major forms of HR effects: selective sweeps and background selection. This chromosome shows reduced levels of silent variability and reduced effectiveness of selection. We show that neither model fits the data on variability. We propose that in large genomic regions with restricted recombination, HR effects among nonsynonymous mutations undermine the effective strength of selection, so that their background selection effects are weakened. This modified model fits the data on variability and also explains why variability in very large nonrecombining genomes is not completely wiped out. We also show that HR effects of this type can produce an individual selection advantage to recombination, as well as greatly reduce the mean fitness of nonrecombining genomes and genomic regions.

In eukaryotes, the disjunction of homologous centromeres in the first division of meiosis results in the independent assortment of genes on different chromosomes; recombinational exchange (gene conversion and reciprocal crossing-over) reshuffles the genetic material between the homologous chromosomes contributed by the two parents. These processes cause different sites in the genome to have more or less distinct ancestries, unless recombination is absent or ineffective. Recombination can therefore have important consequences for the effectiveness of selection (Barton, **this volume**). In particular, Fisher (1930, p.103) pointed out that there may be an evolutionary cost to recombination. When two loci are each polymorphic for two alleles with epistatic fitness effects that create linkage disequilibrium (LD), recombination reduces the frequencies of the selectively favorable combinations of alleles. He concluded that if there is genetic variability in the frequency of recombination, this "…will always tend to diminish recombination, and therefore to increase the intensity of linkage in the chromosomes… ." Subsequent theoretical work has put this verbal argument on a firm theoretical basis (Zhivotovsky et al. 1994; Otto and Lenormand 2002).

Why, therefore, does the genome not "congeal" to a state of zero recombination (Turner 1967)? There is evidence pointing to a countervailing selective advantage to recombination; for example, several ecological factors correlate with rates of crossing-over. Mammalian species with long development times tend to have higher rates of crossing-over per chromosome than fast-developing species (Burt and Bell 1987; Sharp and Hayman 1988), and highly self-fertilizing species of plants tend to have higher rates of cytologically detectable crossovers than related outcrossing species (Roze and Lenormand 2005). Recombination rates therefore appear to vary in response to selective pressures, so that we need to search for population genetic processes that relate recombination to higher fitness. Several credible candidates for this have been identified (Barton, **this volume**). The challenge is to identify biological patterns that indicate that the level of recombination has evolutionary consequences and that also shed light on which processes may be involved.

We argue that a variety of lines of evidence suggest that, as proposed by Felsenstein (1974), Hill–Robertson (HR) effects have a major role in causing the evolutionary effects of recombination. Hill and Robertson (1966) showed that selection at one site in the genome impedes the action of selection at another site, especially when recombination between them is rare or absent (**Fig. 1**). This is because a finite population cannot contain all possible combinations of variants at different sites. If mutations arise in different individuals, a favorable variant at one site will generally be present in a genotype with a deleterious variant at another site, i.e., negative LD exists among selectively favorable variants. Because recombination breaks down this LD, it enhances the population's ability to respond to selection. **Table 1** lists the main types of HR effects.

A useful way to understand HR effects is to consider them in terms of the effective population size N_e. This is essentially the number of individuals in the population that successfully transmits genes to the next generation and is often much smaller than the number of individuals of breeding age (Wright 1931; Charlesworth 2009). Selection implies the existence of heritable variance in fitness among individuals; this reduces N_e because genes are preferentially transmitted through the fittest members of the population (Robertson 1961). A nucleotide site that is closely linked to another site with variants that are under selection experiences an especially large effect,

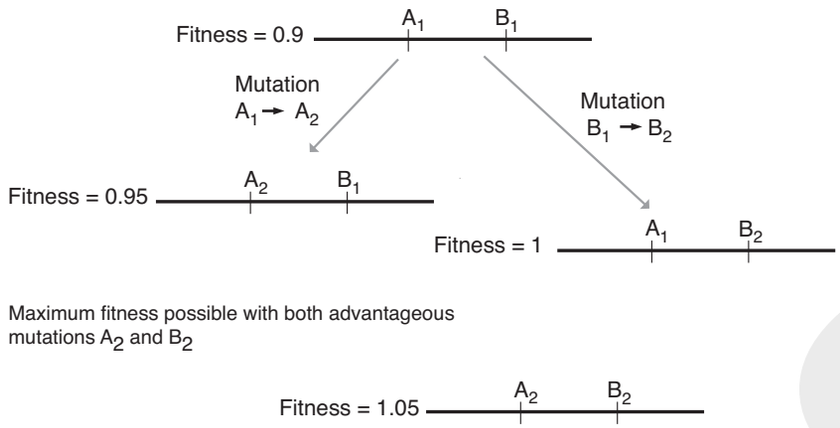

Figure 1. Hill–Robertson interference between the action of selection for advantageous mutations at two sites in a genome when recombination between them is rare or absent. The diagram shows two loci or sites, A and B, at which advantageous mutations arise in a haploid population, and each of which increases fitness. The highest fitness is achieved when both advantageous mutations are present in an individual.

because the influence of the variance in fitness at one site on the behavior of closely linked sites is maintained for many generations (Santiago and Caballero 1998; Barton, **this volume**).

The equilibrium level of diversity at neutral nucleotide sites is equal to the product of $4N_e$ and the mutation rate u provided that $4N_e u << 1$ (Kimura 1971). In addition, the probability that genetic drift fixes a deleterious mutation that reduces the fitness of its homozygous carriers by s (the selection coefficient) is close to the value for a neutral mutation when $N_e s < 1$ but is negligible when $N_e s >> 1$ (Fisher 1930, ch. 5; Kimura 1962). Similarly, the chance that a selectively favorable mutation with selective advantage s becomes established in a population is close to the neutral value when $N_e s << 1$ but approaches the value for an infinitely large population when $N_e s > 1$. Not all features of HR effects can be interpreted simply in terms of a reduction in N_e (Comeron et al. 2008; Kaiser and Charlesworth 2009), but it nevertheless provides a useful heuristic for their interpretation.

Reduced N_e caused by HR effects is thus expected to cause a reduction in the level of variability with respect to neutral or nearly neutral nucleotide variants. It will also cause loci to accumulate more slightly deleterious mutations, and fix fewer advantageous ones, than when HR effects are absent. The following patterns that are

Table 1. Main categories of HR effects

1. Interference by Favorable Mutations (Selective Sweeps)
The spread of a favorable mutation drags to fixation any closely linked neutral or deleterious mutant alleles initially associated with it, so that successive adaptive substitutions in a low recombination region of the genome can lead to a loss in neutral or nearly neutral variability and the fixation of slightly deleterious mutations at many loci. In addition, the spread of a favorable mutation at one locus can prevent the spread of a favorable mutation at another, closely linked locus.
2. Interference by Deleterious Mutations (Background Selection)
Deleterious mutations are assumed to enter the population at sites distributed over the genomic region in question and to be removed by selection with near certainty. A neutral or weakly selected mutation that arises in a nonrecombining section of the genome has a nonzero chance of survival only if it arises on a chromosome free of these mutations. This accelerates the fixation of weakly deleterious mutations and retards the fixation of advantageous mutations.
3. Muller's Ratchet
This involves the stochastic loss from a finite population of the class of chromosomes carrying the fewest deleterious mutations. In the absence of recombination and back mutation, this class of chromosome cannot be restored. The next best class then replaces it and is in turn lost in a process of successive irreversible steps. Each such loss is quickly followed by fixation of a deleterious mutation on the chromosome. Mutations at most sites remain close to their equilibrium frequencies.
4. Mutual Interference among Weakly Selected Sites (Weak HR Effects)
With a very large number of closely linked sites, subject to reversible mutation between favored and disfavored alleles, the mean level of adaptation can be strongly reduced in nonrecombining regions. This is because a mutual interference exists among the different sites under selection, allowing selectively deleterious variants to become much more frequent than expected under mutation-selection equilibrium frequencies as a result of genetic drift.

For further details, see Otto and Lenormand (2002), Comeron et al. (2008), Charlesworth and Charlesworth (2009, ch. 10), and Barton (**this volume**).

consistent with these expectations have been uncovered.

1. Regions of the genome with low levels of genetic recombination often show low levels of genetic diversity (see, e.g., Presgraves 2005).
2. Species with low levels of genome-wide recombination, such as highly self-fertilizing species of animals and plants, also show reduced genetic diversity (see, e.g., Charlesworth 2003).
3. These reductions in diversity are often associated with reduced levels of adaptation at the molecular level (see, e.g., Presgraves 2005; Moran et al. 2008).

We examine in detail here some examples of such evidence from our recent work with *Drosophila*. In addition, we describe recent theoretical work that helps to resolve some contradictions between the theoretical predictions and the data

THE RELATION OF RECOMBINATION TO GENETIC VARIATION

Previous Work with *Drosophila*

About 20 years ago, it was found that within-population variability was unusually low in regions of the *Drosophila* genome with low levels of crossing-over (Aguadé et al. 1989; Stephan and Langley 1989). Begun and Aquadro (1992) showed that a high correlation exists between the estimated level of variability in a gene and the local rate of recombination determined from the standard genetic map, whereas divergence at silent sites showed no relation to recombination rates. These observations have been replicated in more recent studies, for example, by Presgraves (2005).

Some form of HR effect seems to be the only credible explanation of these patterns. Begun and Aquadro (1992) favored hitchhiking effects caused by the spread of favorable mutations (Maynard Smith and Haigh 1974), often now referred to as "selective sweeps" (Berry et al. 1991). Stephan (1995) showed that it is possible to fit the observed relation between recombination rate and level of variability by this model; this has recently been extended to large *Drosophila melanogaster* polymorphism data sets (Andolfatto 2007). However, the alternative mode of hitchhiking by selection against recurrent deleterious mutations ("background selection," Table 1) also fits the data on the relation between levels of variability and local recombination rates (Charlesworth 1996). Attempts to discriminate between selective sweeps and background selection have largely been inconclusive.

Genetic Diversity on the Dot Chromosome of *Drosophila Americana*

We have recently revisited this question (Betancourt et al. 2009), using the close relative of *D. virilis*, *D. americana*, for a survey of within-species variation and divergence among species at 14 genes on the small "dot" chromosome (Muller's element F), a chromosome that shows highly reduced levels of crossing-over (Ashburner et al. 2005). A data set on variability at 18 genes on other chromosomes is available for the purpose of comparison (Maside and Charlesworth 2007). Genetic data show that there is little population subdivision in *D. americana* and little evidence for demographic effects such as population expansion that complicate the interpretation of population genetic data (McAllister 2002; Maside and Charlesworth 2007); this makes it appropriate material for population genetic studies.

Silent nucleotide site diversity on the *D. americana* dot chromosome is about 17 times lower than the genome-wide average, a reduction in variability similar to that in other *Drosophila* species (Berry et al. 1991; Jensen et al. 2002; Wang et al. 2002, 2004; Sheldahl et al. 2003). As expected, the polymorphism data show that dot chromosome loci have a very low, but nonzero, recombination rate, as was seen in the other species. These recombination events are probably due to gene conversion rather than crossovers (Langley et al. 2000; Gay et al. 2007).

Interpretation of the Results

Coalescent simulations show that a recent selective sweep is not compatible with the observed distribution of frequencies of nucleotide site variants on the *D. americana* dot chromosome (Betancourt et al. 2009): There are too many intermediate-frequency variants compared with what is expected after a selective sweep on a nonrecombining chromosome (Braverman et al. 1995; Simonsen et al. 1995). Can background selection explain these data? The expected reduction in diversity for the dot chromosome can be determined from the classical background selection equation (Hudson and Kaplan 1995; Nordborg et al. 1996). For this purpose, we need to know the distribution of selection coefficients against deleterious mutations. Estimates of the parameters of this distribution for nonsynonymous mutations can be obtained from polymorphism data (Loewe and Charlesworth 2006; Loewe et al. 2006; Keightley and Eyre-Walker 2007; Sawyer et al. 2007). These data show that there is a wide distribution of the selective effects of deleterious mutations but the mean selection coefficient against a segregating amino acid mutation is extremely small (of the order of 10^{-5} to 10^{-4}). The classical background selection model with these estimates predicts that the dot chromosome should have ~1000-fold lower variation than the other autosomes (Loewe and Charlesworth 2007), rather than the 17-fold lower value in *D. americana*, and similar values for the other cases cited above.

EXPLAINING THE OBSERVED PATTERNS OF VARIABILITY

Reformulating the Background Selection Model

Both of the standard explanations for reduced variability in regions with low levels of recombination appear to be incompatible with our data on patterns of variability on

the dot chromosome. To try and resolve this paradox, we have reexamined the theory of background selection in a large, low recombination genomic region. The standard model assumes that the frequencies of deleterious mutant variants involved are close to those expected under mutation-selection balance equilibrium in an infinite population (Charlesworth et al. 1993; Hudson and Kaplan 1995; Nordborg et al. 1996). When recombination rates are extremely low, however, the model predicts a larger reduction in variability than is found in simulations (Charlesworth et al. 1993; Nordborg et al. 1996; Gordo et al. 2002). This suggests that low recombination may cause HR interference among the sites involved (for which $N_e s > 1$ when there are no HR effects), undermining the effectiveness of selection on these sites and causing the frequencies of deleterious mutations to drift up to much higher values than with mutation-selection balance. HR interference of this type has been studied previously by Monte Carlo simulations of selection at many linked sites (for review, see Comeron et al. 2008). These studies used fixed selection coefficients at each site and were designed primarily to model weak selection on codon usage (except for Tachida 2000).

In view of the recent increase in our knowledge of the intensity of purifying selection on nonsynonymous mutations (see above), it seemed important to model HR effects with realistic selective effects to see whether they can explain the *Drosophila* data on variability in genomic regions that lack crossing-over. We have performed Monte Carlo simulations of randomly mating populations with either normal or reduced rates of recombination (Kaiser and Charlesworth 2009). Haploid populations of 1000 individuals (equivalent to $N = 500$ diploids) were modeled. The state of a site under selection is represented as 1 or 0, where 0 is wild type and 1 is a deleterious alternative variant (**Fig. 2**). The fitness effect of a mutation at the ith chromosomal site under selection is denoted by s_i; the fitness, w, of an individual carrying a set of mutations is given by the standard multiplicative model (Haldane 1937), such that $\ln(w) = \Sigma_i \ln(1 - s_i)$.

Pairs of adjacent selected sites followed by neutral sites were distributed along the chromosome, with the total number of sites (L) varying between simulations. The selected sites correspond to first and second codon positions, where all mutations were assumed to be nonsynonymous; the neutral sites correspond to third codon positions that experience only synonymous mutations. Mutations at both types of sites arise at a rate u per base pair in each direction. This reversible mutation model applies to nucleotide mutations and allows the population to reach statistical equilibrium between drift, mutation, and selection (McVean and Charlesworth 2000). When the evolutionary forces are all weak, measures of the deterministic forces scaled by multiplying their values by N_e completely describe the system if time is measured in units of N_e generations (Ewens 2004). We can thus infer the behavior of large natural populations from our small simulated populations by using these scaled parameter values (McVean and Charlesworth 2000).

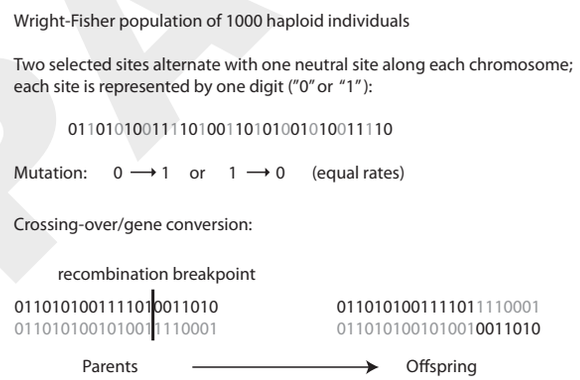

Figure 2. Schematic view of the representation of chromosomes and the simulation methods.

We chose mutation rates, recombination rates between adjacent sites, and a distribution of selection coefficients such that the products of N and the relevant parameter values are similar to those for genes in regions of normal recombination in a *Drosophila* population (Loewe and Charlesworth 2007). As a measure of the effectiveness of background selection, we used the ratio B of the mean pairwise diversity at the simulated neutral sites, relative to the theoretical equilibrium value for a population free of HR effects ($\pi = 4Nu$; Kimura 1971). **Figure 3** (top) shows the effects of selection at the "nonsynonymous" sites on B. For noncrossover regions, there is a rapid initial decline of B with L, but B is always much larger than predicted by the classical background selection formula (Fig. 3, bottom) (Hudson and Kaplan 1995; Nordborg et al. 1996). Importantly, when there is no crossing-over, B levels off at a value of ~0.015 for $L > 640{,}000$ sites. This suggests that the HR effects between sites under selection progressively undermine the effectiveness of selection as more selected sites are packed into a region where crossing-over is absent, so that additional selected sites eventually have no further effect on variability at the linked neutral sites.

Selection also distorts the gene genealogies at linked sites, especially when a large number of sites are under selection (Gordo et al. 2002; Williamson and Orive 2002), so that the reduction in N_e is not a complete descriptor of HR effects, as mentioned earlier. This distortion can be examined using Tajima's D_T statistic, which measures the difference between the estimate of variability from the mean number of sequence differences between all pairs of alleles in a sample and the estimate from the number of segregating sites in the sample (Tajima 1989). This has been used to test for selective sweeps, because these are expected to cause negative D_T values, reflecting an excess of rare variants (Braverman et al. 1995; Simonsen et al. 1995). Our simulations show that D_T for neutral sites is negative and increases in magnitude with L; with no crossing-over, it approaches its maximum value with the largest L values that we have simulated (Kaiser and Charlesworth 2009).

How do the simulation results relate to observations on genomic regions with low levels of recombination? As described in the previous section, the observed mean diver-

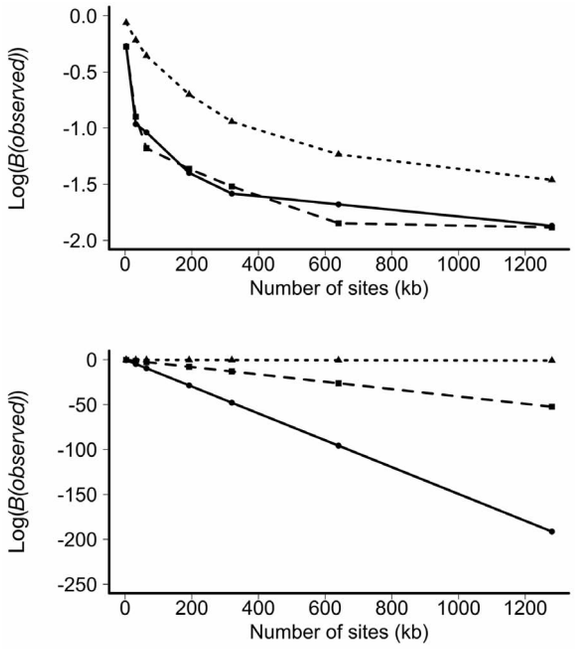

Figure 3. Effects of Hill–Robertson interference among strongly selected mutations on levels of diversity at linked neutral sites. (Dotted line) Free recombination, (dashed line) gene conversion only, (solid line) no recombination. (*Top*) $B(observed) = \pi/(4Nu)$ plotted against the number of sites. $B(observed)$ decreases with an increasing number of sites L. An asymptotic value of ~1.5% is reached for large values of L. (*Bottom*) The logarithm to base 10 of $B(expected)$ as a function of L, where $B(expected)$ is calculated from the standard background selection formula. With an increasing number of sites, neutral diversity is expected to decline exponentially, and the rate of decline is greater if recombination rates are low.

sity value on the dot chromosome in several *Drosophila* species is ~5% of the genome-wide average, close to the value of 6.5% in our simulations of a non-crossing-over region of this size (Fig. 3, top). In addition, in *D. americana*, the observed value of D_T is not significantly different from the predicted value for a noncrossover region of this size (Kaiser and Charlesworth 2009).

Similarly, the neo-Y chromosome of *D. miranda* completely lacks recombination, and it contains ~3.7 Mb of coding sequence, of which approximately one-half are nonfunctional and thus unlikely to cause HR effects (Bachtrog et al. 2008). The mean average silent-site diversity for genes on the *D. miranda* neo-Y is ~1% of the value for their homologs on the recombining neo-X chromosome (Bartolomé and Charlesworth 2006), which is quite close to the predicted value with large L (Kaiser and Charlesworth 2009). There is a large negative D_T for the neo-Y (Bartolomé and Charlesworth 2006), which has been interpreted as having been caused by a recent selective sweep (Bachtrog 2004). However, this D_T value is also consistent with our modified model of background selection (Kaiser and Charlesworth 2009).

The simulations also show that the frequencies of sites fixed for deleterious nonsynonymous variants are greatly increased in large regions with reduced recombination, with a corresponding reduction in the mean fitness of the population (**Table 2**), as found previously in HR models with weak selection and fixed selection coefficients (McVean and Charlesworth 2000; Comeron et al. 2008). Even a small nonrecombining genomic region can experience a noticeable reduction in its mean fitness, as illustrated in **Figure 4**, where a population with a nonrecombining chromosome of only 32 kb in length (21,333 sites under selection) equilibrates at a natural logarithm of mean fitness of –7.78. If we retain the same $N_e s$ values but rescale the population size to 1 million, which is reasonable for a *Drosophila* population, most selection coefficients are very small, so that the expression for $\ln(w)$ given above is well approximated by $-\Sigma_i s_i$. Using this simplification, the mean of log fitness in the simulations corresponds to a log mean fitness for *Drosophila* of $-7.78 \times 500/1000000 = -0.0039$, whereas the equilibrium log mean fitness of a freely recombining population with the same parameters is –0.0001, using the for-

Table 2. Effect of zero recombination on the proportion of selected sites with deleterious mutations and on the mean fitness of the population

Length of chromosome (kb)	Proportion of sites carrying a deleterious mutation[a]	Mean s at fixed sites[b] ($\times 10^5$)	Log fitness[c]	Relative reduction in log fitness[d]
3.2	0.00104	0.16	0	0
32	0.0168	0.63	–0.002	0.002
64	0.0254	0.83	–0.009	0.009
192	0.0377	1.19	–0.057	0.056
320	0.0464	1.35	–0.134	0.131
640	0.0533	1.58	–0.360	0.358
1280	0.0620	1.80	–0.952	0.948

[a]This is for a chromosome chosen randomly from the population after 10,000 generations.

[b]Mean s is scaled to a population size of 10^6 (the simulation value is multiplied by $500/10^6$).

[c]This is the negative of the number of nonsynonymous sites (two-thirds of column one) multiplied by the product of columns two and three. It provides a lower bound to the equilibrium reduction in fitness per nonsynonymous site, because the mean selection coefficient at sites that are fixed is lower than average.

[d]This is the difference in the estimated log fitness for the nonrecombining chromosome and the equilibrium log fitness for a freely recombining chromosome. The latter is calculated from the product of the mutation rate per site by the number of nonsynonymous sites (Haldane 1937).

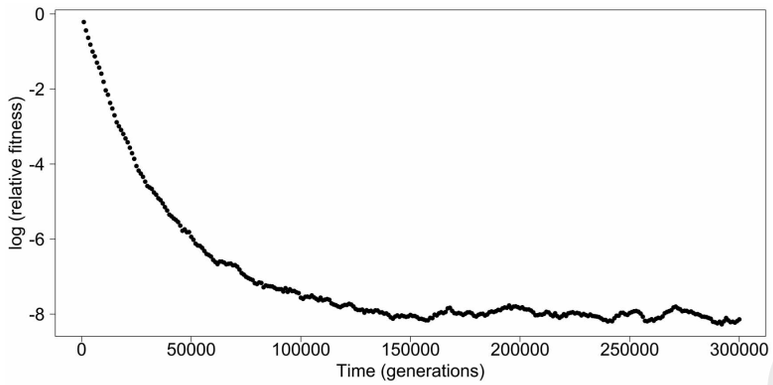

Figure 4. The decline in the natural logarithm of the mean fitness of a population without recombination, relative to the expected fitness under free recombination, plotted against the number of generations since the start (when the population was at equilibrium), for simulations without recombination and a chromosome length of 32 kb. The initial decline is linear, similar to what is found with Muller's ratchet, but eventually, statistical equilibrium is reached.

mula of Haldane (1937). The relative reduction in mean fitness is thus ~0.0038.

It is harder to estimate this reduction for cases with large numbers of sites, because the time to reach equilibrium with respect to mean fitness becomes very long. Table 2 shows some estimates of this reduction. With large L, there is a large reduction in mean fitness for a nonrecombining population, with the mean fitness approaching 39% of the free recombination value. The equilibrium mean fitness of a completely asexual population of a higher eukaryote, where a whole genome of more than 1 million nonsynonymous sites is nonrecombining, must therefore be extremely small, so that this effect may contribute to the apparent long-term evolutionary disadvantages of asexual lineages (Maynard Smith 1978; Bell 1982; Otto and Lenormand 2002; Normark et al. 2003).

EVIDENCE FOR A REDUCED EFFECTIVENESS OF SELECTION IN LOW RECOMBINATION GENOMIC REGIONS

These results lead naturally to the question of the extent to which levels of adaptation at the protein and DNA sequence level are indeed reduced in low recombination genomes or genomic regions. A severe reduction in effective population size should be reflected in a reduction in the efficacy of both purifying selection and positive selection, as well as a reduction in diversity. Various studies have found evidence for such effects, primarily in nonrecombining genomes or chromosomes. For example, codon usage bias is reduced in regions of the *Drosophila* genome that lack crossing-over, especially the dot chromosome (Kliman and Hey 1993, 2003; Marais et al. 2003; Haddrill et al. 2007), as we also found for the *D. americana* dot chromosome (Betancourt et al. 2009).

Studies of the relationships between local recombination rate and protein sequence variation and between-species divergence in *Drosophila* (Betancourt and Presgraves 2002; Presgraves 2005; Haddrill et al. 2007) also suggest that recombination affects the efficiency of selection on amino acid sequences, with reduced rates of adaptive evolution in regions of low recombination and relaxed selection against deleterious mutations. We used our *D. americana* data to ask if dot chromosome loci experience relaxed purifying selection on amino acid mutations. As expected, protein sequence divergence between *D. americana* and two related species, *D. virilis* and *D. ezoana*, is elevated on the dot chromosome (Betancourt et al. 2009).

This pattern can be explained by relaxed purifying selection on the dot chromosome causing the fixation of slightly deleterious amino acid variants, but it is also possible that it is due to a higher rate of fixation of beneficial mutations, although this is contrary to theoretical expectations. We can distinguish between these alternatives using the *D. americana* polymorphism data, because elevated levels of protein polymorphism relative to silent polymorphism indicate relaxed purifying selection: The *D. americana* dot chromosome loci indeed show such an effect (Betancourt et al. 2009). The average strength of purifying selection acting on the heterozygous carriers of segregating amino acid variants was estimated to be $N_e s \approx 4$ for dot loci and $N_e s \approx 29$ for nondot autosomal loci. Provided the two sets of loci experience similar levels of functional constraints, these results suggest that selection against amino acid variants is less effective in low recombination regions.

We also investigated whether adaptive evolution is similarly compromised on the *D. americana* dot chromosome. We used the combination of polymorphism and divergence data to estimate α, the proportion of nonsynonymous between-species differences caused by positive selection (Fay et al. 2002; Smith and Eyre-Walker 2002; Bierne and Eyre-Walker 2004). Using a maximum likelihood implementation of this method (Welch 2006), we found a significantly lower α estimate for the dot chromosome loci than for the other genes. This not only suggests that adaptive evolution may be compromised on the dot chromosome, but also excludes the possibility that the elevated protein sequence for dot chromosome loci is due to more frequent selectively driven substitutions.

DISCUSSION

The data that we have presented on sequence divergence and polymorphism on the dot chromosome of *D. americana* show that both levels of variability and the effectiveness of selection on protein sequences and codon usage are significantly reduced. These findings are in agreement with those previously reported for this small (~80 genes), non-crossing-over component of the *Drosophila* genome, but ours is the most comprehensive study that combines both polymorphism and divergence data for this chromosome. We observed an apparent absence of amino acid sequence differences that have been fixed by positive selection, in contrast to the ~60% fraction for genes on other chromosomes. This agrees with the proposal of Betancourt and Presgraves (2002) and Presgraves (2005) that recombination accelerates adaptive protein sequence evolution.

The main caveat is that we cannot exclude the possibility that the two sets of genes which we have studied differ in properties that affect the rate of protein sequence. We note, however, that the major determinant of the rate of protein sequence evolution is the level of gene expression, with low expression genes showing higher rates of nonsynonymous substitutions (Drummond and Wilke 2008). There is, however, no significant difference in expression levels between the dot chromosome genes and the rest of the genome in *D. virilis* (Betancourt et al. 2009), so that this factor can be ruled out. Indeed, Haddrill et al. (2008) found significantly *higher* expression levels for genes on the *D. melanogaster* dot chromosome compared with the rest of the genome, and they suggested that this might reflect an adaptation to compensate for the lower functionality of protein sequences on this chromosome. It is interesting to note that a protein Painting of fourth (POF) has been characterized that binds specifically to the dot chromosome in *D. melanogaster* and appears to increase the expression of genes on this chromosome (Larsson et al. 2004; Johannson et al. 2007).

The only way of definitively dealing with this difficulty is to exploit systems in which the same genes can be compared in different recombinational environments. This is possible for homologous genes located on the two different sex chromosomes, when the Y or W chromosome does not recombine in the heterogametic sex. In the case of the *D. miranda* neo-Y and neo-X chromosomes mentioned above, there is evidence for accelerated protein sequence evolution associated with relaxed purifying selection on the nonrecombining neo-Y chromosome (Bartolomé and Charlesworth 2006; Bachtrog et al. 2008). The reduced levels of gene expression on this chromosome, and losses of gene function due to major mutational lesions, do not account for this effect (Bachtrog 2006). Similarly, accelerated protein sequence evolution and reduced variability have been observed on the W chromosome of birds (Berlin and Ellegren 2006) and the recently evolved Y chromosome of the white campion *Silene latifolia* (Marais et al. 2008). In all of these cases, the same genes are being compared between the two recombinational environments.

This strongly suggests that selective forces in a low recombination environment do indeed reduce N_e, leading to an impaired effectiveness of selection. It is difficult to be sure which of the factors listed in Table 1 is likely to be the most important cause. However, our study of the *D. americana* dot chromosome shows that it is apparently impossible to account for its reduced N_e by a recent selective sweep—we observe too many variants at intermediate frequencies to be consistent with such an event. As mentioned earlier, the levels of variability on both the dot chromosome and the *D. miranda* neo-Y chromosome are also inconsistent with the classical background selection model.

We propose that this paradox can be resolved by invoking HR interference among the sites subject to reversible mutation and purifying selection, when a large number of such sites are included in a low recombination genomic region. Our simulation results show that this produces weakening of the effective strength of selection on nonsynonymous variants. The HR effects mean that these variants are more likely to drift to intermediate frequencies and can become fixed more easily than with normal levels of recombination. This reduction in the effectiveness of selection means that the mutations in question have reduced effects on variability at linked neutral sites. As we discussed above, it seems likely that an asymptotic state is reached as the extent of a nonrecombining region increases, whereby adding more selected sites into a low recombination region has little or no effect on levels of variability. This may well account for the observation that the *D. miranda* neo-Y has a silent-site diversity value that is 1% of the neo-X value, despite the very large number of functional genes that it carries (Bachtrog et al. 2008).

We can also ask whether this type of process may provide a selective advantage to recombination at the individual level (see Barton, **this volume**), i.e., will selection resist the invasion of a freely recombining population by a genetic factor or chromosome rearrangement that reduces recombination or favor the invasion of a low recombination population by a modifier that increases it? Keightley and Otto (2006) investigated a similar model, but assumed unidirectional mutation from wild-type to deleterious alleles. This type of system cannot reach an equilibrium, and so is more similar to Muller's ratchet (Table 1) than ours (Gordo and Campos 2008).

We have therefore explored the possibility that HR interference among relatively strongly selected mutations may cause individual selection on modifiers of recombination, using simulations similar to those of Keightley and Otto (2006), but with populations at statistical equilibrium under reversible mutation, selection, and drift, similar to those described above. We introduced a single copy of a modifier allele that either completely suppresses recombination (if the initial population has a normal level of recombination) or increases it (if the population has zero recombination). To save computer time, a fixed selection coefficient was assigned to each site. The results are shown in **Table 3**. The advantage of increased recombination and disadvantage of decreased recombination tend to level off as the number of selected

Table 3. Fates of modifiers of recombination with Hill–Robertson effects

$2N$	$2Nu$	$2Ns$	L	P_{fix}/P_{neu}	2 × s.e.
A. Initial population with map length 90 cM ($2Nr$ = 450 for the whole chromosome); a modifier that completely suppresses recombination is introduced at equilibrium.					
500	0.01	10	10,000	0.75	0.27
500	0.01	10	20,000	0.80	0.28
500	0.01	10	50,000	0.28	0.17
500	0.01	10	100,000	0.08	0.09
500	0.01	10	200,000	0.15	0.12
B. Initial population with no recombination; a modifier that increases map length to 90 cM is introduced at equilibrium. (The modifier is located at one end of the chromosome.)					
500	0.01	10	1,000	1.25	0.50
500	0.01	10	5,000	1.50	0.77
500	0.01	10	10,000	3.00	1.09
500	0.01	10	20,000	6.00	2.43
500	0.01	10	40,000	5.25	2.28
500	0.01	10	50,000	6.56	1.19
500	0.01	10	100,000	9.21	2.04
500	0.01	10	200,000	10.27	2.58
500	0.01	10	300,000	13.09	2.51
500	0.01	10	400,000	11.20	2.09

10,000 to 20,000 simulations were run for each parameter set.

sites increases, similar to the effects on neutral diversity that we have described. The results show that HR effects among nonsynonymous mutations can have a significant influence on individual-level selection for recombination.

This raises the question of why some regions of the genome have low frequencies of recombination if selection generally favors increased recombination. The answer must lie in a selective advantage to recombination suppression that is sufficient to overcome selection in favor of recombination. In the case of Y chromosomes, a plausible scenario is that suppression of crossing-over between incipient X and Y chromosomes is advantageous because it prevents alleles that are favored in one sex, but deleterious in the other, recombining into the "wrong" sex (Charlesworth et al. 2005). It is less clear why certain regions of the genome, such as centromeres and telomeres, are associated with suppression of crossing-over (Sherman and Stack 1995; Gerton et al. 2000; Ashburner et al. 2005). (The lack of crossing-over on the dot chromosome probably reflects the fact that its small size means that its euchromatin is adjacent to both the telomere and the centromere.)

Centromeres in most species are compound structures, containing repetitive sequences around which the structure that binds the spindle fiber to microtubules forms (Charlesworth et al. 1986; Ashburner et al. 2005). Telomeres are also made up of repetitive units, of a different type from the centromere (Chan and Blackburn 2004). Unequal crossing-over between the repeat units would produce aberrant numbers of repeats, which could lead to aberrant chromosome segregation in mitosis and meiosis, resulting in aneuploid cells and reduced fitness (Charlesworth et al. 1986). This could result in a selective advantage to reduced crossing-over near centromeres and telomeres (Charlesworth et al. 1986). Additionally, exchanges near centromeres (even without unequal crossing over) may directly interfere with centromere disjunction in meiosis, again leading to aneuploidy. There is evidence for this from the smut fungus *Microbotryum violaceum* (Cattrall et al. 1978), yeast, *Drosophila*, and humans (Rockmill et al. 2006). The suppression of crossing-over in these genomic regions is thus likely to have an adaptive basis.

ACKNOWLEDGMENTS

A.J.B. was supported by a research grant from the UK Biotechnology and Biological Sciences Research Council, B.C. by the Royal Society, I.G. by the Fundação para a Ciência e Tecnologia, Portugal, and V.B.K. by the School of Biological Sciences, University of Edinburgh.

REFERENCES

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

Andolfatto P. 2007. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Res* **17:** 1755–1762.

Ashburner M, Golic KG, Hawley RS. 2005. Drosophila: *A laboratory handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bachtrog D. 2004. Evidence that positive selection drives Y-chromosome degeneration in *Drosophila miranda*. *Nat Genet* **36:** 518–522.

Bachtrog D. 2006. Expression profile of a degenerating neo-Y chromosome in *Drosophila*. *Curr Biol* **16:** 1694–1699.

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

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

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

Bell G. 1982. *The masterpiece of nature*. Croom-Helm, London.

Berlin S, Ellegren H. 2006. Fast accumulation of nonsynonymous mutations on the female-specific W chromosome in birds. *J Mol Evol* **62:** 66–72.

Berry AJ, Ajioka JW, Kreitman M. 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129:** 1111–1117.

Betancourt AJ, Presgraves DC. 2002. Linkage limits the power of natural selection. *Proc Natl Acad Sci* **99:** 13616–13620.

Betancourt AJ, Welch JJ, Charlesworth B. 2009. Reduced effectiveness of selection caused by lack of recombination. *Curr Biol* **19:** 655–660.

Bierne N, Eyre-Walker A. 2004. The genomic rate of adaptive amino-acid substitutions in *Drosophila*. *Mol Biol Evol* **21:** 1350–1360.

Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphism. *Genetics* **140:** 783–796.

Burt A, Bell G. 1987. Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* **326:** 803–805.

Cattrall ME, Baird ML, Garber ED. 1978. Genetics of *Ustilago violacea*. III. Crossing over and nondisjunction. *Bot Gaz* **13:** 266–270.

Chan SR, Blackburn EH. 2004. Telomeres and telomerase. *Philos Trans R Soc Lond B Biol Sci* **359:** 109–121.

Charlesworth B. 1996. Background selection and patterns of

genetic diversity in *Drosophila melanogaster*. *Genet Res* **68:** 131–150.
Charlesworth D. 2003. Effects of inbreeding on the genetic diversity of plant populations. *Philos Trans R Soc Lond B Biol Sci* **358:** 1051–1070.
Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* **10:** 195–205.
Charlesworth B, Charlesworth D. 2009. *Elements of evolutionary genetics*. Roberts, Greenwood Village, CO.
Charlesworth B, Langley CH, Stephan W. 1986. The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* **112:** 947–962.
Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134:** 1289–1303.
Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95:** 118–128.
Comeron JM, Williford A, Kliman RM. 2008. The Hill-Robertson effect: Evolutionary consequences of weak selection in finite populations. *Heredity* **100:** 19–31.
Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **13:** 341–352.
Ewens WJ. 2004. *Mathematical population genetics. 1. Theoretical introduction*. Springer, New York.
Fay J, Wyckhoff GJ, Wu C-I. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415:** 1024–1026.
Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* **78:** 737–756.
Fisher RA. 1930. *The genetical theory of natural selection*. Oxford University Press, Oxford.
Gay J, Myers S, McVean G. 2007. Estimating meiotic gene conversion rates from population genetic data. *Genetics* **177:** 881–894.
Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci* **97:** 11383–11390.
Gordo I, Campos RA. 2008. Sex and deleterious mutations. *Genetics* **179:** 621–626.
Gordo I, Navarro A, Charlesworth B. 2002. Muller's ratchet and the pattern of variation at a neutral locus. *Genetics* **161:** 835–848.
Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome Biol* **8:** R18.
Haddrill PR, Waldron FM, Charlesworth B. 2008. Elevated levels of expression associated with regions of the *Drosophila* genome that lack crossing over. *Biol Lett* **4:** 758–761.
Haldane JBS. 1937. The effect of variation on fitness. *Am Nat* **71:** 337–349.
Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res* **8:** 269–294.
Hudson RR, Kaplan NL. 1995. Deleterious background selection with recombination. *Genetics* **141:** 1605–1617.
Jensen MA, Charlesworth B, Kreitman M. 2002. Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster* and *D. simulans*. *Genetics* **160:** 493–507.
Johannson A-M, Stenberg P, Berhhardsson C, Larsson J. 2007. Painting of fourth and chromosome-wide regulation of the 4th chromosome in *Drosophila melanogaster*. *EMBO J* **26:** 2307–2316.
Kaiser VB, Charlesworth B. 2009. The effects of deleterious mutations on evolution in non-recombining genomes. *Trends Genet* **25:** 9–12.
Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* **177:** 2251–2261.
Keightley PD, Otto SP. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443:** 89–92.
Kimura M. 1962. On the probability of fixation of a mutant gene in a population. *Genetics* **47:** 713–719.
Kimura M. 1971. Theoretical foundations of population genetics at the molecular level. *Theor Popul Biol* **2:** 174–208.
Kliman RM, Hey J. 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol Biol Evol* **10:** 1239–1258.
Kliman RM, Hey J. 2003. Hill-Robertson interference in *Drosophila melanogaster:* Reply to Marais, Mouchiroud and Duret. *Genet Res* **81:** 89–90.
Langley CH, Lazzaro BP, Phillips W, Heikkinen E, Braverman JM. 2000. Linkage disequilibria and the site frequency spectra in the *su(s)* and *su(w^a)* regions of the *Drosophila melanogaster X* chromosome. *Genetics* **156:** 1837–1852.
Larsson J, Svensson MJ, Stenberg P, Mäkitalo M. 2004. Painting of fourth in genus *Drosophila* suggests autosome-specific gene regulation. *Proc Natl Acad Sci* **101:** 9278–9733.
Loewe L, Charlesworth B. 2006. Inferring the distribution of mutational effects on fitness in *Drosophila*. *Biol Lett* **2:** 426–430.
Loewe L, Charlesworth B. 2007. Background selection in single genes may explain patterns of codon bias. *Genetics* **175:** 1381–1393.
Loewe L, Charlesworth B, Bartolomé C, Nöel V. 2006. Estimating selection on nonsynonymous mutations. *Genetics* **172:** 1079–1092.
Marais G, Mouchiroud D, Duret L. 2003. Neutral effect of recombination on base composition in *Drosophila*. *Genet Res* **81:** 79–87.
Marais GAB, Nicolas M, Bergero R, Chambrier P, Kejnovsky E, Moneger F, Hobza R, Widmer A, Charlesworth D. 2008. Evi - dence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *Curr Biol* **18:** 545–549.
Maside X, Charlesworth B. 2007. Patterns of molecular variation and evolution in *Drosophila americana* and its relatives. *Genetics* **176:** 2293–2305.
Maynard Smith J. 1978. *The evolution of sex*. Cambridge Uni - versity Press, Cambridge.
Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* **23:** 23–35.
McAllister BF. 2002. Chromosomal and allelic variation in *Drosophila americana:* Selective maintenance of a chromosomal cline. *Genome* **45:** 13–21.
McVean GAT, Charlesworth B. 2000. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155:** 929–944.
Moran NA, McCutcheon JP, Nakabatchi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42:** 165–190.
Nordborg M, Charlesworth B, Charlesworth D. 1996. The effect of recombination on background selection. *Genet Res* **67:** 159–174.
Normark BB, Judson OP, Moran NA. 2003. Genomic signatures of ancient asexual lineages. *Biol J Linn Soc* **79:** 69–84.
Otto SP, Lenormand T. 2002. Resolving the paradox of sex and recombination. *Nat Rev Genet* **3:** 256–261.
Presgraves D. 2005. Recombination enhances protein adaptation in *Drosophila melanogaster*. *Curr Biol* **15:** 1651–1656.
Robertson A. 1961. Inbreeding in artificial selection programmes. *Genet Res* **2:** 189–194.
Rockmill B, Voelkel-Meiman K, Roeder GS. 2006. Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in *Saccharomyces cerevisiae*. *Genetics* **174:** 1745–1754.
Roze D, Lenormand T. 2005. Self-fertilization and the evolution of recombination. *Genetics* **170:** 840–857.
Santiago E, Caballero A. 1998. Effective size and polymorphism of linked neutral loci in populations under selection. *Genetics* **149:** 2105–2117.
Sawyer SA, Parsch J, Zhang Z, Hartl DL. 2007. Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. *Proc Natl Acad Sci* **104:** 6504–6510.

Sharp PJ, Hayman DL. 1988. An examination of the role of chiasma frequency in the genetic system of marsupials. *Heredity* **60:** 77–85.

Sheldahl LE, Weinreich DM, Rand DM. 2003. Recombination, dominance and selection on amino-acid polymorphisms in the *Drosophila* genome: Contrasting patterns on the X and fourth chromosomes. *Genetics* **165:** 1195–1208.

Sherman JD, Stack SM. 1995. Two-dimensional spreads of synaptonemal complexes from solanaceous plants. VI. High-resolution recombination map for tomato (*Lycopersicon esculentum*). *Genetics* **141:** 683–708.

Simonsen KL, Churchill GA, Aquadro CF. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141:** 413–429.

Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* **415:** 1022–1024.

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, Langley CH. 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.

Tachida H. 2000. DNA evolution under weak selection. *Gene* **261:** 3–9.

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

Turner JRG. 1967. Why does the genome not congeal? *Evolution* **21:** 645–656.

Wang W, Thornton K, Berry A, Long M. 2002. Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science* **295:** 134–137.

Wang W, Thornton K, Emerson JJ, Long M. 2004. Nucleotide variation along the *Drosophila simulans* fourth chromosome. *Genetics* **166:** 1783–1794.

Welch JJ. 2006. Estimating the genomewide rate of adaptive protein evolution in *Drosophila*. *Genetics* **173:** 821–827.

Williamson SM, Orive ME. 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.

Zhivotovsky LA, Feldman MW, Christiansen FB. 1994. Evolution of recombination among multiple selected loci: A generalized reduction principle. *Proc Natl Acad Sci* **91:** 1079–1093.