

## SHORT REVIEW

# The Hill–Robertson effect: evolutionary consequences of weak selection and linkage in finite populations

JM Comeron<sup>1,2</sup>, A Williford<sup>1</sup> and RM Kliman<sup>3</sup>

<sup>1</sup>Department of Biological Sciences, University of Iowa, IA, USA; <sup>2</sup>Carver Center for Comparative Genomics, University of Iowa, IA, USA and <sup>3</sup>Department of Biological Sciences, Cedar Crest College, Allentown, PA, USA

The ‘Hill–Robertson (HR) effect’ describes that linkage between sites under selection will reduce the overall effectiveness of selection in finite populations. Here we discuss the major concepts associated with the HR effect and present results of computer simulations focusing on the linkage effects generated by multiple sites under weak selection. Most models of linkage and selection forecast differences in effectiveness of selection between chromosomes or chromosomal regions involving a number of genes. The abundance and physical clustering of weakly selected mutations across genomes, however, justify the investigation of HR effects at a very local level and we pay particular

attention to linkage effects among selected sites of the same gene. Overall, HR effects caused by weakly selected mutations predict differences in effectiveness of selection between genes that differ in exon–intron structures and across genes. Under this scenario, introns might play an advantageous role reducing intragenic HR effects. Finally, we summarize observations that are consistent with local HR effects in *Drosophila*, discuss potential consequences on population genetic studies and suggest future lines of research.

*Heredity* (2008) **100**, 19–31; doi:10.1038/sj.hdy.6801059; published online 19 September 2007

**Keywords:** weak selection; recombination; gene structure; gene conversion bias; interference selection; codon bias

## Introduction

In the context of discussing the advantage of sexual organisms compared to asexuals, Fisher (1930) presented one of the first arguments on the evolutionary effect that selection at one locus impinges on other loci. Fisher argued (pp 121–122) that the prospect of a beneficial mutation ultimately prevailing, and so leading to evolutionary progress, depends upon the presence of other mutations in the population: the presence of either beneficial or deleterious mutations in the genetic background would slow down the fixation of a beneficial mutation, reducing the rate of adaptive evolution.

### The Hill–Robertson effect and the effective population size

In a fundamental study, Hill and Robertson (1966) used computer simulations to investigate multilocus interactions under realistic conditions, including selection and recombination in finite populations. In particular, Hill and Robertson studied a case with two loci with the possibility of random genetic drift, where they investigated the probability of fixation of a beneficial mutation in the presence of another segregating beneficial

mutation. The study confirmed Fisher’s predictions and showed that selection at one locus *interferes* with selection at a second beneficial mutation reducing its probability of fixation. That is, in natural populations (with finite size), selection at more than one site should cause an overall reduction in the effectiveness of selection. The magnitude of this effect depends on selection intensities and the initial frequencies of the alleles. More important, the degree of interference increases with genetic linkage between the loci under selection. The corollary of these observations, called ‘the Hill–Robertson (HR) effect’ by Felsenstein (1974), states that linkage between sites under selection will reduce the overall effectiveness of selection in natural populations (Hill and Robertson, 1966; Birky and Walsh, 1988).

Interestingly, Hill and Robertson (see also Robertson (1960, 1961)) rationalized the effects of interference among multiple loci under selection in terms of the increased (heritable) variance in fitness caused by selection in the genetic background, equating this to an increase in genetic drift relative to conditions with selection acting solely at one site. Thus the concept that interference between sites under selection will cause a reduction in the efficacy of selection can be investigated in terms of reduced effective population size ( $N_e$ ) that defines the increased genetic drift associated with a particular mutation or site. A more general description of the HR effect is then often presented in  $N_e$  terms: in the presence of selection  $N_e$  will vary with recombination, with reduced  $N_e$  in genomic regions with reduced

Correspondence: Dr JM Comeron, Department of Biological Sciences, University of Iowa, 212 Biology Building, Iowa City, IA 52242, USA.

E-mail: josep-comeron@uiowa.edu

Received 15 March 2007; revised 24 July 2007; accepted 6 August 2007; published online 19 September 2007

recombination rates relative to genomic regions with higher recombination (Hill and Robertson, 1966; Felsenstein, 1974; Birky and Walsh, 1988; Charlesworth *et al.*, 1993; Kliman and Hey, 1993b; Hilton *et al.*, 1994; Kondrashov, 1994; Caballero and Santiago, 1995; Barton, 1995a; Otto and Barton, 1997; Wang *et al.*, 1999).

The concept of  $N_e$  was introduced by Wright (1931, 1938) as an attempt to parameterize the dynamics of a complicated population process under an idealized evolutionary model (that is, the Wright–Fisher model). The study of the HR effect in terms of variation in  $N_e$  is convenient because it describes a reduction in effectiveness of selection (expressed by the product of  $N_e$  and the selection coefficient  $s$ ;  $\gamma = N_e s$ ) and also forecasts testable predictions about the level of intraspecific neutral variation  $\theta$  (described by the product of  $N_e$  and the mutation rate  $u$ ;  $\theta = 4 N_e u$  in diploids). As we discuss below, variation in  $N_e$  is indeed a useful simplification of the process caused by interference between sites under selection, although the HR effect generates population dynamics that cannot be solely described by differences in  $N_e$  (Comeron and Kreitman, 2002).

#### HR and the advantage of recombination

The two most prominent features of Hill and Robertson's study are (1) finite populations and (2) selection acting simultaneously at more than one locus. Clearly, these two features make the analysis of the HR effect particularly pertinent to the study of the evolution of recombination and to the limits of selection and adaptation in real biological systems (Haldane, 1957; Felsenstein, 1974; Felsenstein and Yokoyama, 1976; Barton, 1995a,b; Otto and Barton, 1997, 2001; Otto and Lenormand, 2002; Barton and Otto, 2005; Roze and Barton, 2006). Felsenstein (1974) examined the HR effect in the context of the evolutionary advantage of recombination, arguing that recombination breaks down negative linkage disequilibrium (LD) generated by selection and random genetic drift, thereby increasing the rate of adaptation (the probability of fixation of beneficial alleles). Otto and Barton (1997) pioneered analytical approximations that showed the advantage of modifiers for increased recombination in a two-locus case in finite populations. Later studies have presented more sophisticated analytical predictions and/or simulated more complex scenarios (Hey, 1998; Otto and Barton, 2001; Iles *et al.*, 2003; Barton and Otto, 2005; Keightley and Otto, 2006; Martin *et al.*, 2006; Roze and Barton, 2006). Altogether these studies have deepened our appreciation of the decisive role that finite population size and selection at multiple loci play explaining the advantage of sex and recombination.

#### HR and other models of linkage and selection

The HR effect is often used to describe the consequences of linkage reducing the effectiveness of selection regardless of the selective regime. The HR effect, as originally described (Hill and Robertson, 1966; Felsenstein, 1974), is specific in that it is caused by the segregation of mutations under selection. Strongly selected mutations (either beneficial or deleterious) are not expected to segregate within a population for long and hence the HR effect is frequently associated with weak/moderate

selection, when  $N_e$  plays a role on the fate of selected mutations.

It must be emphasized, however, that interference between segregating mutations is not the only selective paradigm that forecasts variation in  $N_e$  (and consequently in polymorphism and effectiveness of selection) in association with variation in recombination rates (that is, the consequences of the HR effect). Muller's ratchet (Muller, 1964) describes the stochastic accumulation of deleterious mutations as a consequence of drift and complete linkage in a model with deleterious mutations (no advantageous or back mutations, ever). The hitchhiking (HH) and pseudohitchhiking (pHH) models describe the spread of strongly favorable mutations (selective sweeps), possibly dragging deleterious mutations to fixation while eliminating extant variability, including other advantageous alleles (Maynard Smith and Haigh, 1974; Kaplan *et al.*, 1989; Gillespie, 2000; Kim and Stephan, 2003). The background selection (BGS) model focuses on negative selection on strongly deleterious mutations and the consequent removal of linked neutral and weakly selected mutations from populations (Charlesworth *et al.*, 1993; Charlesworth, 1994; Hudson and Kaplan, 1995). The differences among these models can be attributed to the particular selective scenarios rather than to their qualitative consequences. All of these models, to a different degree, predict that genetic linkage will cause a reduction in  $N_e$  and, consequently, reduced levels of neutral polymorphism, efficacy of selection and rate of adaptation; that is, in one way or another, all predict the general HR effect.

#### Genomic units associated with the HR effect

The positive relationship between recombination and  $N_e$  predicted by these models of selection and linkage is not trivial because it changes the genomic units classically associated with the advantage of sex and recombination. The general concept of  $N_e$  (Wright, 1931, 1938) is connected to diverse species-specific factors (unequal number of the two sexes, temporal variation in census population size, variance in mating success and so on), all of them with genome-wide consequences. Debates over the consequences of recombination (or lack thereof) in sexual and asexual organisms often invoke differences in  $N_e$  (all else being equal) and effectiveness of selection that apply to whole genomes. In species with sexual reproduction, different chromosomes usually vary in recombination rates per physical unit. Recombination rates also vary along chromosomes in the great majority of species, and this variation can occur over rather short physical distances. For instance, Cirulli *et al.* (2007) recently showed that crossover rate varies from 1.4 to 52 cM Mb<sup>-1</sup> within a 2 Mb region in *Drosophila pseudoobscura*. Therefore, HR effects predict not only a 'species'  $N_e$  but an  $N_e$  that will vary systematically within a genome, following changes in recombination. In all, the HR effect implies a change in the genomic units traditionally associated with  $N_e$  and effectiveness of selection, from whole genomes to chromosomes to chromosomal regions involving several genes. As reviewed below, the HR caused by weak selection further refines the genomic scale of HR effects, playing a previously unrecognized role at a much more local scale, the scale of individual genes, individual exons or even

shorter distances across exons (Comeron *et al.*, 1999; Comeron and Kreitman, 2002).

#### Direct and indirect evidence of HR in eukaryotes

Several lines of evidence support the concept that HR effects play a detectable role influencing the effectiveness of selection and  $N_e$ . Theoretical predictions forecast that mutations under weak selection will be highly sensitive to small differences in  $N_e$  and therefore they represent the ideal type of mutations to investigate the consequences of HR (Ohta and Kimura, 1971; Li, 1987). In *Drosophila*, as well as in many other species, weak selection can be investigated by studying synonymous mutations and synonymous codon usage (Moriyama and Hartl, 1993; Kliman and Hey, 1993b; Akashi, 1995, 1996; Moriyama and Powell, 1996, 1998; Akashi and Schaeffer, 1997; Powell and Moriyama, 1997; Comeron *et al.*, 1999; Kliman, 1999; Comeron and Kreitman, 2000, 2002; Begun, 2001; McVean and Vieira, 2001). Kliman and Hey (1993b) presented the first genomic analysis designed to investigate variation in effectiveness of selection in association with changes in rates of crossover, showing a reduced degree of adaptation to optimal synonymous codon usage in genes located in genomic regions with reduced rates of crossing-over in *D. melanogaster*.

More recent studies have confirmed the main observation of reduced selection at synonymous sites in regions with strongly reduced rates of recombination, even after correcting for possible nonselective influences on nucleotide composition (Kliman and Hey, 1993a; Comeron *et al.*, 1999; Zurovcova and Eanes, 1999; Comeron and Kreitman, 2002; Hey and Kliman, 2002; Marais and Piganeau, 2002; Marais *et al.*, 2003; Haddrill *et al.*, 2007). It should be noted however that the relationship between codon bias and recombination rate in *Drosophila* could also be influenced by nonselective factors. For instance, biased resolution of heteroduplex DNA that arises during crossover may have evolutionary effects that mimic weak selection (Nagylaki, 1983). In some eukaryotes heteroduplex DNA appears to be biased toward resolution as a GC base pair (that is, GC-biased gene conversion) and one could propose that the same bias might occur in *Drosophila*, explaining a positive relationship between recombination rate and use of preferred (G- and C-ending) codons (Marais *et al.*, 2001, 2003; Marais and Piganeau, 2002; Kliman and Hey, 2003a). Nevertheless, a GC-biased gene conversion model (Nagylaki, 1983) predicts that noncoding regions would show (1) a continuous increase in G + C content with recombination and (2) a reduction in rates of evolution in genomic regions with high recombination and G + C content. In *D. melanogaster*, the correlation between recombination rate and either codon bias or noncoding G + C content is not observed in regions with very high recombination rates (Kliman and Hey, 2003) and substitution rates at noncoding regions are not reduced in regions with high recombination (Singh *et al.*, 2005). Altogether, these observations are more consistent with a relaxation of long-range HR effects in regions of high recombination coupled with a recent increase in mutational biases toward AT (Kern and Begun, 2005; Akashi *et al.*, 2006).

Other studies support the detectable role of HR effects across genomes, with variable efficacy of selection in regions with different rates of recombination. Analyses

of rates of protein evolution between *Drosophila* species show that genes located in genomic regions with strongly reduced recombination have an excess of fixed deleterious mutations and a deficit of fixed advantageous mutations compared to highly recombining genomic regions (Hilton *et al.*, 1994; Takano, 1998; Comeron and Kreitman, 2000; Betancourt and Presgraves, 2002; Zhang and Parsch, 2005; Haddrill *et al.*, 2007). Also, the study of recent nonrecombining chromosomes (for example, neo-Y chromosome in *Drosophila miranda*) reveals an equivalent pattern with reduced effectiveness of selection (Bachtrog and Charlesworth, 2002; Bachtrog, 2003). Finally, artificial selection experiments have revealed that recombination favors the response to selection, increasing the rate of adaptation while populations with nonrecombining chromosomes accumulate deleterious mutations (Carson, 1958; Felsenstein, 1965; McPhee and Robertson, 1970; Rice, 1994; Barton and Charlesworth, 1998; Lynch, 1998; Fridolfsson and Ellegren, 2000; Rice and Chippindale, 2001).

The approximation that HR causes a reduction of  $N_e$  is further validated by studies of the level of neutral polymorphism ( $\theta$ ) across genomes. Genomic regions with reduced rates of crossing-over, and likely overall recombination, show reduced neutral  $\theta$  in many different eukaryotic genomes (Aguadé *et al.*, 1989; Stephan and Langley, 1989, 1998; Berry *et al.*, 1991; Begun and Aquadro, 1992; Martin-Campos *et al.*, 1992; Langley *et al.*, 1993; Aguadé and Langley, 1994; Stephan, 1994; Aquadro *et al.*, 1994; Moriyama and Powell, 1996; Nachman, 1997; Dvorak *et al.*, 1998; Kraft *et al.*, 1998; Nachman *et al.*, 1998; Zurovcova and Eanes, 1999; Bachtrog and Charlesworth, 2000; Przeworski *et al.*, 2000; Andolfatto and Przeworski, 2001; Baudry *et al.*, 2001; Jensen *et al.*, 2002; Wang *et al.*, 2002; Cutter and Payseur, 2003). Mitochondrial genomes, which lack recombination in many species, also show reduced level of polymorphisms (Weinreich and Rand, 2000). In *Drosophila*, as in most organisms investigated, such a reduction in levels of polymorphism cannot be accounted for by reduced mutation rates since there is no decline in rates of neutral divergence in most species (this is however not the case in primates (Hellmann *et al.*, 2003)). Altogether, these studies support the concept that genetic linkage reduces  $N_e$  in the presence of selection (Maynard Smith and Haigh, 1974; Ohta and Kimura, 1975; Birky and Walsh, 1988; Kaplan *et al.*, 1989; Stephan *et al.*, 1992; Charlesworth *et al.*, 1993; Gillespie, 1994; Hudson and Kaplan, 1995; Nordborg *et al.*, 1996; Barton, 1998; Przeworski *et al.*, 2000).

#### The Hill–Robertson effect caused by many weakly selected mutations

This review focuses on the Hill–Robertson effect generated by weak/moderate selection. Weakly selected mutations are of particular interest for at least two main reasons. First, genomes are likely to contain a large number of weakly selected sites, with a significant fraction of segregating genetic variation under weak selection. Second, weakly selected sites are likely to be physically clustered in exons and regulatory regions. This physical clustering is expected to amplify the HR effect because it decreases the opportunity for recombina-

nation events. The abundance and clustering of weakly selected mutations warrants the investigation of the broader implications of interference among these mutations, including a possible explanation for observed patterns of adaptation, levels of extant variation at genes and flanking neutral regions, and specific features of gene and genome architecture. That is, in this review we will use the term HR to specifically address the consequences of interference caused by linkage between multiple weakly selected sites. Although not discussed here we would also like to indicate that in finite populations selection at one locus might (under certain circumstances) influence the effectiveness of selection at genetically unlinked loci (Robertson, 1960, 1961; Cohan, 1984), a possibility that has not been fully explored.

Hill and Robertson (1966) indicated that no algebraic solution to the study of interference between sites under selection in finite populations was available at that time, and that most of the information had to come from Monte Carlo simulations of the evolutionary process. Substantial advances have been made with analytical frameworks now describing two-locus scenarios that include selection, drift and partial or total linkage (Otto and Barton, 1997; Barton and Otto, 2005; Roze and Barton, 2006). Unfortunately, there is no general analytical treatment describing the dynamics of interference among many alleles under selection and drift, with variable degrees of selection and linkage. Thus, forward computer simulations are still fundamental to the study of realistic situations. We will review simulation results of this evolutionary interference and discuss genomic and evolutionary patterns observed in *Drosophila* that support the concept that evolutionary patterns and rates, as well as some genomic and gene features might be shaped by the collective effects of many weakly selected mutations.

#### Multilocus forward simulations

In general, forward computer simulations follow a Wright–Fisher model, with an ideal population (diploid or haploid) and constant populations size ( $N$ ) that undergoes repeated cycles of mutation, possibly recombination, random mating, selection and sampling (Felsenstein, 1974; Li, 1987; Hey, 1998; Comeron *et al.*, 1999; Gillespie, 2000; McVean and Charlesworth, 2000; Tachida, 2000; Weinreich and Rand, 2000; Piganeau *et al.*, 2001; Comeron and Kreitman, 2002). Each new generation is obtained by choosing  $2N$  chromosomes with probability proportional to their relative fitness and randomly pairing these  $2N$  chromosomes. The mutational process allows for two (or more) allelic states at a site that are beneficial (preferred) or detrimental (unpreferred).

As noted by Hill and Robertson (1966), and as predicted by diffusion theory (Ewens, 1979; Li, 1987), the evolutionary process of mutation, selection and recombination can be ‘completely described’ by the product of  $N_e$  and individual rates or coefficients: mutation rate ( $u$ ), selection coefficient ( $s$ ) and recombination rate ( $c$ ). Hence, the evolutionarily pertinent parameters to study the effects of mutation, selection and linkage can be described as  $\theta$  ( $\theta = 4 N_e u$  in diploids),  $\gamma$  ( $\gamma = N_e s$ ) and  $\rho$  ( $\rho = N_e c$ ), respectively. At a practical level, the simulation of a relatively small population can give us quantitatively valuable information on much

larger, natural populations, as long as  $\theta$ ,  $\gamma$  and  $\rho$  are kept constant and  $\theta$  is small. Further, simulation times increase exponentially with  $N$  as both the time per generation, the number of generations required to obtain equilibria ( $> 10N/\theta$ ; (Tachida, 2000)) and the number of generations between independent measures ( $> N$  generations apart) increase with  $N$ . In practice, the smallest simulated population size is somewhat restricted by the sample size ( $n$ ) used to estimate intraspecific variables (estimates of polymorphism, frequency of variants, LD and so on). This is because several of these estimates are somewhat biased when  $n$  represents a large fraction of the population, usually  $> 5\%$  of the simulated number of chromosomes (Tachida, 2000; Comeron and Kreitman, 2002).

The selective model applied in these simulations can vary extensively. For instance, Li (1987) investigated haploid populations with either additive or multiplicative effects over sites. Multiplicative and additive effects (over sites) predict fitness with  $i$  preferred sites ( $F_i$ ) of  $(1+s)^i$  and  $1+is$ , respectively and the fitness effect for each additional preferred variant  $F_{i+1}/F_i$  is  $(1+s)$  for the multiplicative and  $(1+s)/(1+is)$  for the additive models. Therefore, multiplicative effects over sites is a required selective regime when trying to investigate the possible HR effects of an increasing number of sites under selection (Piganeau *et al.*, 2001); the use of additive effects would decrease the degree of selection per site with increasing number of sites beyond the possible influence of HR effects (decreasing the effectiveness of selection). Most simulation studies of the HR effect have assumed diploid populations, semidominance (or genic selection) and multiplicative effects over sites.

#### Simulation results of the HR effect

The first multilocus simulation of weakly selected mutations with beneficial and deleterious variants at mutation-selection-drift equilibrium was presented by Li (1987). Li’s simulations showed that under complete linkage the proportion of sites with the beneficial or preferred variant at equilibrium decreased when the number of sites under selection increased. This result was consistent with the interpretation of the HR effect, according to which linked selected sites cause a reduction in selection intensity. The simulation of 300 linked sites under weak selection ( $\gamma = Ns = 2$ ) suggested a reduction in frequency of preferred sites equivalent to a reduction of  $\sim 35\%$  in  $N_e$  relative to theoretical predictions of the mutation-selection-drift (MSD) model for single sites (see Table 1).

Later, multilocus simulations of diploid populations assuming semidominance and multiplicative effects over sites showed that the reduction in  $N_e$  was observed, not only with linked sites, but under a wide, realistic range of recombination rates (Comeron *et al.*, 1999). As the number of sites under weak selection increased, the frequency of preferred sites decreased for any given recombination rate, including expected rates for species such as *Drosophila*. Based on these simulation results, Comeron *et al.* (1999) proposed that HR effects could have a detectable influence at a very local and/or gene-specific level, as opposed to chromosomal or regional effects as previously assumed. Indeed, HR effects acting at a gene-specific scale could explain unexpected

**Table 1** Abbreviations and brief descriptions of evolutionary models

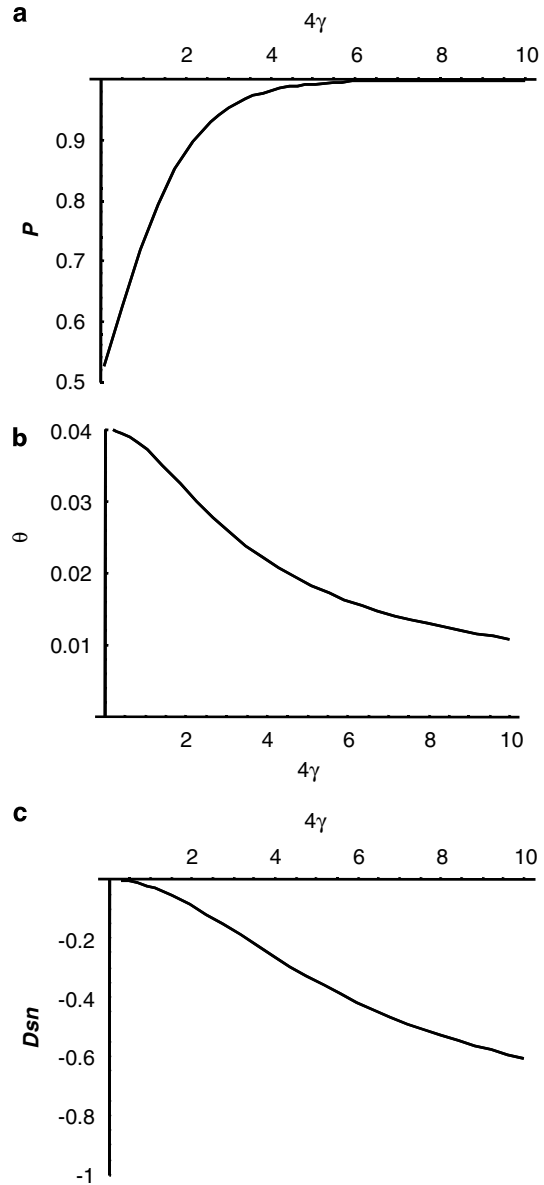
MSD	Mutation-selection-drift (MSD) equilibrium. Theoretical predictions under an infinite-site model, based on a single-site model in the absence of linkage
HR	The Hill–Robertson (HR) effect describes the interference between linked sites under selection in finite populations. The overall result is a reduction in the effectiveness of selection
HH	The Hitchhiking (HH) model describes the effects on linked sites of rare, strongly advantageous mutations when they increase in frequency within a population (i.e. selective sweep)
PHH	The pseudo-Hitchhiking (pHH) model describes the effects on linked sites of a steady stream of adaptive substitutions (a series of selective sweeps)
BGS	The background selection (BGS) model describes the effects on linked sites caused by selection against strongly deleterious mutations

patterns of divergence, polymorphism and codon bias in *Drosophila* genes. Rates of synonymous evolution ( $K_s$ ) increase with the length of coding sequence (CDS) and levels of synonymous polymorphism have been found to be lower in genes with reduced codon bias (Comeron and Aguade, 1996; Moriyama and Powell, 1996; Comeron *et al.*, 1999). Also, measures of codon usage bias decrease as the length of CDSs increase in *Drosophila* as well as in other eukaryotes (Moriyama and Powell, 1998; Comeron *et al.*, 1999; Duret and Mouchiroud, 1999). Altogether, these features were inconsistent with classic models of weak selection for single sites (MSD model), but directly predicted by a model with local HR effect, with variable  $N_e$  (and  $\theta$  and  $\gamma$ ) in association with gene-specific features such as CDS length.

Tachida (2000), McVean and Charlesworth (2000) and Comeron and Kreitman (2000, 2002) extended the analysis of HR effects using forward simulations to study multiple population and evolutionary parameters. These studies focused on levels of polymorphism and divergence, frequency distribution of mutations, average fitness, patterns of linkage equilibrium, heterogeneous effects across regions and possible effects of gene architecture. The major results of these multilocus studies are summarized below, with three variables always at play: (1) the number of sites under selection ( $L$ ), (2) selection intensity ( $\gamma$ ) and (3) recombination ( $\rho$ ). We pay particular attention to the consequences of HR effect that cannot be solely explained in terms of reduction in  $N_e$ . For illustrative purposes, we show expectations under MSD (in the absence of interference) equilibrium (Figure 1) and the simulation results for a case with complete linkage, variable  $L$  and  $4\gamma = 4$  (Figure 2).

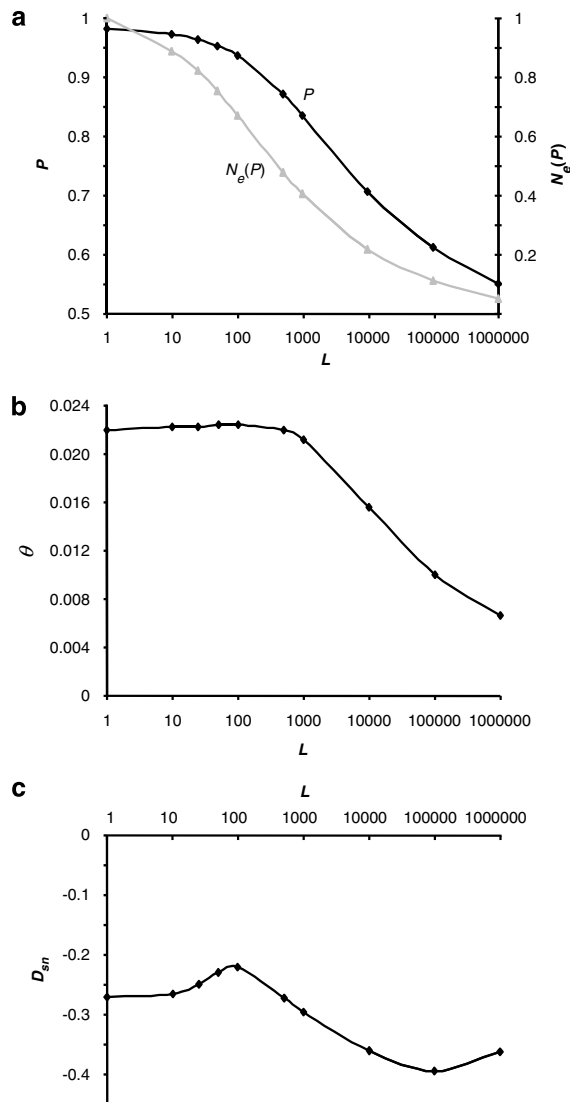
#### HR reduces the frequency of sites with the beneficial/preferred variant

Under MSD, the frequency of sites with the beneficial or preferred variant ( $P$ ) increases with selection relative to the frequency expected purely by mutational tendencies (Figure 1a) (Li, 1987; Bulmer, 1991; McVean and Charlesworth, 1999). Linked sites under selection easily cause detectable HR effects and make  $P$  to be smaller than that predicted by MSD. Increasing  $L$  increases HR and



**Figure 1** Expectations under mutation-selection-drift (MSD) equilibrium for single sites (in the absence of interference) in diploid organisms. The MSD model investigated assumes beneficial and deleterious variants, scaled selection coefficient  $\gamma$  ( $\gamma = N_e s$ ) and semidominance. (a) Frequency of sites with the beneficial or preferred variant ( $P$ ). (b) Polymorphism level ( $\theta$ ) as measured by the number of segregating sites in a sample size ( $n$ ) of 10 chromosomes (with neutral expectations of  $\theta = 0.04$ ). (c) Measures of the deviation in allele frequency spectrum from neutral expectations based on Tajima’s  $D$  (Tajima, 1989) values. Because  $D$  is influenced by the number of segregating sites ( $S$ ) and HR also varies  $S$ , we have followed Schaeffer (2002) and used a modification of Tajima’s  $D$  statistic ( $D_{sn}$ ) that is independent of  $S$  and  $n$  to better investigate the frequency of mutations in a sample.  $D_{sn}$  is calculated as the ratio of Tajima’s  $D$  to its theoretical minimum value (Schaeffer, 2002), where  $D_{sn} = (ak - S)/(ak_{\min} - S)$ ,  $k_{\min} = 2S/n$  and  $a = \sum_{i=1}^{n-1} \frac{1}{i}$ .

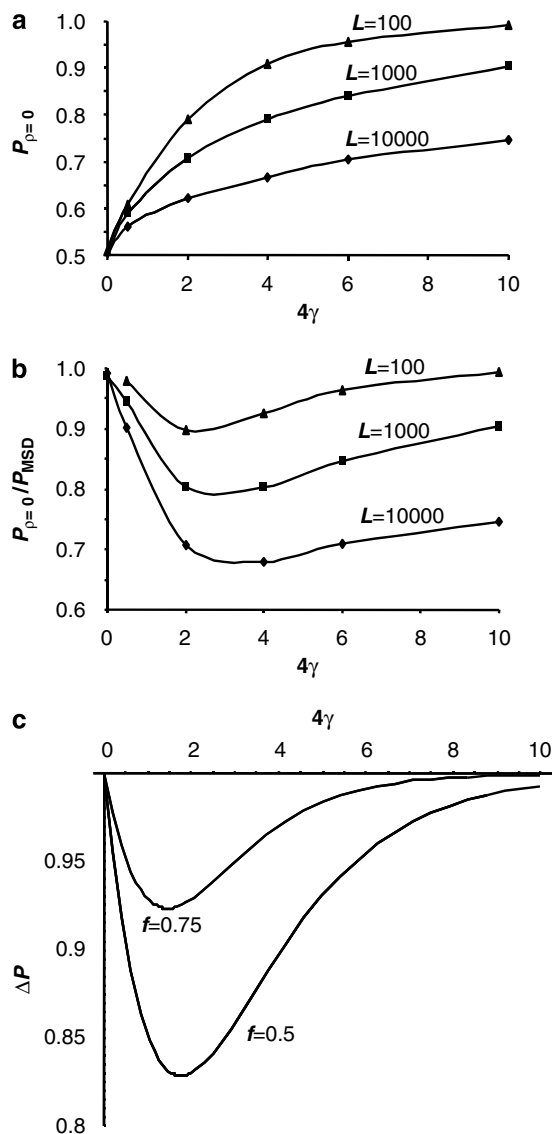
reduces  $P$ , particularly when  $L > 50$ . (Figure 2a) (Li, 1987; Comeron *et al.*, 1999; McVean and Charlesworth, 2000; Comeron and Kreitman, 2002). To better quantify the consequences of HR it is informative to study the relative



**Figure 2** Simulation results of the HR effect for a case with complete linkage ( $\rho=0$ ), scaled selection coefficients  $\gamma=1$  and variable number of sites under selection ( $L$ ) (see text and Comeron and Kreitman (2002) for simulation details). (a) Frequency of sites with the preferred variant ( $P$ ) and estimates of the relative  $N_e$  that would cause the observed reduction in  $P$  ( $N_e(P)$ ). (b) Polymorphism levels ( $\theta$ ). (c) Modified Tajima's  $D$  statistic ( $D_{sn}$ ; see Figure 1 for details).

reduction in  $N_e$  that would cause the observed reduction in  $P$  ( $N_e(P)$  in Figure 2). In this case, with  $4\gamma=4$  and complete linkage, simulations of 100 sites generate a  $P$  that is expected for  $N_e$  (and  $\gamma$ ) that is 67% of its original value; one million linked sites cause a very strong HR effect, with an  $N_e$  and effective  $\gamma$  only 5% its original value (see also McVean and Charlesworth, 2000).

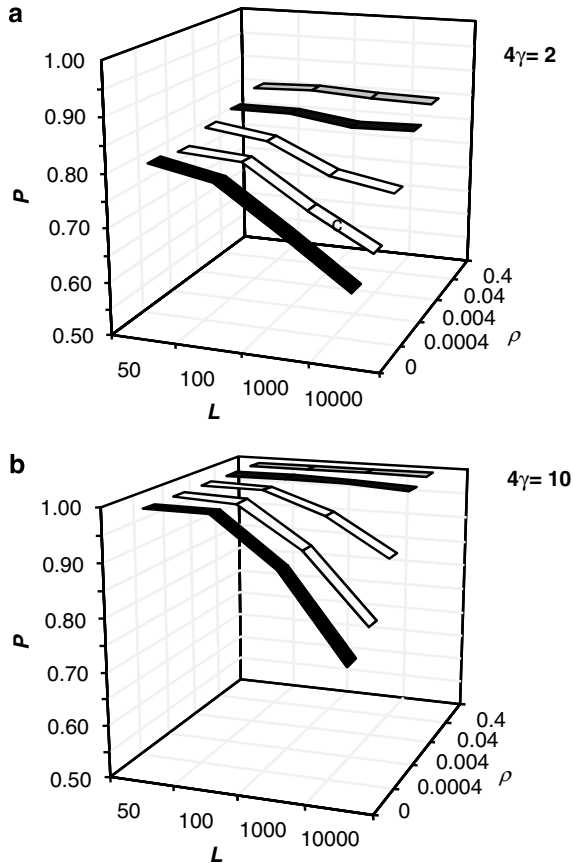
For any given  $\gamma$ , the longer the sequence under selection, the greater the reduction in  $P$  (Figure 3a). Interestingly, this reduction in  $P$  is not constant across  $\gamma$  and Figure 3b illustrates this point by showing  $P$  relative to its expected value under MSD; the largest reduction in  $P$  is observed at intermediate selection intensities (see also (McVean and Charlesworth, 2000)). This is because the expected increase in  $P$  with  $\gamma$  is not linear (Figure 1a), but is rapid for small selection intensities ( $4\gamma < 2$ ) and



**Figure 3** (a) Simulation results showing the influence on the frequency of sites with the preferred variant ( $P$ ) of variable number of sites under selection ( $L$ ) and selection coefficients ( $\gamma$ ) when linkage is complete ( $\rho=0$ ). (b) Simulation results showing  $P$  relative to its expected value under a mutation–selection–drift (MSD) models for single sites. (c) MSD predictions on the expected reduction in  $P$  ( $\Delta P$ ) when  $N_e$  is reduced by a factor  $f$  ( $f=0.75$  or  $f=0.5$ ).

very slow for larger selection intensities ( $4\gamma > 5$ ). Figure 3c depicts these theoretical expectations when  $N_e$  is reduced by a factor  $f$  ( $f=0.75$  or  $f=0.5$ ), with an expected maximum reduction in  $P$  ( $\Delta P$ ) when  $4\gamma$  is between 1 and 4. In other words, the study of HR effects based on changes in the frequency of the beneficial variant (in our case  $P$ ) will be most informative when selection intensities are intermediate (within the realm of weak selection;  $0.1 < \gamma < 10$ ).

Increasing recombination reduces HR, for any given  $L$  and/or  $\gamma$ , and makes the observed  $P$  closer to that predicted by MSD (Figure 4) (Comeron *et al.*, 1999; McVean and Charlesworth, 2000; Comeron and Kreitman, 2002). When the number of sites is small, recombination mostly restores the frequency of preferred sites expected under MSD. For sequences with many

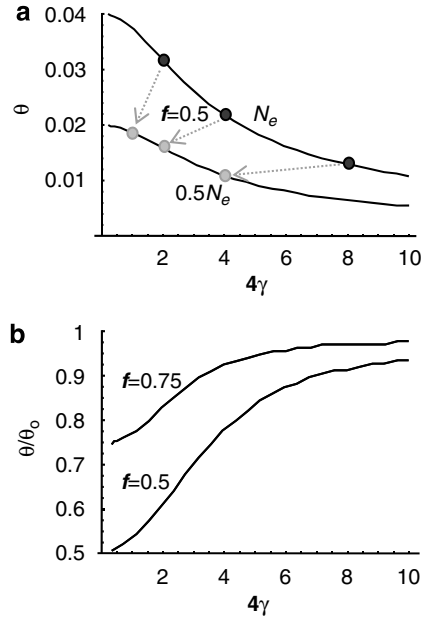


**Figure 4** Simulation results showing the influence of variable number of sites under selection ( $L$ ) and variable-scaled recombination rates ( $\rho = N_e c$ ) on the frequency of sites with the preferred variant ( $P$ ).

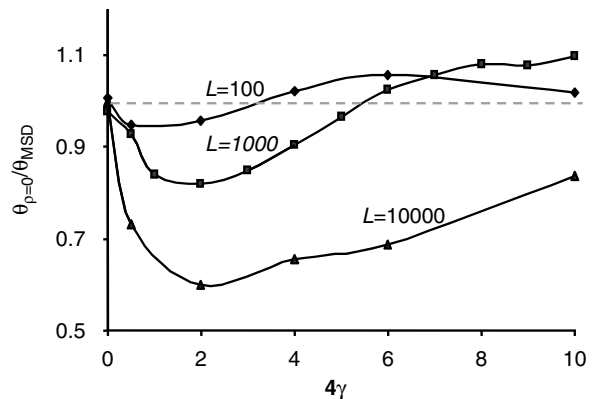
sites under selection, however, even realistically high levels of recombination for most species are not sufficient to completely remove the effects of interference. The magnitude of these residual effects will depend on selection intensity, with maximum consequences for intermediate ( $4\gamma = 2$ ) selection (McVean and Charlesworth, 2000; Cameron and Kreitman, 2002).

#### HR reduces polymorphism levels (but not always)

According to MSD predictions, polymorphism levels (as measured by estimates of  $\theta$ ) at sites under weak selection decrease relative to neutral levels ( $\theta_0$ ) when selection increases (Figure 1b). The expected influence of HR effects on  $\theta$  is however not direct because a reduction in  $N_e$  has a dual effect. On the one hand, a reduction in  $N_e$  results in a reduction in neutral polymorphism ( $\theta_0$ ) and, hence, it will reduce the maximum level under MSD. On the other hand, a reduction in  $\gamma$  will result in an increase in  $\theta$  relative to its maximum, neutral value ( $\theta/\theta_0$ ). In all, theory predicts that the cumulative effect of the reduction in  $N_e$  reducing both  $\theta_0$  and  $\gamma$  will always be the reduction in  $\theta$  (Figures 5a and b) (Cameron and Kreitman, 2002); only a reduction in  $\gamma$  that is not associated with a reduction in  $N_e$  can cause an increase in  $\theta$ . If HR would exert its effects solely through the reduction in  $N_e$ , the only possible expected outcome is a reduction in  $\theta$  and never an increase.



**Figure 5** Predicted effects of a reduction in  $N_e$  on levels of polymorphism ( $\theta$ ) under an MSD model for single sites and for different values of selection ( $\gamma$ ). (a) Overall  $\theta$  predicted when  $N_e$  is reduced by a factor ( $f$ ) of 0.5; note that  $\theta$  under neutrality varies from 0.04 to 0.02. (b) Predicted  $\theta$  relative to its neutral ( $\theta_0$ ) value ( $\theta/\theta_0$ ) when  $N_e$  is reduced by a factor  $f$  ( $f = 0.75$  or  $f = 0.5$ ).



**Figure 6** Simulation results showing the effects of complete linkage ( $\rho = 0$ ) on polymorphism ( $\theta$ ) levels relative to mutation-selection-drift (MSD) predictions for single sites. Results shown for different number of selected sites ( $L = 100, 1000$  and  $10000$ ) and different values of selection ( $\gamma$ ). Dashed line indicates no HR effects, with  $\theta_{\rho=0}/\theta_{\text{MSD}} = 1$ .

Figure 2b shows how increasing the number of sites alters the observed  $\theta$ , making it smaller than that predicted by MSD when  $4\gamma = 4$ . Interestingly, McVean and Charlesworth (2000) showed that this is not always the case. Figure 6 depicts the influence of different  $\gamma$  on the relative  $\theta$ , showing that  $\theta$  does not always decrease due to HR. When  $L = 1000$  there is a reduction in  $\theta$  for  $4\gamma < 5$  but  $\theta$  is higher than expected by MSD predictions for  $4\gamma > 5$ . Note also that the precise value of  $4\gamma$  at which  $\theta$  crosses MSD expectations depends on the number of sites under selection. These results on  $\theta$  are the first that suggest that HR effects due to weak selection cannot be solely explained by a reduction in  $N_e$ .

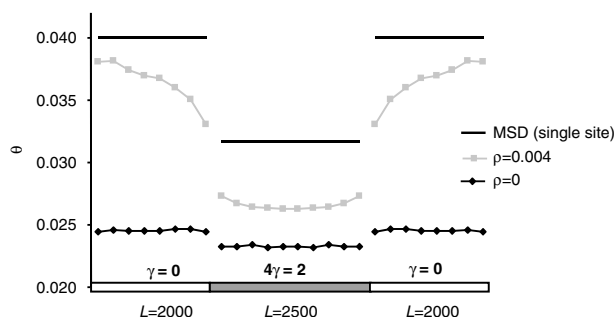
### HR increases the frequency of rare mutations

According to the MSD model, the frequency of mutations (that is, derived states at polymorphic sites) will decrease when selection intensity increases. This leads to negative Tajima's  $D$  (Tajima, 1989) values, a measure of the deviation in allele frequency spectrum from neutral expectations (Figure 1c). Note that we have used a modification of Tajima's  $D$  statistic ( $D_{sn}$ ) that is not influenced by the number of segregating sites or number of chromosomes (Schaeffer, 2002); see legend in Figure 1c for details). A modified  $D_{sn}$  independent of the number of segregating sites is necessary when comparing  $D$  values, because HR reduces the number of segregating sites (see above). The consequences of a reduction in  $N_e$  (and  $\gamma$ ) are expected to generate  $D_{sn}$  values closer to those expected under neutrality (less negative). Simulations show that when HR effects are weak (Figure 2c;  $L < 500$ )  $D_{sn}$  tends toward the neutral expectation, but becomes more negative than expected when HR effects are more noticeable ( $L > 1000$ ). Equivalent results are observed for different  $\gamma$  ( $0.5 < 4\gamma < 10$ ; data not shown) (McVean and Charlesworth, 2000; Tachida, 2000; Comeron and Kreitman, 2002). The reduction in allele frequencies due to HR effects are contrary to the proposal that HR effects can be investigated simply (and fully) as phenomena caused by a reduction in  $N_e$ .

### HR produces heterogeneous effects across regions under selection, with strongest effects in the central region

The detection of an influence on HR effects by just doubling the number of sites suggested the possibility that sites at the center of a region under selection might show stronger effects than sites at the edge of these same regions (which, in an ideal case, should be influenced by half of the neighboring sites under weak selection). Comeron and Kreitman (2002) found that under realistic conditions of recombination (at least for *Drosophila*), the center of a region under uniform selection showed reduced  $P$ , reduced  $\theta$  and ratios of polymorphism to divergence ( $rpd$ ) closer to neutral expectations relative to the edge of these same regions. These results suggest that HR effects might have very local effects, not only explaining differences in  $N_e$  and effectiveness of selection among genes but also showing variation across exons.

Figure 7 shows the 'center effect' on polymorphism. As expected, complete linkage does not produce any



**Figure 7** Simulation results showing the heterogeneous distribution of polymorphism ( $\theta$ ) levels across a region of 2500 sites under uniform selection ( $4\gamma=2$ ) and adjacent neutral sites. Results shown for complete linkage ( $\rho=0$ ) and partial linkage ( $\rho=0.004$ ), and compared to to mutation-selection-drift (MSD) predictions for single sites (in the absence of interference).

heterogeneity across the region under selection. For weak selection under MSD, the difference in  $\theta$  between central and lateral regions increases with intermediate recombination rates and mostly disappears with high recombination rates ( $\rho > 0.04$ ) when the physical range of interference is expected to be minimal. For intermediate ranges of recombination ( $\rho \sim 0.004$ – $0.04$ ) and weak selection, there is a  $\sim 4\%$  reduction in  $\theta$  in the central region compared to the lateral regions. In terms of expected variation in efficacy of selection and  $N_e$  across regions, the differences are fairly small under the selective scenario investigated but measurable.

### Neutral regions embedded in regions under selection ameliorate HR effects

As a complementary analysis to the results described above, simulations show that the insertion of a neutral sequence into the selected sequence noticeably increases selection intensity, leading to higher  $P$  in the regions under selection (Comeron and Kreitman, 2002). This effect is entirely due to physical separation of selected sites that allows for increase in recombination, relieving the effects of interference. As expected, the magnitude of these effects depends on both the selection coefficients and recombination rates. With high recombination, short intervening neutral sequences suffice to cause evolutionary independence between the two regions under selection. As recombination rates decrease, more neutral sites are needed to eliminate interference between regions.

### HR increases negative linkage disequilibrium

The results of two-locus simulations by Hill and Robertson (1966) indicated that as linkage between the selected loci becomes tighter, the LD becomes more negative (more frequent associations between preferred and unpreferred alleles). McVean and Charlesworth (2000) showed that an equivalent outcome is observed in multilocus simulations, with the extent of LD becoming more negative as distance between selected sites decreases. Increasing any of the three parameters that cause HR (number of sites, selection or genetic linkage) causes an increase in negative LD (Comeron and Kreitman, 2002).

### HR affects adjacent neutral sites

The influence of selection on linked neutral sites has usually been associated with strong selection on either favorable (the HH and pHH models) or deleterious mutations (the BGS model). The study of neutral sites adjacent to regions under weak selection shows that HR can also reduce levels of neutral polymorphism when these neutral sites are physically close to the region under selection (Comeron and Kreitman, 2002). Figure 7 shows the influence of HR on neutral polymorphism ( $\theta_0$ ) for different recombination rates ( $\rho$ ). The strongest reduction in neutral polymorphism is observed under the same conditions that maximize the effect of HR at selected sites: large number of selected sites and low recombination. Increasing recombination leads to a gradual increase in levels of neutral polymorphism, with faster recovery to neutral levels at sites that are farthest away from the region under selection.

Notably, intermediate recombination rates ( $4\gamma=2$ ;  $\rho=0.004$ ) cause a substantial effect on  $\theta_0$  at neutral sequences immediately adjacent to the region under selection, with a reduction of  $\sim 17\%$  in 250 bp regions adjacent to a region with  $L=2500$  and a reduction of  $\sim 8\%$  at a distance of 1000 sites from the region under selection. The frequency of neutral segregating mutations is also influenced, showing a negative  $D_{sn}$  close to the regions under selection (data not shown). Parallel to a reduction in polymorphism and an excess of rare mutations at neutral sites, HR also causes an excess of LD in these adjacent neutral sequences. As noted by Birky and Walsh (1988), however, substitution rates at neutral sites do not change under HR conditions. Therefore, neutral sites influenced by local HR will show lower  $rp_d$  than neutral sites not influenced by HR (Comeron and Kreitman, 2002).

In all, these simulation results demonstrate that models invoking strong selection (that is, the HH, pHH or BGS models) are not the only ones that can generate a reduction in neutral polymorphism in regions with reduced recombination. Further, HR may (under certain circumstances) cause a reduction in  $rp_d$  and an excess of rare variants, two features usually associated with recent events of positive selection and selective sweeps. That is, results of simulations indicate that the HR effect should be considered as a possible cause of departures from neutral expectations when population genetic analyses suggest positive selection, particularly in regions adjacent to exons and genes (see below). The quantitative influence of selection at coding regions on adjacent possibly neutral sites will require further simulation and theoretical analyses.

HR can only partially be explained in terms of increased drift and reduced  $N_e$

The results of simulations described above show that many population features caused by HR are congruent with expectations associated with a reduced  $N_e$ , the approach originally proposed by Hill and Robertson (1966) that has been widely accepted since. Indeed, reduced levels of polymorphism and effectiveness of selection with increasing interference (increasing linkage and/or the number of sites) are two features qualitatively in agreement with reduced  $N_e$ .

Nevertheless, an excess of rare variants and increased LD not only at sites under selection, but also at linked neutral sites, evidence dynamics that cannot be fully explained by reduced  $N_e$  and, in fact, are opposite to predictions based solely on reduced  $N_e$ . Also, the analysis of  $P$  and  $\theta$  under HR provide incongruent estimates of  $N_e$  (McVean and Charlesworth, 2000). For instance, when  $L=10\,000$  with complete linkage (Figure 2), estimates of  $N_e$  represent 42 and 28% of the original  $N_e$  when  $P$  and  $\theta$  are used, respectively. Another approach to investigate the underlying causes associated with HR is the analysis of the time to fixation of advantageous mutations and simulation results show that HR causes weak mutations to take longer to be fixed than expected purely based on estimates of  $N_e$  (using either  $P$  or  $\theta$ ; Williford and Comeron, unpublished data). In conclusion, the simulation results show that HR cannot be solely explained and investigated as a drift-enhancing mechanism equivalent to a simple reduction

in  $N_e$ . The population dynamics under HR interference is compatible with a dynamic ‘allelic traffic’, with many partial selective sweeps that infrequently reach fixation unbroken due to appearance of slightly deleterious mutations in the same chromosome (Comeron and Kreitman, 2002).

#### Simulation results of the HR effect caused by many deleterious mutations

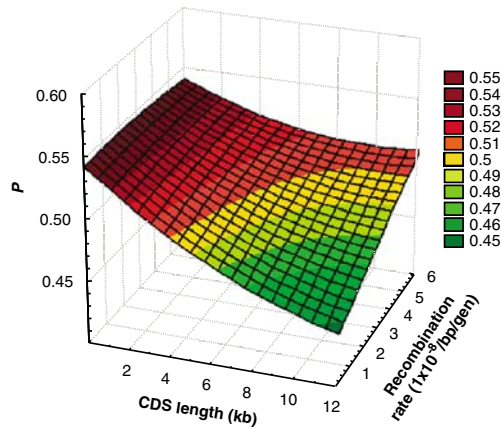
Recently, Loewe and Charlesworth (2007) have investigated the possibility that many moderately deleterious mutations could also generate HR effects at a very local genomic scale. Loewe and Charlesworth use simulation approximations to expand the standard BGS model to incorporate moderate/weak negative selection coefficients (with  $\gamma>1$ ) and, interestingly, include the effects of both crossover and gene conversion. Their results show that most of the intragenic patterns generated by HR effects among very weakly selected mutations can also be expected in a model with only weakly deleterious mutations under realistic conditions of mutation, selection and recombination for *Drosophila*. In fact, heterogeneity of HR effects across regions under selection is predicted to be more extreme under this ‘moderate’ BGS scenario. Additional studies are needed to fully appreciate the influence of moderate selection under a BGS model relative to or in combination with other selective regimes that incorporate beneficial mutations.

#### Genomic and evolutionary analyses

The results of simulations summarized above provide a coherent picture of the effects that selection acting at many weakly selected sites can potentially impinge on patterns of variation and genome structure. But as with any other theoretical framework, the question is whether this model of selection is not only reasonable but realistic, congruent with observed patterns of variation and genome architecture in actual species. Below we summarize several observed correlations that are consistent with the effects produced by HR due to weak selection (see Figure 8).

As indicated, three different observations in *Drosophila* prompted the investigation of HR as a possible mechanism with gene-specific effects: (1) the level of synonymous polymorphism is positively correlated with codon bias while theoretical predictions under MSD forecast a negative relationship (Moriyama and Powell, 1996), (2) gene length is positively correlated with  $K_s$  (Comeron and Aguade, 1996; Comeron and Guthrie, 2005) and (3) gene length is negatively correlated with codon usage bias (Moriyama and Powell, 1998; Comeron *et al.*, 1999; Duret and Mouchiroud, 1999). These three patterns suggest a reduction in effectiveness of selection in long genes associated with a reduction in  $N_e$ , as expected by local HR effects. As indicated above, *a posteriori* simulations showed that HR caused by weak selection could generate such patterns using reasonable parameters for recombination, selection and drift.

The study of evolutionary patterns across genes is particularly appropriate to investigate HR effects because genomic or transcriptional influences can be assumed to be equivalent. Synonymous codon usage bias across long exons in *D. melanogaster* shows a



**Figure 8** Observed relationship between the length of coding sequences (CDS), recombination rate (rate of crossing-over) and the frequency of preferred codons ( $P$ ) in *Drosophila melanogaster* (release 4; September, 2006). The recombination rate for each gene was estimated as previously described (Comeron *et al.*, 1999). Only genes with a single transcript and showing no overlap with adjacent genes were used in the analysis (8967 genes). Three-dimensional surface plot (quadratic fit) obtained using Statistica 6.1 (StatSoft, Inc. (2003)).

reduction in codon bias in the central region that suggests a reduction of  $\sim 10\%$  in  $\gamma$  relative to the edges of these same exons (Comeron and Kreitman, 2002). Moreover, the reduction in the effectiveness of selection in the central region increases with the length of the coding region as predicted by HR effects. The observation that genes with introns show no ‘center effect’ across a CDS also supports this conclusion and rules out a possible inherent reduction in  $\gamma$  in the central regions of CDS. In fact, measures of density of sites under selection (as measured by proportion of coding region in a gene) indicates reduced efficacy of selection at the level of codon bias with increasing density of selected sites under weak selection (Comeron and Kreitman, 2002). Later analyses of codon usage bias across CDSs in *D. melanogaster* revealed similar patterns (Qin *et al.*, 2004). This study reports that the center effect increases with levels of expression, as expected if selection at synonymous sites increases with expression (Duret and Mouchiroud, 1999). Also the center effect lessens when genes with introns are included in the analysis. Interestingly, this detailed analysis shows that the center effect is accompanied by a very narrowly distributed reduction in selection at the very 5' and 3' regions, with an M-shaped distribution of codon bias along the whole CDS, with highest codon bias at  $\sim 50$  codons from both ends of the CDS. This observation resembles that observed in *Escherichia coli* and attributed to conflicting selection pressures (Bulmer, 1988; Eyre-Walker and Bulmer, 1993; Eyre-Walker, 1996), but no direct explanation has been presented so far for *Drosophila*.

A recent study of divergence and polymorphism in *Drosophila* genes that differ in the length of their exons but with similar levels of expression and recombination rates allowed the direct comparison of  $K_s$  and diverse estimates of  $\gamma$  based on patterns of polymorphism. The results showed a significant increase in  $K_s$  and reduced  $\gamma$  in long exons relative to short exons and an equivalent increase in  $K_s$  and reduced  $\gamma$  and the central region of

long exons (Comeron and Guthrie, 2005). Further, the observed ‘center effect’ in long exons accounts for the difference between genes with long and short exons. These results are congruent with the proposal that HR effects play a detectable role at intragenic level in *Drosophila* hence making CDS length, exon length and intron presence and length relevant parameters when studying the evolutionary trends associated with variation in the efficacy of selection (Comeron and Kreitman, 2002).

In another analysis of interference selection across genomes, Hey and Kliman (2002) investigated the possibility of HR effects between adjacent genes in *D. melanogaster*. Codon bias was expected to decrease with gene density under HR effects and such a relationship was observed for genes that are separated by less than 2000 bp; longer intergenic sequences generated the opposite trend. This result suggests that the evolutionary parameters for recombination and selection in *D. melanogaster* cause HR effects that are mostly restricted to single genes with intragenic effects, while influencing a short physical distance upstream and downstream of most genes.

The patterns previously described for *D. melanogaster* are however not universally observed. For instance, Qin *et al.* (2004) investigated intragenic patterns in *Saccharomyces cerevisiae* and prokaryotes, showing that the center effect is not detected in these species. These results should not be taken as evidence against the existence of HR effects because they are directly predicted by the model: with rare recombination or stronger selective events (a likely scenario in species with larger  $N_e$  than *D. melanogaster*), HR is not expected to vary intragenically but to influence larger physical distances, at the level of whole genes and/or gene density. Indeed, yeast genes in regions of increased density and those with longer CDSs show reduced codon bias (Kliman *et al.*, 2003), a result that is congruent with the idea that, in this species, HR has effects at the scale of whole genes or groups of genes with negligible variation at intragenic level. Alternatively the relationship between gene density and codon bias may not reflect HR, *per se*, but rather a distinct form of selection conflict: antagonistic pleiotropy. Kliman *et al.* (2003) observed that gene expression also decreased as gene density increased. While the effectiveness of selection on regulatory elements in noncoding regions shared by gene pairs might be less effective due to HR effects, it is also possible that pleiotropic effects of mutations in overlapping regulatory regions make it more difficult to fine-tune expression of closely spaced gene pairs. Further analyses are required to assess whether the observed correlation between codon bias and gene density may reflect HR effects or a more direct relationship between gene expression and gene density.

Finally, recent evolutionary studies of noncoding regions in *Drosophila* show a detectable fraction of mutations under purifying and positive selection, likely evidencing evolutionary trends of regulatory sequences (Andolfatto, 2005). It is therefore possible that selection at regulatory sequences could also generate local HR effects and future evolutionary studies will assess the magnitude and distribution of these effects on population parameters. These studies should also address the possible effects that selection at exons might impinge

on population parameters and estimates of selection at adjacent non-CDSs and vice versa.

## Conclusions

Here we reviewed the major concepts and simulation results associated with the HR effect, focusing on those caused by weak selection at many sites, and summarized empirical observations that are both predicted by this selective scenario and observed, at least, in *Drosophila*. At a genomic scale, HR caused by weak selection has the potential for explaining very local patterns of polymorphism, divergence and effectiveness of selection, with variation among genes that differ in CDS length and exon–intron structures as well as variation across genes and exons. This is predicted from simulations and observed in *D. melanogaster*. One of the genomic consequences of HR effects at this local scale is that introns might be viewed as modifiers of recombination that lower intragenic HR effects, hence providing a possible evolutionary explanation to the origin and maintenance of introns.

Although the empirical support for HR acting at a very local level across genes and genomes is notable, alternative selection regimes must also be investigated. For instance, the weak selection scenario investigated assumes beneficial and deleterious mutations with uniform selection coefficients, clearly an oversimplification of selective regimes across genes and genomes. Future simulations should incorporate more realistic spectra of selection coefficients, with deleterious and beneficial mutations and include both weak and strong selection. In all, simulation studies are highly informative describing expected trends of HR effects and provide an opportunity to design experimental analyses that will focus on quantifying these effects across genes and genomes and assess their selective causes and consequences.

## Acknowledgements

We thank Ana Llopart, Bruce Walsh and an anonymous reviewer for valuable comments on this manuscript. This work was supported by Roy J Carver Charitable Trust Grant 05-2258 and National Science Foundation Grant DEB-03-44209 to JMC.

## References

- Aguadé M, Langley CH (1994). Polymorphism and divergence in regions of low recombination in *Drosophila*. In: Golding GB (ed). *Non-Neutral Evolution: Theories and Molecular Data*. Chapman & Hall: New York, pp 67–76.
- Aguadé M, Miyashita N, Langley CH (1989). Restriction-map variation at the zeste-tko region in natural populations of *Drosophila melanogaster*. *Mol Biol Evol* **6**: 123–130.
- Akashi H (1995). Inferring weak selection from patterns of polymorphism and divergence at ‘silent’ sites in *Drosophila* DNA. *Genetics* **139**: 1067–1076.
- Akashi H (1996). Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics* **144**: 1297–1307.
- Akashi H, Ko WY, Piao S, John A, Goel P, Lin CF et al. (2006). Molecular evolution in the *Drosophila melanogaster* species subgroup: frequent parameter fluctuations on the timescale of molecular divergence. *Genetics* **172**: 1711–1726.
- Akashi H, Schaeffer SW (1997). Natural selection and the frequency distributions of ‘silent’ DNA polymorphism in *Drosophila*. *Genetics* **146**: 295–307.
- Andolfatto P (2005). Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**: 1149–1152.
- Andolfatto P, Przeworski M (2001). Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. *Genetics* **158**: 657–665.
- Aquadro CF, Begun DJ, Kindahl EC (1994). Selection, recombination and DNA polymorphism in *Drosophila*. In: Golding B (ed). *Non-Neutral Evolution: Theories and Molecular Data*. Chapman & Hall: New York.
- Bachtrog D (2003). Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in *Drosophila*. *Nat Genet* **34**: 215–219.
- Bachtrog D, Charlesworth B (2000). Reduced levels of microsatellite variability on the neo-Y chromosome of *Drosophila miranda*. *Curr Biol* **10**: 1025–1031.
- Bachtrog D, Charlesworth B (2002). Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**: 323–326.
- Barton N (1995a). Linkage and the limits to natural selection. *Genetics* **140**: 821–841.
- Barton NH (1995b). A general model for the evolution of recombination. *Genet Res* **65**: 123–145.
- Barton NH (1998). The effect of hitch-hiking on neutral genealogies. *Genet Res* **72**: 123–133.
- Barton NH, Charlesworth B (1998). Why sex and recombination? *Science* **281**: 1986–1990.
- Barton NH, Otto SP (2005). Evolution of recombination due to random drift. *Genetics* **169**: 2353–2370.
- Baudry E, Kerdelhue C, Innan H, Stephan W (2001). Species and recombination effects on DNA variability in the tomato genus. *Genetics* **158**: 1725–1735.
- Begun DJ (2001). The frequency distribution of nucleotide variation in *Drosophila simulans*. *Mol Biol Evol* **18**: 1343–1352.
- Begun DJ, Aquadro CF (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.
- 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 in *Drosophila*. *Proc Natl Acad Sci USA* **99**: 13616–13620.
- Birky Jr CW, Walsh JB (1988). Effects of linkage on rates of molecular evolution. *Proc Natl Acad Sci USA* **85**: 6414–6418.
- Bulmer M (1988). Codon usage and intragenic position. *J Theor Biol* **133**: 67–71.
- Bulmer M (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**: 897–907.
- Caballero A, Santiago E (1995). Response to selection from new mutation and effective size of partially inbred populations. I. Theoretical results. *Genet Res* **66**: 213–225.
- Carson HL (1958). Response to selection under different conditions of recombination in *Drosophila*. *Cold Spring Harb Symp Quant Biol* **23**: 291–306.
- Charlesworth B (1994). The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet Res* **63**: 213–227.
- Charlesworth B, Morgan MT, Charlesworth D (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- Cirulli ET, Kliman RM, Noor MA (2007). Fine-scale crossover rate heterogeneity in *Drosophila pseudoobscura*. *J Mol Evol* **64**: 129–135.
- Cohan FM (1984). Can uniform selection retard random genetic divergence between isolated conspecific populations? *Evolution* **38**: 495–504.
- Comeron JM, Aguade M (1996). Synonymous substitutions in the Xdh gene of *Drosophila*: heterogeneous distribution along the coding region. *Genetics* **144**: 1053–1062.

- Comeron JM, Guthrie TB (2005). Intragenic Hill–Robertson interference influences selection intensity on synonymous mutations in *Drosophila*. *Mol Biol Evol* **22**: 2519–2530.
- Comeron JM, Kreitman M (2000). The correlation between intron length and recombination in *Drosophila*. Dynamic equilibrium between mutational and selective forces. *Genetics* **156**: 1175–1190.
- Comeron JM, Kreitman M (2002). Population, evolutionary and genomic consequences of interference selection. *Genetics* **161**: 389–410.
- Comeron JM, Kreitman M, Aguade M (1999). Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. *Genetics* **151**: 239–249.
- Cutter AD, Payseur BA (2003). Selection at linked sites in the partial selfer *Caenorhabditis elegans*. *Mol Biol Evol* **20**: 665–673.
- Duret L, Mouchiroud D (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci USA* **96**: 4482–4487.
- Dvorak J, Luo MC, Yang ZL (1998). Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *aegilops* species. *Genetics* **148**: 423–434.
- Ewens WJ (1979). *Mathematical Population Genetics*. Springer-Verlag: Berlin.
- Eyre-Walker A (1996). The close proximity of *Escherichia coli* genes: consequences for stop codon and synonymous codon use. *J Mol Evol* **42**: 73–78.
- Eyre-Walker A, Bulmer M (1993). Reduced synonymous substitution rate at the start of enterobacterial genes. *Nucleic Acids Res* **21**: 4599–4603.
- Felsenstein J (1965). The effect of linkage on directional selection. *Genetics* **52**: 349–363.
- Felsenstein J (1974). The evolutionary advantage of recombination. *Genetics* **78**: 737–756.
- Felsenstein J, Yokoyama S (1976). The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* **83**: 845–859.
- Fisher RA (1930). *The Genetical Theory of Natural Selection*. Oxford University Press: New York.
- Fridolfsson A-K, Ellegren H (2000). Molecular Evolution of the avian CHD1 genes on the Z and W sex chromosomes. *Genetics* **155**: 1903–1912.
- Gillespie JH (1994). Alternatives to the neutral theory. In: Golding B (ed). *Non-Neutral Evolution: Theories and Molecular Data*. Chapman and Hall: New York, pp 1–17.
- Gillespie JH (2000). Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* **155**: 909–919.
- 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.
- Haldane JBS (1957). The cost of natural selection. *J Genet* **55**: 511–524.
- Hellmann I, Ebersberger I, Ptak SE, Paabo S, Przeworski M (2003). A neutral explanation for the correlation of diversity with recombination rates in humans. *Am J Hum Genet* **72**: 1527–1535.
- Hey J (1998). Selfish genes, pleiotropy and the origin of recombination. *Genetics* **149**: 2089–2097.
- Hey J, Kliman RM (2002). Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* **160**: 595–608.
- Hill WG, Robertson A (1966). The effect of linkage on limits to artificial selection. *Genet Res* **8**: 269–294.
- Hilton H, Kliman RM, Hey J (1994). Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. *Evolution* **48**: 1900–1913.
- Hudson RR, Kaplan NL (1995). Deleterious background selection with recombination. *Genetics* **141**: 1605–1617.
- Iles MM, Walters K, Cannings C (2003). Recombination can evolve in large finite populations given selection on sufficient loci. *Genetics* **165**: 2249–2258.
- 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.
- Kaplan NL, Hudson RR, Langley CH (1989). The “hitchhiking effect” revisited. *Genetics* **123**: 887–899.
- Keightley PD, Otto SP (2006). Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**: 89–92.
- Kern AD, Begun DJ (2005). Patterns of polymorphism and divergence from noncoding sequences of *Drosophila melanogaster* and *D. simulans*: evidence for nonequilibrium processes. *Mol Biol Evol* **22**: 51–62.
- Kim Y, Stephan W (2003). Selective sweeps in the presence of interference among partially linked Loci. *Genetics* **164**: 389–398.
- Kliman RM (1999). Recent selection on synonymous codon usage in *Drosophila*. *J Mol Evol* **49**: 343–351.
- Kliman RM, Hey J (1993a). DNA sequence variation at the period locus within and among species of the *Drosophila melanogaster* complex. *Genetics* **133**: 375–387.
- Kliman RM, Hey J (1993b). Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol Biol Evol* **10**: 1239–1258.
- Kliman RM, Hey J (2003a). Hill–Robertson interference in *Drosophila melanogaster*: reply to Marais, Mouchiroud and Duret. *Genet Res* **81**: 89–90.
- Kliman RM, Irving N, Santiago M (2003). Selection conflicts, gene expression, and codon usage trends in yeast. *J Mol Evol* **57**: 98–109.
- Kondrashov AS (1994). Muller’s ratchet under epistatic selection. *Genetics* **136**: 1469–1473.
- Kraft T, Sall T, Magnusson-Rading I, Nilsson NO, Hallden C (1998). Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). *Genetics* **150**: 1239–1244.
- Langley CH, MacDonald J, Miyashita N, Aguade M (1993). Lack of correlation between interspecific divergence and intraspecific polymorphism at the suppressor of forked region in *Drosophila melanogaster* and *Drosophila simulans*. *Proc Natl Acad Sci USA* **90**: 1800–1803.
- Li WH (1987). Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. *J Mol Evol* **24**: 337–345.
- Loewe L, Charlesworth B (2007). Background selection in single genes may explain patterns of codon bias. *Genetics* **175**: 1381–1393.
- Lynch MJB (1998). Deleterious mutation accumulation in organelle genomes. *Genetica* **102–103**: 29–39.
- Marais G, Mouchiroud D, Duret L (2001). Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc Natl Acad Sci USA* **98**: 5688–5692.
- Marais G, Mouchiroud D, Duret L (2003). Neutral effect of recombination on base composition in *Drosophila*. *Genet Res* **81**: 79–87.
- Marais G, Piganeau G (2002). Hill–Robertson interference is a minor determinant of variations in codon bias across *Drosophila melanogaster* and *Caenorhabditis elegans* genomes. *Mol Biol Evol* **19**: 1399–1406.
- Martin G, Otto SP, Lenormand T (2006). Selection for recombination in structured populations. *Genetics* **172**: 593–609.
- Martin-Campos JM, Comeron JM, Miyashita N, Aguade M (1992). Intraspecific and interspecific variation at the y-ac-sc region of *Drosophila simulans* and *Drosophila melanogaster*. *Genetics* **130**: 805–816.
- Maynard Smith J, Haigh J (1974). The hitch-hiking effect of a favorable gene. *Genet Res* **23**: 23–35.

- McPhee CP, Robertson A (1970). The effect of suppressing crossing-over on the response to selection in *Drosophila melanogaster*. *Genet Res* **16**: 1–16.
- McVean GA, Charlesworth B (1999). A population genetic model for the evolution of synonymous codon usage: patterns and predictions. *Genet Res* **74**: 145–158.
- McVean GA, Charlesworth B (2000). The effects of Hill–Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**: 929–944.
- McVean GA, Vieira J (2001). Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*. *Genetics* **157**: 245–257.
- Moriyama EN, Hartl DL (1993). Codon usage bias and base composition of nuclear genes in *Drosophila*. *Genetics* **134**: 847–858.
- Moriyama EN, Powell JR (1996). Intraspecific nuclear DNA variation in *Drosophila*. *Mol Biol Evol* **13**: 261–277.
- Moriyama EN, Powell JR (1998). Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Res* **26**: 3188–3193.
- Muller HJ (1964). The relation of recombination to mutational advance. *Mutat Res* **1**: 2–9.
- Nachman MW (1997). Patterns of DNA variability at X-linked loci in *Mus domesticus*. *Genetics* **147**: 1303–1316.
- Nachman MW, Bauer VL, Crowell SL, Aquadro CF (1998). DNA variability and recombination rates at X-linked loci in humans. *Genetics* **150**: 1133–1141.
- Nagylaki T (1983). Evolution of a finite population under gene conversion. *Proc Natl Acad Sci USA* **80**: 6278–6281.
- Nordborg M, Charlesworth B, Charlesworth D (1996). The effect of recombination on background selection. *Genet Res* **67**: 159–174.
- Ohta T, Kimura M (1971). On the constancy of the evolutionary rate of cistrons. *J Mol Evol* **1**: 18–25.
- Ohta T, Kimura M (1975). The effect of selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect). *Genet Res* **25**: 313–326.
- Otto SP, Barton NH (1997). The evolution of recombination: removing the limits to natural selection. *Genetics* **147**: 879–906.
- Otto SP, Barton NH (2001). Selection for recombination in small populations. *Evolution* **55**: 1921–1931.
- Otto SP, Lenormand T (2002). Resolving the paradox of sex and recombination. *Nat Rev Genet* **3**: 252–261.
- Piganeau G, Westrelin R, Tourancheau B, Gautier C (2001). Multiplicative versus additive selection in relation to genome evolution: a simulation study. *Genet Res* **78**: 171–175.
- Powell JR, Moriyama EN (1997). Evolution of codon usage bias in *Drosophila*. *Proc Natl Acad Sci USA* **94**: 7784–7790.
- Przeworski M, Hudson RR, Di Rienzo A (2000). Adjusting the focus on human variation. *Trends Genet* **16**: 296–302.
- Qin H, Wu WB, Comeron JM, Kreitman M, Li WH (2004). Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes. *Genetics* **168**: 2245–2260.
- Rice WR (1994). Degeneration of a nonrecombining chromosome. *Science* **263**: 230–232.
- Rice WR, Chippindale AK (2001). Sexual recombination and the power of natural selection. *Science* **294**: 555–559.
- Robertson A (1960). A theory of limits in artificial selection. *Proc R Soc Lond B Biol Sci* **153**: 234–249.
- Robertson A (1961). Inbreeding in artificial selection programmes. *Genet Res* **2**: 189–194.
- Roze D, Barton NH (2006). The Hill–Robertson effect and the evolution of recombination. *Genetics* **173**: 1793–1811.
- Schaeffer SW (2002). Molecular population genetics of sequence length diversity in the Adh region of *Drosophila pseudoobscura*. *Genet Res* **80**: 163–175.
- Singh ND, Arndt PF, Petrov DA (2005). Genomic heterogeneity of background substitutional patterns in *Drosophila melanogaster*. *Genetics* **169**: 709–722.
- Stephan W (1994). Effects of genetic recombination and population subdivision on nucleotide sequence variation in *Drosophila ananassae*. In: Golding B (ed). *Non-Neutral Evolution: Theories and Molecular Data*. Chapman and Hall: New York. pp 57–66.
- 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.
- Stephan W, Langley CH (1998). DNA polymorphism in lycopersicon and crossing-over per physical length. *Genetics* **150**: 1585–1593.
- Stephan W, Wiehe THE, Lenz MW (1992). The effect of strongly selected substitutions on neutral polymorphism-analytical results based on diffusion theory. *Theor Popul Biol* **41**: 237–254.
- Tachida H (2000). Molecular evolution in a multisite nearly neutral mutation model. *J Mol Evol* **50**: 69–81.
- Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Takano TS (1998). Rate variation of DNA sequence evolution in the *Drosophila* lineages. *Genetics* **149**: 959–970.
- Wang J, Hill WG, Charlesworth D, Charlesworth B (1999). Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. *Genet Res* **74**: 165–178.
- Wang W, Thornton K, Berry A, Long M (2002). Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science* **295**: 134–137.
- Weinreich DM, Rand DM (2000). Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. *Genetics* **156**: 385–399.
- Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.
- Wright S (1938). Size of population and breeding structure in relation to evolution. *Science* **87**: 430–431.
- Zhang Z, Parsch J (2005). Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression. *Mol Biol Evol* **22**: 1945–1947.
- Zurovcova M, Eanes WF (1999). Lack of nucleotide polymorphism in the Y-linked sperm flagellar dynein gene Dhc-Yh3 of *Drosophila melanogaster* and *D. simulans*. *Genetics* **153**: 1709–1715.