

# Weak Selection and Protein Evolution

Hiroshi Akashi,<sup>1</sup> Naoki Osada, and Tomoko Ohta

Division of Evolutionary Genetics, Department of Population Genetics, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

**ABSTRACT** The “nearly neutral” theory of molecular evolution proposes that many features of genomes arise from the interaction of three weak evolutionary forces: mutation, genetic drift, and natural selection acting at its limit of efficacy. Such forces generally have little impact on allele frequencies within populations from generation to generation but can have substantial effects on long-term evolution. The evolutionary dynamics of weakly selected mutations are highly sensitive to population size, and near neutrality was initially proposed as an adjustment to the neutral theory to account for general patterns in available protein and DNA variation data. Here, we review the motivation for the nearly neutral theory, discuss the structure of the model and its predictions, and evaluate current empirical support for interactions among weak evolutionary forces in protein evolution. Near neutrality may be a prevalent mode of evolution across a range of functional categories of mutations and taxa. However, multiple evolutionary mechanisms (including adaptive evolution, linked selection, changes in fitness-effect distributions, and weak selection) can often explain the same patterns of genome variation. Strong parameter sensitivity remains a limitation of the nearly neutral model, and we discuss concave fitness functions as a plausible underlying basis for weak selection.

**U**NDER the neutral model, newly arising mutations fall into two major fitness classes: strongly deleterious and selectively neutral (Kimura 1968; King and Jukes 1969). The first class is well supported by mutation accumulation experiments (reviewed in Simmons and Crow 1977; Halligan and Keightley 2009) and early DNA sequence comparisons (Grunstein *et al.* 1976; Kafatos *et al.* 1977) and is shared among competing evolutionary models. The novel and controversial aspect of the neutral theory was the proposition that, among mutations that go to fixation, the vast majority are selectively neutral. Advantageous substitutions, although important in phenotypic evolution, are sufficiently rare at the molecular level that they need not be considered to adequately model the process. Under the neutral theory, within- and between-species variation sample two aspects of a process of origination by mutation and changes in gene frequency dominated by drift and, for some mutations, negative selection (Kimura and Ohta 1971a). In contrast,

polymorphism and divergence may be “uncoupled” under selection models (Gillespie 1987).

## *Protein polymorphism and the neutral model: invariance of heterozygosity*

Clear predictions for levels of polymorphism within populations and divergence among species are appealing aspects of the neutral model. However, within a few years of its proposal, the notion of drift-dominated evolution was challenged by overall patterns of allozyme polymorphism and contrasts between DNA and protein divergence.

Although evolutionary geneticists were generally surprised by the extent of naturally occurring variation revealed by allozyme gel electrophoresis in the 1970s (Lewontin 1991), the *lack* of species with high polymorphism levels became a central problem for proponents of the neutral theory (Robertson 1968; Maynard Smith 1970a; Lewontin 1974). The neutral model makes a simple prediction for protein heterozygosity, a summary statistic for levels of naturally occurring variation. Heterozygosity,  $H$ , is defined as the probability of randomly sampling different alleles from a population in two independent trials. Kimura and Crow (1964) determined expected heterozygosity for neutral alleles in an “idealized” population of constant size with random mating among hermaphroditic individuals. Generations are nonoverlapping in this scenario, and offspring are

Copyright © 2012 by the Genetics Society of America  
doi: 10.1534/genetics.112.140178

Manuscript received March 1, 2012; accepted for publication June 11, 2012

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.140178/-/DC1>.

This article is dedicated to the memory of James F. Crow.

<sup>1</sup>Corresponding author: Division of Evolutionary Genetics, Department of Population Genetics, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan. E-mail: hakashi@nig.ac.jp

generated by randomly sampling gametes from the parents. Under such conditions, referred to as a Wright–Fisher population, and in the absence of mutation, genetic drift leads to a loss of heterozygosity at a rate inversely related to the population size. If new mutations arise in each generation, heterozygosity will reach an equilibrium between mutational input and loss by drift:

$$H = 4Nv/(1 + 4Nv),$$

where  $v$  is the per-generation mutation rate to new, neutral alleles, and  $N$  is the number of diploid individuals in the population. If all amino acid changes are either strongly deleterious or selectively neutral,  $v$  can be replaced with the product of the fraction of neutral mutations ( $f_n$ ) and the total mutation rate ( $u$ ). To employ results for Wright–Fisher populations to predict heterozygosity (and other aspects of genetic drift) in populations that violate assumptions of the model,  $N$  is replaced by the “effective” population size ( $N_e$ ) (see Charlesworth 2009). Incorporating both substitutions gives an expected heterozygosity of  $H = 4N_e f_n u / (1 + 4N_e f_n u)$ .

Heterozygosity under the neutral model can be predicted given values for mutation rate and effective population size. Estimates for both parameters are crude, but Maynard Smith (1970b) and Nei and Graur (1984) predicted that levels of protein heterozygosity in large natural populations should approach the upper limit of 1. Although some species show close to zero allozyme variation, very few studies found  $H > 0.30$ . Species such as humans, *Drosophila melanogaster*, and *Escherichia coli* show roughly similar levels of protein polymorphism although their historical population sizes presumably differ greatly. This “invariance of heterozygosity” (Lewontin 1974) was argued as strong evidence against the neutral model.

### Protein divergence and the neutral model

The neutral model also makes simple predictions for evolutionary divergence. For proteins, new mutations fall into two main fitness classes: strongly deleterious mutations that natural selection quickly eliminates from populations and neutral mutations that drift to fixation with probabilities equal to their initial frequency. Neutral mutations have smaller fixation probabilities in larger populations than in smaller populations, but this difference is exactly matched by the higher mutational input in larger populations, and the expected rate of neutral divergence is simply the mutation rate (and is independent of population size) (Wright 1938).

The initial motivation for Kimura’s proposal of the neutral theory was absolute rates of protein evolution that appeared to violate theoretical upper limits for adaptive fixations (Kimura 1968), but the molecular clock was quickly adopted as one of the strongest pieces of evidence supporting neutral evolution (Kimura and Ohta 1971a). The clock-like nature of protein evolution (Zuckerandl and Pauling 1965) was considered to be in accord with the neutral theory’s predic-

tion of lack of dependence on population size and inconsistent with adaptive models that predict rate dependence on both population sizes and the particular ecology of species. To account for the molecular clock, Kimura and Ohta (1971a,b) proposed that mutation rates are approximately constant per year. In addition, they suggested that the fraction of neutral mutations varies among proteins (this explains the among-protein rate variation) but remains relatively constant over time (this explains the clock-like divergence for each protein) (Kimura and Ohta 1971b). Kimura’s views on the importance of weakly selected mutations fluctuated over his career (Kimura and Takahata 1995); we will refer to the initial formulation discussed above as the “neutral model.”

Contrasting patterns of protein and DNA divergence forced a reconsideration of the fit of protein molecular clocks to the strict neutral model (Ohta 1972a). DNA divergence, accessed through DNA–DNA hybridization (Laird and McCarthy 1968; Kohne 1970), appeared to show a generation time effect (greater divergence in lineages experiencing larger numbers of generations per year) in contrast to the absolute-time dependence of protein evolution. The ratio of DNA to protein divergence varied considerably (>10-fold) among lineages. Under the neutral model, this ratio gives an estimate of  $1/f_n$  for proteins (assuming  $f_n = 1$  for DNA divergence), and such variation violated a tenet of the neutral model (Kimura and Ohta 1971a).

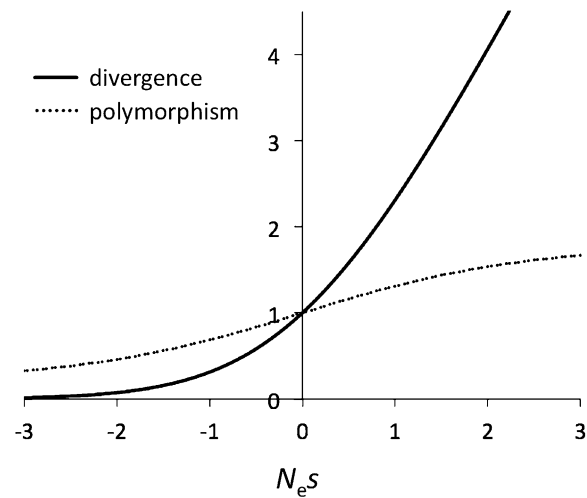
Ohta proposed that “nearly neutral” mutations could explain both the upper limit of protein heterozygosity and excess variation in protein divergence scaled to DNA divergence. The efficacy of selection depends on the product of selection coefficient and effective population size ( $N_e s$ ), referred to as “selection intensity” or “scaled selective effect.” The model assumes a large fraction of newly arising mutations with “borderline” fitness effects between clearly deleterious ( $N_e s \ll -1$ ) and selectively neutral ( $|N_e s| \ll 1$ ). Amino acid changes with subtle effects on protein folding could fall into this category (Barnard *et al.* 1972; Ohta 1973). For mutations in the nearly neutral range, the balance between the influence of genetic drift and natural selection is strongly dependent on effective population size. Figure 1 shows expected levels of within- and between-species variation for mutations with small fitness effects. Positive and negative selection increase and decrease expected levels of polymorphism and divergence, respectively, but the impact of selection is greater on divergence than on polymorphism. Interestingly, the rate of increase in polymorphism declines with the magnitude of positive selection. Although more strongly selected advantageous alleles have a reduced probability of loss by drift, they also show a shorter transit time within populations (and thus have a smaller probability of being sampled). Slightly deleterious mutations,  $N_e s \approx -1$ , have non-negligible probabilities of being sampled within populations and going to fixation relative to neutral mutations, but fixation probabilities drop to essentially zero for  $N_e s < -3$ . Ohta defined nearly neutral

mutations as those “whose selection coefficients are so small that their behavior is not very different from strictly neutral mutants. Operationally, this is defined by  $|N_e s| < 1$  (Ohta 1972b). The relative strengths of selection and genetic drift shift gradually near  $|N_e s| \approx 1$ , so “near neutrality” cannot be precisely defined (especially for  $s > 0$ ). In this section, we employ Ohta’s definition of effectively neutral mutations in the range  $|N_e s| < 1$  and consider only deleterious mutations.

Under weak selection, expected heterozygosity becomes  $H \approx 4N_e f'_n u / (1 + 4N_e f'_n u)$ , where  $f'_n$  from the neutral model is substituted by the fraction of nearly neutral mutations ( $f'_n$ ).  $f'_n$  is assumed (under the original neutral model) to be generally constant among lineages but  $f'_n$  is very sensitive to effective population size. For nearly neutral mutations,  $f'_n$  decreases as  $N_e$  increases; thus, heterozygosity increases more slowly as a function of population size than for neutral alleles (Ohta 1974; Ohta and Kimura 1975). Excess rare (low frequency) variants observed for allozyme data in a number of species (Latter 1975; Ohta 1975; Chakraborty *et al.* 1980) are consistent with slightly deleterious protein polymorphism.

Neutral and adaptive scenarios were also proposed to account for the invariance of allozyme heterozygosity. The expected relationship between  $H$  and  $N_e$  assumes that populations have reached a steady-state level of polymorphism. Nei and Graur (1984) developed a demographic hypothesis to explain the upper limit on heterozygosity; population sizes may fluctuate considerably on evolutionary timescales, and effective population size is especially sensitive to strong reductions. Consistently low heterozygosities result from bottlenecks during glaciation events; current population samples for many species reflect the recovery of neutral variation in expanding populations since the last glaciation. The approach to equilibrium heterozygosity can require 4–8N generations after population size changes (Nei and Graur 1984; Tajima 1989), and frequency spectra will be skewed toward rare polymorphisms during the recovery phase. Maynard Smith and Haigh (1974) showed that adaptive fixations can cause bottleneck-like reductions in variation and excess rare variants at neutral sites that are genetically linked to advantageous mutations, a process termed “genetic hitchhiking.” Deleterious mutations also reduce levels of neutral variation at linked sites (Charlesworth *et al.* 1995). Population bottlenecks and linked selection may contribute to the upper limit of heterozygosity for neutral variation, but weak selection in protein evolution also predicts differences in *ratios* of protein to synonymous polymorphism among populations (see below).

Nearly neutral mutations were also invoked to explain variation in rates of protein evolution. As noted above, DNA divergence appeared to be reduced in lineages with long generation times whereas protein evolution showed a more clock-like dependence on absolute time. Ohta and Kimura (1971) and Ohta (1972a) postulated an inverse relationship between generation time and effective population size to explain this discordance [e.g., among mammals, humans

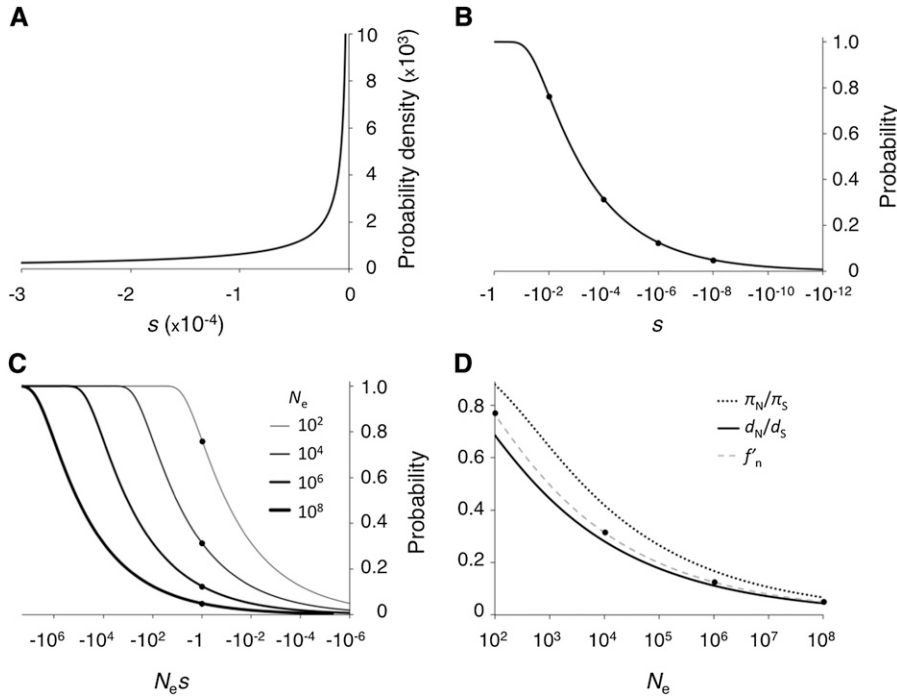


**Figure 1** Polymorphism and divergence under weak selection. Expected levels of nucleotide diversity and DNA divergence (each relative to neutral mutations) are shown. The dotted line represents nucleotide diversity (probability of observing a polymorphism at a given nucleotide site in a pair of randomly chosen chromosomes) and is calculated using sampling formulas from Sawyer and Hartl (1992), assuming an infinite-sites mutation model and constant  $N_e$ . The solid line shows the fixation probabilities of mutations (Kimura 1962). The plots assume directional (genic) selection with fitness values 1,  $1 + 1/2s$ , and  $1 + s$  for the homozygote for the ancestral allele, heterozygote, and homozygote for a new mutation, respectively. The plots assume independent evolution among sites and are based on Kimura (1983, p. 44).

and elephants have longer generation times and presumably smaller population sizes than rodents (see Chao and Carr 1993)]. If mutation rates are expressed in units of generations, then the rate of protein divergence (nonsynonymous substitutions per site) in absolute time is  $f'_n u g$ , where  $u$  is the mutation rate per generation and  $g$  is the number of generations per year. An inverse relationship between  $f'_n$  and  $g$  can produce clock-like protein evolution if  $f'_n g$  is roughly constant. Because lineages that undergo fewer generations per year also experience less effective negative selection against slightly deleterious mutations, the ratio of protein to DNA divergence is positively correlated with generation time.

### Structure of the Nearly Neutral Model

The nearly neutral model posits a distribution of selection coefficients with a large fraction of new mutations with fitness effects near the reciprocal of population sizes found in nature. This section will give an illustration of the model and its predictions. Figure 2A shows an example of the probability distribution of selective effects (DSE) of new mutations that puts a substantial density in the nearly neutral range for a wide range of population sizes,  $10^2 < N_e < 10^8$ . Initially, we considered only deleterious mutations. It is important to note that “weakly selected” refers to the magnitude of  $N_e s$  rather than to the functional or fitness effects of mutations. This distinction is critical because empirical studies define “strong” and “weak” effects relative either



**Figure 2** Example evolutionary patterns under slightly deleterious mutations. (A) A probability density function for negative selection coefficients (gamma distribution with shape parameter 0.2 and scale parameter 0.05). The area under the curve gives the proportion of mutations in a given fitness range. This distribution of  $s$  was chosen to allow substantial increases in the effectively neutral proportion for population sizes in the range  $10^2$ – $10^8$  and is assumed in plots in B, C, and D. Under this distribution of selective effects (DSEs), <25% of newly arising mutations have  $s < -0.01$  and <2% of mutations have  $s < -0.1$ . (B) Cumulative distribution function for selective effects of new mutations. y-axis values are the total areas under the curve in A for  $x < s < 0$ .  $f_n$ , the proportion of “effectively neutral” mutations,  $-1 < N_e s \leq 0$ , for a given population size is the y-axis value at  $x = 1/N_e$  (values are marked for  $N_e$  of  $10^2$ ,  $10^4$ ,  $10^6$ ,  $10^8$ ). (C) Cumulative distribution function for  $N_e s$ . y-axis values are the areas under the DSE curve for  $x < N_e s < 0$  in A. Curves are shown for  $N_e$  of  $10^2$ ,  $10^4$ ,  $10^6$ , and  $10^8$  (thicker lines represent larger population sizes).  $f_n$  values are indicated (as solid circles) for each population size. (D) Polymorphism and divergence as

a function of  $N_e$ . Expected DNA diversity ( $\pi_N/\pi_S$ , dotted line) and divergence ( $d_N/d_S$ , solid line) are shown. The dashed line shows  $f_n$ , and values for  $N_e$  of  $10^2$ ,  $10^4$ ,  $10^6$ , and  $10^8$  are marked. Expected divergence is smaller than  $f_n$  because selection reduces fixation rates for slightly deleterious mutations within this range.  $\pi_N/\pi_S$  values are higher than  $f_n$  because mutations in the range  $N_e s < -1$  contribute to polymorphism (Figure 1). These plots assume independent evolution among sites.

to an absolute scale or to the sensitivity of the assays used. The limit of detection for laboratory studies, even in large-scale fitness assays for microbes, is usually  $s > 0.001$  (e.g., Dykhuizen and Hartl 1983; Lind *et al.* 2010; Hietpas *et al.* 2011). “Weakly selected” alleles in such studies would be well outside the nearly neutral range in many natural populations; mutations of  $|N_e s| \approx 1$  in populations with effective sizes  $N_e > 10^4$  may be undetectable in fitness and phenotype assays.

Figure 2B shows the cumulative distribution function for the DSE shown in Figure 2A. Expected levels of polymorphism and divergence are functions of scaled selective effects,  $N_e s$  (Figure 1), and Figure 2C shows the cumulative distribution functions for  $N_e s$  across a range of population sizes. Many mutations that fall into the “effectively neutral” range in small populations are strongly selected in larger populations, and  $f_n$  varies considerably as a function of  $N_e$ . Expected polymorphism and divergence patterns under this DSE are shown in Figure 2D. Polymorphism is less reduced under weak selection than divergence but both decrease at substantial rates over a wide range of  $N_e$ .

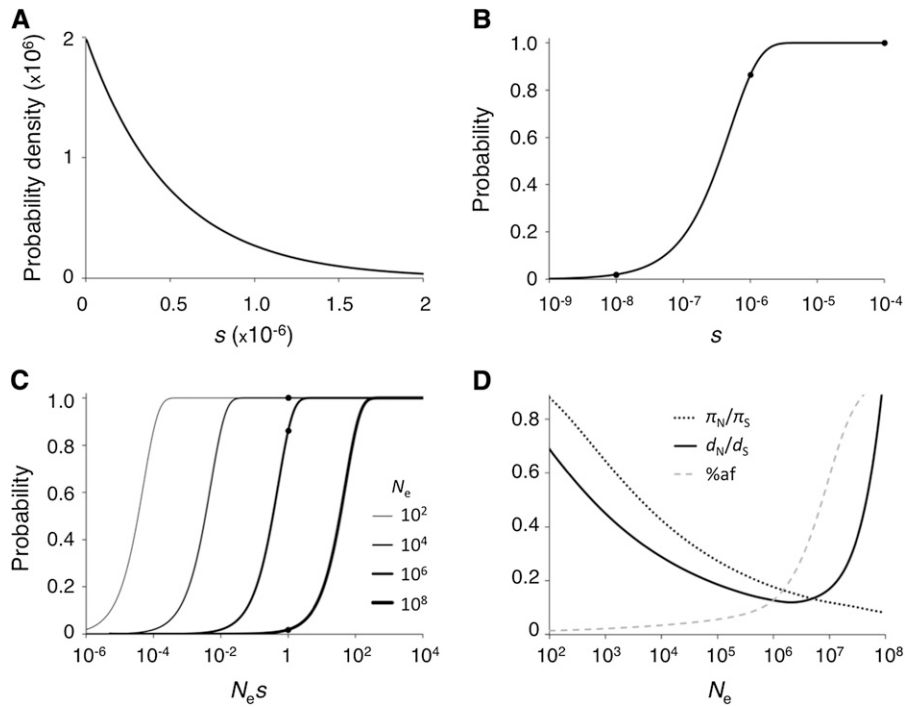
In the analyses above, we have assumed no new mutations that enhance fitness,  $s > 0$ . Adaptive evolution may often occur in response to environmental change. Because slightly deleterious fixations go to fixation at appreciable rates, nearly neutral models must allow for slightly advantageous mutations, including back mutations and compensatory mutations (Ohta 1972b, 1973; Latter 1975; Ohta and Tachida 1990), even in the absence of pressure for novel function. Compensatory evolution retards the slow decline in fitness by slightly

deleterious fixations and will be more effective if some compensatory mutations have large positive fitness effects (i.e., if they can compensate for multiple deleterious fixations).

Figure 3A shows an example DSE for advantageous mutations. Even a small fraction of such mutations can have a substantial impact on divergence (we assume that 99% of new mutations are drawn from the deleterious DSE shown in Figure 2A and that 1% of mutations are drawn from the DSE shown in Figure 3A). Expected polymorphism and divergence under the combined deleterious and adaptive DSE are shown in Figure 3D. Advantageous mutations contribute little to polymorphism, and DNA diversity shows a negative relationship with  $N_e$  similar to the deleterious mutations case. However, larger populations have smaller proportions of deleterious, and larger proportions of adaptive, fixations. The fixation rate of new mutations begins to *increase* as a function of population size when the proportion of effectively selected positive mutations ( $N_e s > 1$ ) becomes large. The effect of population size on expected polymorphism and divergence can be highly sensitive to the particular DSE; we present a single example to illustrate relationships among DSEs, population size, and evolutionary patterns under the nearly neutral model.

## Evidence for Weak Selection in Protein Evolution

Both DNA variation data and statistical methods for testing evolutionary mechanisms have proliferated tremendously in the past decade, and we summarize key findings related to the nearly neutral model below.



**Figure 3** Example evolutionary patterns for slightly deleterious and advantageous mutations. (A) Probability density function for positive selection coefficients (gamma distribution with shape parameter 1 and scale parameter  $5 \times 10^{-7}$ ). (B) Cumulative distribution function for selective effects of new mutations. y-Axis values are the total areas under the curve for  $0 < s < x$  in A.  $f'_n$  for a given population size is the y-axis value at  $x = 1/N_e$  (values are marked for  $N_e$  of  $10^4$ ,  $10^6$ ,  $10^8$ ). Almost all beneficial mutations are effectively neutral in  $N_e$  of  $10^2$  and  $10^4$ , and almost none are effectively neutral in  $N_e$  of  $10^8$ . (C) Cumulative distribution function for  $N_e s$  or scaled selective effects. y-Axis values are the areas under the DSE curve for  $0 < N_e s < x$  in A. Curves are shown for  $N_e$  values of  $10^2$ ,  $10^4$ ,  $10^6$ , and  $10^8$  (thicker lines represent larger population sizes).  $f'_n$  values are shown as solid circles for each population size. (D) Polymorphism and divergence as a function of  $N_e$  for a distribution of fitness effects that combines the density functions in Figure 2A (99% of new mutations) and Figure 3A (1% of new mutations). The dashed line shows the proportion of advantageous fixations. Expected DNA diversity ( $\pi_N/\pi_S$ , dotted line) and divergence ( $d_N/d_S$ , solid line) are shown. These predictions assume independent evolution among sites.

### Population size and rates of evolution

Negative relationships between rates of protein evolution and population size are a clear prediction of near neutrality; Ohta (1972a) showed such correlations in *Drosophila* and mammals. Although generation time may be a noisy predictor of  $N_e$  and DNA–DNA hybridization gives only rough measures of distance, Ohta’s study is notable for introducing an early form of comparison of nonsynonymous and putatively neutral divergence to test evolutionary models. More recent analyses often employ the ratio of nonsynonymous to synonymous DNA divergence per site ( $d_N/d_S$ ) to estimate  $f_n$  or  $f'_n$  under the neutral and nearly neutral models, respectively.

Several genome-scale comparisons support the weak-selection prediction of inverse relationships between  $d_N/d_S$  and population size. Faster protein evolution in primates compared to rodents was observed in early studies limited to small numbers of genes (Li *et al.* 1987; Ohta 1995), and the pattern has been confirmed in larger-scale analyses (Chimpanzee Sequencing and Analysis Consortium 2005; Rhesus Macaque Genome Sequencing and Analysis Consortium *et al.* 2007). In addition, substitutions among amino acids with greater physicochemical differences have been more frequent in primates than in rodents (Zhang 2000; Eyre-Walker *et al.* 2002; Hughes and Friedman 2009). Analyses of nuclear genes in a broader range of mammals revealed a twofold variation in  $d_N/d_S$  and negative relationships between  $d_N/d_S$  and estimates of population size (Lindblad-Toh *et al.* 2005; Kosiol *et al.* 2008; Ellegren 2009). Wright and Andolfatto (2008) noted that such a relationship extends to comparisons among pairs of bacteria, *Drosophila*, plants, and mammals. The relationship is strong, but DSEs may not be conserved across distantly related organisms.

Independent comparisons between host-dependent bacterial lineages and their free-living relatives show consistently faster evolution in endosymbionts (Moran 1996; Wernegreen and Moran 1999) as well as pathogens (Andersson and Andersson 1999; Warnecke and Rocha 2011). Wernegreen (2011) also noted a higher ratio of radical-to-conservative amino acid replacement changes in insect symbionts relative to free-living relatives. Vertically transmitted symbionts, such as *Buchnera* in aphids, experience population bottlenecks during transmission from parent to embryo as well as population-size fluctuations of the host species. In addition, symbiont genomes experience limited opportunities for recombination; genetic linkage among selected mutations reduces the efficacy of natural selection in a manner similar to reduced  $N_e$  (Hill and Robertson 1966; Felsenstein 1974; Birky and Walsh 1988; Charlesworth 1994; Barton 1995). Low heterozygosity at synonymous positions is consistent with small population sizes of bacterial endosymbionts (Funk *et al.* 2001; Abbot and Moran 2002; Herbeck *et al.* 2003).

Because reduced  $N_e$  is associated with the ecology of host dependence, shifts in DSEs should also be considered as a cause of elevated protein divergence. In particular, “relaxed selection,” an increased density of mutations with very small effects, is plausible in an intracellular environment that may be more stable than environments of free-living microbes (especially for insect and animal hosts) and from which many metabolites can be obtained directly (Moran 1996). In this scenario, an increase in the density of mutations with fitness effects in the range  $-1 < N_e s < 0$  reflects an elevated proportion of amino acid changes with smaller selection coefficients rather than (or in addition to) a

reduced  $N_e$ . Interestingly, several fast-evolving endosymbiont lineages show high expression of heat-shock proteins (McCutcheon and Moran 2011). GroEL, an ATP-dependent bacterial chaperonin that assists in protein folding, is among the most abundant proteins in independently derived endosymbionts. Elevated heat-shock protein expression may be part of a compensatory response to slightly deleterious fixations that destabilize proteins (Moran 1996; van Ham *et al.* 2003; Fares *et al.* 2004) and could be an example of a relatively small number of adaptive changes counteracting a much larger number of deleterious fixations. GroEL overexpression can enhance growth rates of laboratory strains of bacteria that have accumulated mutations but show a cost in the ancestral strains (Moran 1996; Fares *et al.* 2002). Interestingly, this cost is apparent only when ancestral strains are grown in amino-acid-limited environments. Elevated heat-shock protein expression may be favored in endosymbiont lineages that have accumulated destabilizing mutations. Enhanced protein folding should reduce the fitness effects of further destabilizing amino acid changes (*i.e.*, shift the distribution of selection coefficients) (Tokuriki and Tawfik 2009a). Thus, elevated heat-shock protein expression and accumulation of destabilizing mutations may form a positive feedback loop until the benefits of further chaperonin overexpression no longer outweigh the costs.

A number of cases of accelerated divergence in small populations have been documented for proteins encoded in mitochondrial genomes. Island bird species show elevated rates of protein evolution relative to their mainland counterparts (Johnson and Seger 2001), and similar patterns hold for other vertebrates as well as invertebrates (Woolfit and Bromham 2005). Higher  $d_N/d_S$  in obligately asexual lineages than in sexual lineages of *Daphnia* (Paland and Lynch 2006) and freshwater snails (Johnson and Howard 2007; Neiman *et al.* 2010) is consistent with elevated Hill–Robertson interference with reduced opportunities for recombination [between mitochondrial DNA (mtDNA) and nuclear genomes] and/or greater frequencies of founder events in asexual lineages (Glémin and Galtier 2012). However, asexual lineages are often short-lived, tip branches on the phylogeny, and  $d_N/d_S$  for these lineages may include a greater fraction of polymorphic mutations than in sexual lineages. Both  $d_N/d_S$  and ratios of radical-to-conservative amino acid fixations are positively correlated to body mass for mtDNA-encoded proteins (13 genes) among >100 species of mammals (Popadin *et al.* 2007). If body mass and  $N_e$  are inversely related (Damuth 1981), these patterns support slightly deleterious fixations in large mammals. In many of the examples above, lineages that are thought to differ in  $N_e$  also differ in their ecology; differences in DSEs remain a plausible alternative, or a contributing factor, to heterogeneity in protein divergence. In addition, Bazin *et al.* (2006) found little evidence for positive correlations between mtDNA heterozygosity and proxies for effective population size (allozyme and nuclear DNA variation) across a wide range of taxa. They suggested that complete genetic linkage

in mtDNA genomes and recurrent positive selection (Gillespie 2001) may underlie such patterns. However, within-taxa comparisons show positive correlations between mtDNA heterozygosity and nuclear genome variation in mammals (Mulligan *et al.* 2006; Nabholz *et al.* 2008) and other animals (Piganeau and Eyre-Walker 2009). Variation in mutation rates, adaptive evolution, and/or biased transmission may contribute to within- and between-taxa differences for mtDNA polymorphism.

Within-genome comparisons can also test for relationships between  $N_e$  and protein evolution. Near neutrality predicts faster protein evolution in regions of reduced recombination, and such patterns have been confirmed in yeast (Connallon and Knowles 2007; Weber and Hurst 2009) and *Drosophila* (Campos *et al.* 2012; Mackay *et al.* 2012). Because these studies compare evolutionary rates among different genes, differences in DSEs may contribute to heterogeneity. Such differences may be related to the functional categories of genes that reside in regions of low recombination and/or to their expression level. Interestingly, in *Drosophila*, subsets of loci such as male-biased genes (Betancourt and Presgraves 2002; Presgraves 2005; Zhang and Parsch 2005) and *D. melanogaster* lineage accelerated genes (Larracuente *et al.* 2008) show the opposite trend: positive correlations between recombination and  $d_N/d_S$  consistent with limits to adaptive evolution under low recombination. Primates show no correlation between  $d_N/d_S$  and the rate of recombination (Bullaughay *et al.* 2008), but relationships between local recombination rates and DNA diversity in the human genome are weak (Hellmann *et al.* 2005).

Interpretations of among-lineage and within-genome  $d_N/d_S$  tests of near neutrality are complicated by recent evidence that a majority of protein changes are adaptively fixed in many species (reviewed in Fay 2011). Advantageous mutations go to fixation at higher rates in larger  $N_e$  when they are weakly selected and when adaptive evolution is mutation-limited. In such cases,  $d_N/d_S$  can show positive relationships with  $N_e$  (Figure 3D) even in the presence of slightly deleterious mutations.

### Population size and protein polymorphism

Several assumptions are necessary when attributing variation in protein divergence to near neutrality. Advantageous substitutions should be rare, and DSEs must remain constant among the lineages compared. In addition, relative ancestral effective population sizes must be known. Some studies assume relationships between  $N_e$  and the ecology of organisms (*e.g.*, host-cell dependence, island habitats), and other studies estimate  $N_e$  from current polymorphism levels. The former approach leaves open the possibility of associations between the “life style” of organisms and their DSEs. The latter method allows inference of  $N_e$  in roughly the last  $4N$  generations. In most cases, molecular divergence measured by  $d_N/d_S$  occurred at a much deeper time scale. Finally, inferring  $f_n$  from  $d_N/d_S$  requires an assumption of neutral

divergence at synonymous sites and accurate estimation of  $d_S$ . Both natural selection (reviewed in Akashi 2001; Chamary *et al.* 2006; Hershberg and Petrov 2008; Plotkin and Kudla 2010) and biased gene conversion (reviewed in Birdsell 2002; Marais 2003; Duret and Galtier 2009) can cause fixation biases in synonymous divergence. In addition, estimates of  $d_S$  can differ considerably among methods when base composition is biased and/or when divergence is large (Dunn *et al.* 2001; Bierne and Eyre-Walker 2003; Aris-Brosou and Bielawski 2006).

Weak selection also predicts inverse relationships between polymorphism levels (scaled to neutral values) and  $N_e$  (see Figures 2 and 3). The DNA analog of heterozygosity in allozyme data is “nucleotide diversity,” the number of differences (per nucleotide site) in a randomly selected pair of chromosomes from a population ( $\pi$ ). Under neutral protein evolution, equilibrium DNA diversity at nonsynonymous sites is

$$\pi_N = 4 N_e f_n u,$$

where  $u$  is the total mutation rate (Kimura 1969; Watterson 1975; Tajima 1983). If all synonymous mutations are neutral,  $\pi_N/\pi_S$  provides an estimate of  $f_n$ , which, under the neutral model, should be generally constant among species. Under near neutrality,  $\pi_N/\pi_S$  gives an estimate of  $f'_n$ , the proportion of effectively neutral mutations that are strongly dependent on  $N_e$  [note that  $f'_n$  estimates from polymorphism include mutations with a larger range of selection intensities than  $f_n$  estimated from divergence (see Figure 1)]. Fluctuations in population size and/or linked selection may contribute to the invariance of allozyme heterozygosity, but such factors do not affect expected  $\pi_N/\pi_S$  for neutral protein polymorphism.

Tests of weak selection that compare polymorphism among populations are robust to some of the assumptions required in divergence tests. Estimates of population size from present-day populations should be more accurate for the relevant  $N_e$ . Since advantageous mutations make small contributions to polymorphism (under directional selection), comparisons of  $\pi_N/\pi_S$  are less affected by adaptive evolution than divergence (Figure 3). In addition, although neutrality at synonymous sites (or other control classes such as introns) may not be strictly valid, weak selection in the control class will have less impact on polymorphism than fixation. Finally,  $\pi_N/\pi_S$  comparisons can be performed among closely related species (or even among populations of the same species if they are evolving independently). The critical assumption of DSE identity is more likely to hold among closely related organisms than among more distantly related species.

Elyashiv *et al.* (2010) found negative relationships between  $\pi_N/\pi_S$  and estimates of population size among closely related yeast species.  $\pi_S$  varies roughly 10-fold among essentially independently evolving populations of *Saccharomyces cerevisiae* and *S. paradoxus*, suggesting simi-

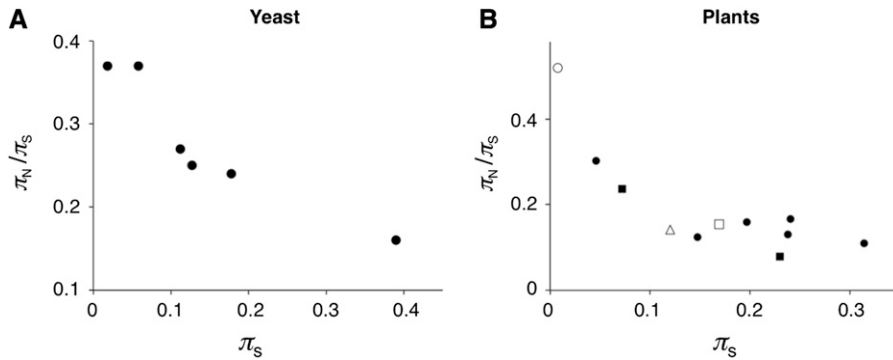
lar variation in  $N_e$ . Although data were available for only six populations,  $\pi_N/\pi_S$  varies considerably among populations (from 0.16 to 0.37) and shows a negative correlation with  $\pi_S$ . Piganeau and Eyre-Walker (2009) showed negative correlations between  $\pi_N/\pi_S$  and  $\pi_S$  for mtDNA genes across more distantly related species within several taxa (they examined >1700 species including fish, birds, mammals, insects, and mollusks).

$\pi_N/\pi_S$  ratios for mammal and plant nuclear genes also show negative associations with estimates of population size. In comparisons limited to a small number of genes, Satta (2001) found higher  $\pi_N/\pi_S$  in humans than in chimps; this difference has been confirmed in genome-scale analyses (Hvilsom *et al.* 2012).  $\pi_N/\pi_S$  is also higher in humans (0.35) than in mice (0.19; Halligan *et al.* 2010) and rabbits (0.05; Carneiro *et al.* 2012), which is consistent with substantial differences in  $f'_n$ . Among a larger number of plant species,  $\pi_N/\pi_S$  varies by roughly fourfold among species and is inversely related to estimates of  $N_e$  (which ranges from  $\sim 3 \times 10^4$  to  $\sim 7 \times 10^5$ ) (Gossmann *et al.* 2010; Slotte *et al.* 2010; Strasburg *et al.* 2011).

Figure 4 shows protein polymorphism comparisons among yeasts and plants. Both yeast and plants show negative relationships between  $\pi_N/\pi_S$  and  $\pi_S$  as predicted under weak selection.  $\pi_N/\pi_S$  is lower among *Drosophila* species than among mammals (both overall and among homologous genes involved in the metabolic process) (Petit and Barbadilla 2009), but patterns within *Drosophila* are unclear (the numbers of sampled populations is low; see supporting information, Table S1). Polymorphism patterns for individual species can be attributed to differences in DSEs rather than population size [e.g., high  $\pi_N/\pi_S$  in humans has been attributed to relaxed selection (Takahata 1993; Satta 2001)]. Sampling species with a breadth of population sizes in a larger number of taxa will be necessary to establish  $N_e$  as a key factor determining the proportions of weakly and strongly deleterious mutations and to test whether near neutrality underlies the invariance of protein heterozygosity.

### **Within-lineage contrasts of polymorphism and divergence**

The examples above employed mostly among-lineage contrasts in the evolutionary dynamics of protein and DNA mutations to test mechanisms of molecular evolution. Within-lineage comparisons offer an alternative approach to test for natural selection acting at its limit of efficacy. We first consider approaches restricted to polymorphism data. As discussed above, frequency spectra for allozyme variation are often skewed toward an excess of rare (low frequency) variants in mammals and *Drosophila* (Latter 1975; Ohta 1975; Chakraborty *et al.* 1980). These studies compared data for a single class of mutations to a theoretical expectation for neutral alleles at steady state (see also Watterson 1977); rejection of the null hypothesis is consistent with either selection or departures from equilibrium. Bulmer



**Figure 4** Levels of nonsynonymous and synonymous DNA polymorphism among populations. DNA diversity for nonsynonymous mutations (scaled to DNA diversity for synonymous mutations) is plotted against DNA diversity for synonymous mutations (an estimate of population size).  $\pi_S$  is a proxy for population size if mutation rates are similar among the species compared. Note that statistical analyses of such data must account for the contribution of  $\pi_S$  to both axes (Piganeau and Eyre-Walker 2009; Elyashiv *et al.* 2010). Common symbols in each plot indicate the same set of genes compared among species. Data are shown for taxa for

which six or more independent populations have been sampled for  $\geq 20$  nuclear genes. See Table S1 for species names, sample numbers, number of loci, and references (as well as data for a more limited number of *Drosophila* species).

(1971) took a different approach in directly comparing observed frequency distributions among classes. At loci segregating for more than two alleles, common allozyme variants consistently showed intermediate electrophoretic mobility whereas rare alleles tended to show extreme (both slow and fast) mobility. The statistical approach compared patterns for within-locus mutation classes and rejected equal fitness effects among mobility classes (the neutral model predicts no association between frequency and mobility) in favor of natural selection. Because electrophoretic mobility is primarily dependent on the overall charge of proteins, Bulmer’s findings are consistent with stronger deleterious effects of radical amino acid mutations relative to conservative changes.

Comparisons among functional categories of mutations that are interspersed in DNA can resolve overlapping predictions for demographic and selection scenarios in tests of the neutral model. Sawyer *et al.* (1987) proposed direct comparisons between the site frequency spectra for nonsynonymous and synonymous DNA polymorphisms (interspecific data were not available). Their study showed an excess of rare amino acid variants among naturally occurring alleles of an *E. coli* gene. Sawyer *et al.* (1987) argued that inference of natural selection from direct comparisons among classes of mutations from the same gene is robust. Because factors that affect variation in a given genetic region (*e.g.*, demographic history or genetic linkage to selected sites) have similar impacts on synonymous and nonsynonymous mutations, differences in their frequency spectra can be attributed to differences in their fitness effects. A similar argument for robustness applies to the Bulmer test discussed above.

McDonald and Kreitman (1991) included divergence data in direct comparisons of mutation classes interspersed in DNA. Their test divides variable sites into two classes (polymorphic mutations are pooled into a single class and compared to fixed differences). The test relies on the sensitivity of the ratio of the numbers of polymorphic-to-fixed differences ( $r_{pd}$ ) to even very weak selection (Figure 1). Slightly deleterious amino acid polymorphisms elevate nonsynonymous  $r_{pd}$ , whereas adaptive protein evolution has the opposite effect.

We will refer to the general category of approaches that compare frequency spectra and fixations between classes of mutations interspersed in DNA as “population genetics of interspersed mutations” (PGIM) tests. Numerous studies have reported PGIM patterns consistent with slightly deleterious amino acid changes. Excess rare polymorphisms and higher  $r_{pd}$  for nonsynonymous compared to synonymous variation appear to be general patterns in viral (Edwards *et al.* 2006; Pybus *et al.* 2006; Hughes 2009) and bacterial genomes (Hughes 2005; Charlesworth and Eyre-Walker 2006; Rocha *et al.* 2006; Hughes *et al.* 2008) as well as in animal mtDNA (Ballard and Kreitman 1994; Nachman *et al.* 1996; Rand and Kann 1996; Hasegawa *et al.* 1998; Wise *et al.* 1998; Weinreich and Rand 2000; Gerber *et al.* 2001; Subramanian *et al.* 2009). Similar patterns have been noted in nuclear genes from yeast (Doniger *et al.* 2008; Liti *et al.* 2009), *Drosophila* (Akashi 1996; Fay *et al.* 2002; Loewe *et al.* 2006; Begun *et al.* 2007; Shapiro *et al.* 2007; Haddrill *et al.* 2010; Andolfatto *et al.* 2011), humans (Cargill *et al.* 1999; Sunyaev *et al.* 2000; Hughes *et al.* 2003; Williamson *et al.* 2005; Boyko *et al.* 2008; Keightley and Halligan 2011; Subramanian 2012), mice (Halligan *et al.* 2010), and plants (Bustamante *et al.* 2002; Nordborg *et al.* 2005; Foxe *et al.* 2008; Fujimoto *et al.* 2008; Gossmann *et al.* 2010; Slotte *et al.* 2010; Branca *et al.* 2011; Strasburg *et al.* 2011). Excess rare polymorphism could reflect very strongly deleterious mutations if the number of sampled chromosomes is very high, but sample sizes in most of these studies do not approach such levels. Interpretations of PGIM patterns often assume random sampling from a panmictic population, but some of the studies above are likely to include alleles from structured populations; in such cases, local adaptation could contribute to excess protein polymorphism. Balancing selection, more generally, is consistent with high nonsynonymous  $r_{pd}$  and, under some conditions, with excess rare variants (Gillespie 1994a). Genome scans have detected little evidence for major contributions of long-term balancing selection in humans (Andrés *et al.* 2009) and *Drosophila* (Wright and Andolfatto 2008), but the genomic signal for other forms of balancing selection may be more subtle (Charlesworth 2006).



The null hypothesis of PGIM tests is identical distributions of selection intensities among classes of DNA mutations. Additional information, such as an assumption of neutral mutations for one of the classes, allows inference of the sign of selection coefficients. The statistical power to detect selection can be improved by including both frequency spectra and divergence data [*i.e.*, by combining the Sawyer *et al.* (1987) and McDonald and Kreitman (1991) approaches] and by inferring frequencies for newly arisen polymorphisms (Akashi 1999a; Bustamante *et al.* 2001). Such information also allows inference of the relative contributions of slightly deleterious, effectively neutral, and advantageous substitutions to genome evolution (Charlesworth 1994; Akashi 1999b; Fay *et al.* 2001; Bustamante *et al.* 2002; Smith and Eyre-Walker 2002; Sawyer *et al.* 2003, 2007). Recent approaches jointly estimate demographic parameters (*e.g.*, population size changes) and distributions of  $N_e$ s (Williamson *et al.* 2005; Keightley and Eyre-Walker 2007; Boyko *et al.* 2008; Eyre-Walker and Keightley 2009; Schneider *et al.* 2011; Wilson *et al.* 2011). Although the number of taxa that have been examined remains limited, current findings are generally consistent with an  $N_e$  effect on the efficacy of negative selection on nonsynonymous polymorphism (discussed above). The proportion of adaptive protein fixations among lineages also shows some indications of a positive association with population size (Gossmann *et al.* 2012). Such associations are expected for both weakly selected beneficial alleles and more strongly selected alleles when rates of adaptive evolution are mutation-limited. Among mammals, PGIM studies support ~60% adaptive protein fixations in mice and rabbits (Halligan *et al.* 2010; Carneiro *et al.* 2012) compared to a point estimate of 0% (and a rough maximum estimate of ~40%) in humans (Eyre-Walker and Keightley 2009). Adaptive proportions range from 0 to 50% among sunflower species and show positive relationships with estimates of  $N_e$  (Strasburg *et al.* 2011). Several studies estimate >50% adaptive evolution in three *Drosophila* species believed to have large population sizes (Smith and Eyre-Walker 2002; Bierne and Eyre-Walker 2004; Welch 2006; Maside and Charlesworth 2007; Haddrill *et al.* 2008; Schneider *et al.* 2011; Wilson *et al.* 2011), but adaptive proportions are also high in a species with a substantially lower estimate of  $N_e$  (Bachtrog 2008).

McDonald and Kreitman (1991) discussed a potential caveat in PGIM inference of adaptive protein evolution. Spurious evidence for positive selection can arise from a combination of weakly deleterious mutations and changes in population size (see also Ohta 1993; Eyre-Walker 2002; Hughes 2008). Current approaches to estimate selection intensity assume that  $N_e$ s has remained constant over the period that sampled polymorphisms and fixed differences have arisen. However, consider a case where fixed differences accumulate in a small population that subsequently increases so that the sampled polymorphisms come from the larger population whereas a large proportion of the fixed differ-

ences occurred in the smaller, ancestral  $N_e$ . In our example (Figure 2), a population size of  $10^2$  has a >2.5-fold higher  $d_N/d_S$  than a population of  $10^4$ . Under such a scenario, an excess of slightly deleterious fixations will reduce nonsynonymous  $r_{pd}$  relative to the constant population size case. This pattern would be interpreted as evidence for adaptive evolution under assumptions of constant DSE and effective population size. Current approaches to jointly estimate demographic parameters and distributions of  $N_e$ s do not attempt to adjust for excess amino acid fixations in cases of smaller ancestral populations.

Slightly deleterious fixations in ancestral populations remain a viable alternative explanation for excess replacement fixations in PGIM tests. McDonald and Kreitman (1991) argued that such scenarios require extensive parameter tuning (particular combinations of DSE, population sizes, and timing of population size change) and favored the simpler explanation of adaptive evolution. Fay *et al.* (2002) claimed that considerable among-gene variation in nonsynonymous  $r_{pd}$  favors adaptive evolution rather than slightly deleterious fixations. This argument assumes similar DSEs among genes, but the basis for this assumption is unclear. DSEs are likely to be determined by structural features of proteins (*e.g.*, ratio of solvent exposed to buried residues) as well as their expression patterns, and both factors may vary widely among genes. Eyre-Walker and Keightley (2009) suggested that the scenario of deleterious fixations in small ancestral populations is unlikely to explain low protein  $r_{pd}$  across multiple lineages. This argument is valid if high rates of adaptive protein evolution are consistently found among species with independent population histories but does not address evidence for high rates of adaptive protein evolution in particular lineages or cases where support for adaptive evolution varies among species (*e.g.*, plants). Incorporating a model of codon bias evolution in tests of positive selection at other site classes may help to distinguish between deleterious and adaptive fixations (Akashi 1999a).

## Compensatory Protein Evolution

Determining the contribution of alleles of small effect in adaptive evolution is a challenge for polymorphism/divergence analyses. The notion that adaptive evolution should “advance by the shortest and slowest steps” (Darwin 1859) remains a central assumption in many theories of adaptation (reviewed in Orr 2005). Further developments in statistical approaches may provide a means to infer DSEs among positively selected mutations, but disentangling the proportion of adaptive fixations and the magnitude of their effects may prove difficult (Boyko *et al.* 2008; Schneider *et al.* 2011).

Weakly beneficial mutations may include compensatory changes that maintain fitness in the face of abundant slightly deleterious fixations. The term “compensatory” has been used in a number of contexts, and we will first attempt to clarify some terminology. Kimura (1985) focused on scenarios of

simultaneous substitutions of relatively strongly deleterious and compensatory mutations. Because fitness remains constant and genetic drift drives double-mutant substitution, we refer to such cases as “compensatory neutral” (CN). Weakly deleterious mutations can go to fixation at appreciable rates, and we refer to the restoration of fitness by subsequent, positively selected mutations as “compensatory weak selection” (CWS) (Osada and Akashi 2012). Note, however, that such mutations can confer large fitness benefits if they compensate for multiple deleterious fixations. Finally, several studies of artificial protein evolution have found that adaptive amino acid fixations are followed by beneficial fixations that restore pleiotropic deleterious effects of the initial substitution (reviewed in Andersson and Hughes 2010). The latter class often includes mutations that elevate protein stability (Wang *et al.* 2002; Ortlund *et al.* 2007; Tokuriki *et al.* 2008). If the initial substitution is positively selected ( $s > 0$  despite pleiotropic effects), we refer to subsequent beneficial fixations as “compensatory modifiers” (CM). The distinction among classes of compensatory evolution is important because, under CWS, weak deleterious fixations create a necessity for positive selection even in the absence of environmental change.

Compensatory substitutions may be prevalent in genome evolution. Co-evolution at nucleotide sites that correspond to paired regions in RNA stem structures supports both CN and CWS (Chen *et al.* 1999; Meer *et al.* 2010). PGIM comparisons between translationally preferred and unpreferred synonymous codons are consistent with CWS in the maintenance of *Drosophila* and yeast codon bias (Akashi 1995; Liti *et al.* 2009), and similar contrasts between putative fitness classes suggest CWS maintenance of nucleosome occupancy patterns in yeast (Kenigsberg *et al.* 2010). Conserved expression patterns despite gain and loss of binding sites in regulatory regions (Ludwig *et al.* 2000; reviewed in Weirauch and Hughes 2010) and turnover of splicing enhancers in mammals (Ke *et al.* 2008) have been attributed to mutation-selection-drift balance (CWS).

Laboratory studies support an abundance of epistatic mutations that can restore loss of fitness in proteins (Burch and Chao 1999; Moore *et al.* 2000; Poon 2005; Poon *et al.* 2005), but identifying compensatory evolution in nature has proven challenging. Charlesworth and Eyre-Walker (2007) tested for CWS by comparing rates of evolution (mostly mtDNA-encoded proteins) among pairs of species that inhabit mainlands and islands. They found higher rates of protein evolution in mainland relative to island populations in cases where the mainland populations are thought to be derived from island populations. Such patterns are consistent with compensatory evolution in larger (mainland) populations following an accumulation of slightly deleterious fixations in smaller (island) populations. However, such patterns could also reflect population size effects on mutation-limited adaptive evolution in novel habitats.

Co-evolutionary patterns can reveal compensatory protein evolution (Haag 2007). Mutations that are strongly del-

eterious in one species can be fixed in a different genetic background (presumably containing compensatory mutations) in other species. Patterns consistent with such a scenario have been documented in mammals (Kondrashov *et al.* 2002) and insects (Kulathinal *et al.* 2004). Several early studies identified within-protein correlations among amino acid positions but did not account for phylogenetic relationships (Korber *et al.* 1993; Neher 1994). More recent studies have identified co-occurring substitutions within a phylogeny (Pollock 1999; Fukami-Kobayashi *et al.* 2002; Dimmic *et al.* 2005; Duthel *et al.* 2005; Yeang and Haussler 2007). Such patterns support within-protein epistatic interactions but do not distinguish among CN, CM, and CWS scenarios. Mapping the temporal order of substitutions (given dense sampling of species) can identify sequentially occurring substitutions consistent with CM or CWS (Bazykin *et al.* 2006; Kryazhimskiy *et al.* 2011; Osada and Akashi 2012), but distinguishing between these scenarios may be limited to particular biological contexts.

The difficulty of predicting fitness effects of amino acid changes has been a major limitation to testing compensatory evolution. The notion that protein evolution may, in large part, reflect mutation-selection-drift balance among slightly deleterious destabilizing and weakly beneficial, stabilizing amino acid changes (Tokuriki and Tawfik 2009b; Goldstein 2011; Wylie and Shakhnovich 2011) may allow tests of CWS if mutations falling into these fitness classes can be predicted.

### Concave Fitness Functions and the Paradox of Near Neutrality

Most of this review has focused on patterns of protein and DNA variation that test weak selection. However, one of the strongest objections to the nearly neutral model has been a theoretical one—its reliance on a particular distribution of selective effects. We show an example of a DSE (Figure 2A) that gives nearly neutral dynamics over a wide range of population sizes (Figure 2D). However, such dynamics hold only for a specific range of shape parameters; small changes in the DSE can cause a lack of dependence of polymorphism and divergence on population size and/or a complete lack of evolution in large populations (Nei and Graur 1984; Gillespie 1987; Takahata 1993). Thus, support for weak selection among classes of mutations and across taxa is paradoxical. Population sizes and ecological circumstances vary widely among species, and relationships between DNA mutations and fitness are determined by different factors for different types of mutations. Why should DSEs in nature so often show the precise characteristics required for nearly neutral evolution?

Linked selection may constrain effective population sizes in natural populations. Larger populations experience a higher input of non-neutral mutations (reduced  $f'_n$  and elevated within-population mutation rate to advantageous alleles). Depending on recombination rates and DSEs, levels

of neutral polymorphism at sites linked to selected sites may show little relationship with population size (Gillespie 2000). Although the number of species for which both direct estimates of mutation rates and measures of DNA diversity in natural populations are available is small, estimates of  $N_e$  range over three orders of magnitude (Charlesworth 2009). Linked selection is likely to constrain among-lineage variation in  $N_e$  (especially in regions of limited recombination) but does not appear to be sufficient to account for the paradox of near neutrality.

General properties of phenotype–fitness relationships may underlie weak selection. Wright (1929, 1934) proposed that diminishing returns in fitness as phenotypes approach optima explain the prevalence of genetic dominance among wild-type alleles and recessive effects of newly arising mutations. His model examined relationships between the catalytic activity of an enzyme (a phenotypic value) and flux through a biochemical pathway (directly related to fitness) and showed that the fitness benefit for a given increase in activity is greater for low-activity alleles than for alleles functioning close to an optimum. Kacser and Burns (1981) confirmed hyperbolic relationships between activity and flux under enzyme kinetic models and gave supporting evidence from biochemical studies. The Kacser–Burns model applies only to particular scenarios of enzyme catalysis (Savageau 1992; Wilkie 1994; Omholt *et al.* 2000), but concave fitness functions may hold more generally. Classic studies by Crow and co-workers showed an “inverse heterozygous-homozygous effect” for deleterious mutations in *Drosophila*; deleterious mutations of large effect tend to be recessive (large  $s$ , small  $h$ ) whereas alleles of small effect tend to show heterozygous effects (small  $s$ , large  $h$ ) (Greenberg and Crow 1960; Simmons and Crow 1977). Such patterns are consistent with diminishing returns because fitness vs. phenotype relationships are close to linear (additive) within restricted ranges of phenotypic values but show strong nonlinearity over wider ranges. Gene-deletion studies in yeast support inverse relationships between fitness (growth rate) and heterozygous effects across functional categories of genes (Phadnis 2005; Agrawal and Whitlock 2011). Gene knockouts were tested (missense mutations may not show the same associations), but the findings are consistent with concave fitness functions for a variety of traits (including noncatalytic functions). Non-additive relationships between protein stability and fitness (Bershtein *et al.* 2006; Tokuriki and Tawfik 2009b; Wylie and Shakhnovich 2011) could underlie diminishing returns epistasis.

Hartl *et al.* (1985) related concave fitness functions to mechanisms of molecular evolution. Under diminishing returns, DSEs change as character values approach optima. Changes in activity will have smaller fitness effects (both positive and negative) as enzymes approach optimum activity, and Hartl and co-workers predicted that characters in natural populations should be on plateaus of concave fitness functions. Gillespie (1994b) found that concave fitness functions can lead to phenotypic values close to their optima and to dynamics indistinguishable from neutral evolution when

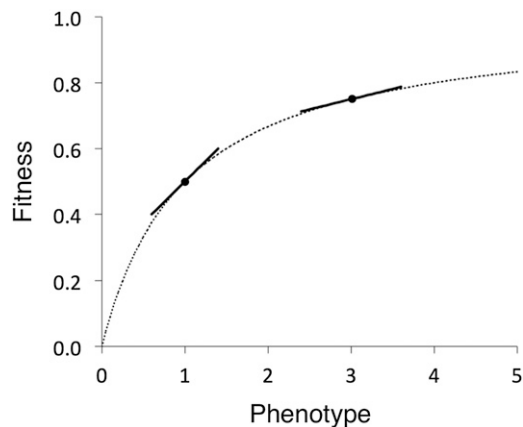
phenotypes of new mutations are normally distributed around parental values (*i.e.*, equal proportions of positively and negatively selected mutations). However, if mutations often have small phenotypic effects and tend to be deleterious, then characters will evolve to equilibrium values below their optima (Akashi 1996; Hartl and Taubes 1998). Smaller populations will evolve to lower equilibrium values than larger populations, and a wide range of population sizes could show a balance among weak selection, drift, and mutation; *i.e.*, DSEs will evolve in different population sizes to a point where many mutations show  $|N_e s| \approx 1$  (Figure 5). Such a scenario has been proposed to account for patterns of codon usage bias (Li 1987; Akashi 1995; Kondrashov *et al.* 2006). The correlation between evolutionary rate and population size can be weak under concave fitness relationships but will depend on how DSEs change with phenotypic values (Cherry 1998). Concave fitness functions are an appealingly simple explanation for widespread weak selection, but evidence for this form of epistasis remains mixed. Some mutation accumulation studies support increasing negative effects with declining fitness whereas others support the opposite pattern (reviewed in de Visser *et al.* 2011).

## Discussion

Although support for weak selection in protein evolution has accumulated in diverse taxa, adaptive protein evolution and changes in the distribution of fitness effects can often make predictions that overlap those of near neutrality. Within-population comparisons (excess rare amino acid variants) provide some of the strongest evidence for weak selection because positive, directional selection should have little impact on polymorphism data. Associations between levels of protein polymorphism and estimates of  $N_e$  among closely related populations also support weak selection because DSEs are likely to be conserved on relatively short time scales.

Distinguishing adaptive, neutral, and nearly neutral modes of molecular evolution remains challenging. All three mechanisms enjoy sufficient support to be considered among potential explanatory factors in almost all studies that infer mechanisms of genome evolution. For polymorphism analyses, near neutrality can cause differences in levels of variation and in frequency spectra among functional categories of mutations or among populations. In among-species comparisons, weak selection may underlie considerable variation in rates and patterns of divergence. Slightly deleterious amino acid mutations can reduce the power to detect adaptive protein evolution but can also generate a false signal of adaptive fixations under particular scenarios of population size change. Weak selection in the comparison class in tests of adaptive protein evolution (often synonymous DNA changes) can also contribute to false signals of positive selection (Akashi 1999a).

Compensatory evolution, and small fitness effects in adaptive evolution more generally, may be important factors in protein evolution but have been difficult to test. Evidence



**Figure 5** Concave fitness functions and near neutrality. The curve  $y = x/(1 + x)$  shows a hypothetical relationship between fitness and phenotypic values of a trait. The slope of the curve at a given point determines the fitness effect of small phenotypic changes (slopes are shown for phenotypic values of 1 and 3). The slope decreases as a function of the phenotypic value (*i.e.*, the distribution of  $s$  changes with character values). If a large fraction of mutations have small phenotypic effects and if the rate of mutation to deleterious alleles is higher than the rate to advantageous mutations, populations will evolve to a point on the curve where slightly deleterious mutations that move the population away from the optimum will be balanced by weak positive selection. The left and right points marked in the figure correspond to equilibrium points in species with small and large population sizes, respectively (this assumes constant mutation rates and population sizes).

for such factors is stronger at synonymous sites, in RNA genes, and in noncoding DNA than in protein evolution in part because functional and fitness effects of mutations can be predicted (*e.g.*, translationally preferred vs. unpreferred synonymous changes, destabilizing and stabilizing changes in RNA stem structures). Further incorporation of biophysical principles may be essential for advancing our understanding of protein evolution. Models of protein evolution that account for both stability and activity effects have the potential to predict fitness effects of amino acid changes (Depristo *et al.* 2005; Bloom *et al.* 2007b; Tokuriki and Tawfik 2009b; Wilke and Drummond 2010; Goldstein 2011; Wylie and Shakhnovich 2011) and may help to determine the proportion of compensatory substitutions among adaptive fixations. This information, in combination with experimental approaches (*e.g.*, Bershtein *et al.* 2006; Bloom *et al.* 2007a; Araya and Fowler 2011), will allow tests of Ohta's (1973) prediction of fundamental relationships among protein primary structure, folding, and weak selection.

## Acknowledgments

We thank Wen-Ya Ko, Boyang Li, Neha Mishra, Michael Turelli, and two anonymous reviewers for many helpful comments on the manuscript. We also thank Richard Kliman for providing *Drosophila* polymorphism data and Noriko Yamauchi for technical support.

## Literature Cited

- Abbot, P., and N. A. Moran, 2002 Extremely low levels of genetic polymorphism in endosymbionts (*Buchnera*) of aphids (*Pemphigus*). *Mol. Ecol.* 11: 2649–2660.
- Agrawal, A. F., and M. C. Whitlock, 2011 Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* 187: 553–566.
- 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., 1999a Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. *Genetics* 151: 221–238.
- Akashi, H., 1999b Within- and between-species DNA sequence variation and the “footprint” of natural selection. *Gene* 238: 39–51.
- Akashi, H., 2001 Gene expression and molecular evolution. *Curr. Opin. Genet. Dev.* 11: 660–666.
- Andersson, D. I., and D. Hughes, 2010 Antibiotic resistance and its cost: Is it possible to reverse resistance? *Nat. Rev. Microbiol.* 8: 260–271.
- Andersson, J. O., and S. G. Andersson, 1999 Insights into the evolutionary process of genome degradation. *Curr. Opin. Genet. Dev.* 9: 664–671.
- Andolfatto, P., K. M. Wong, and D. Bachtrog, 2011 Effective population size and the efficacy of selection on the X chromosomes of two closely related *Drosophila* species. *Genome Biol. Evol.* 3: 114–128.
- Andrés, A. M., M. J. Hubisz, A. Indap, D. G. Torgerson, J. D. Degenhardt *et al.*, 2009 Targets of balancing selection in the human genome. *Mol. Biol. Evol.* 26: 2755–2764.
- Araya, C. L., and D. M. Fowler, 2011 Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.* 29: 435–442.
- Aris-Brosou, S., and J. P. Bielawski, 2006 Large-scale analyses of synonymous substitution rates can be sensitive to assumptions about the process of mutation. *Gene* 378: 58–64.
- Bachtrog, D., 2008 Similar rates of protein adaptation in *Drosophila miranda* and *D. melanogaster*, two species with different current effective population sizes. *BMC Evol. Biol.* 8: 334.
- Ballard, J. W., and M. Kreitman, 1994 Unraveling selection in the mitochondrial genome of *Drosophila*. *Genetics* 138: 757–772.
- Barnard, E. A., M. S. Cohen, M. H. Gold, and J. K. Kim, 1972 Evolution of ribonuclease in relation to polypeptide folding mechanisms. *Nature* 240: 395–398.
- Barton, N. H., 1995 Linkage and the limits to natural selection. *Genetics* 140: 821–841.
- Bazin, E., S. Glémin, and N. Galtier, 2006 Population size does not influence mitochondrial genetic diversity in animals. *Science* 312: 570–572.
- Bazykin, G. A., J. Dushoff, S. A. Levin, and A. S. Kondrashov, 2006 Bursts of nonsynonymous substitutions in HIV-1 evolution reveal instances of positive selection at conservative protein sites. *Proc. Natl. Acad. Sci. USA* 103: 19396–19401.
- Begun, D. J., A. K. Holloway, K. Stevens, L. W. Hillier, Y.-P. Poh *et al.*, 2007 Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* 5: e310.
- Bershtein, S., M. Segal, R. Bekerman, N. Tokuriki, and D. S. Tawfik, 2006 Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444: 929–932.

- Betancourt, A. J., and D. C. Presgraves, 2002 Linkage limits the power of natural selection in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 99: 13616–13620.
- Bierne, N., and A. Eyre-Walker, 2003 The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: implications for the correlation between the synonymous substitution rate and codon usage bias. *Genetics* 165: 1587–1597.
- Bierne, N., and A. Eyre-Walker, 2004 The genomic rate of adaptive amino acid substitution in *Drosophila*. *Mol. Biol. Evol.* 21: 1350–1360.
- Birdsell, J. A., 2002 Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Mol. Biol. Evol.* 19: 1181–1197.
- Birky, C. W., and J. B. Walsh, 1988 Effects of linkage on rates of molecular evolution. *Proc. Natl. Acad. Sci. USA* 85: 6414–6418.
- Bloom, J. D., Z. Lu, D. Chen, A. Raval, O. S. Venturelli *et al.*, 2007a Evolution favors protein mutational robustness in sufficiently large populations. *BMC Biol.* 5: 29.
- Bloom, J. D., A. Raval, and C. O. Wilke, 2007b Thermodynamics of neutral protein evolution. *Genetics* 175: 255–266.
- Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 4: e1000083.
- Branca, A., T. D. Paape, P. Zhou, R. Briskine, A. D. Farmer *et al.*, 2011 Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proc. Natl. Acad. Sci. USA* 108: E864–E870.
- Bullaughay, K., M. Przeworski, and G. Coop, 2008 No effect of recombination on the efficacy of natural selection in primates. *Genome Res.* 18: 544–554.
- Bulmer, M. G., 1971 Protein polymorphism. *Nature* 234: 410–411.
- Burch, C. L., and L. Chao, 1999 Evolution by small steps and rugged landscapes in the RNA virus phi6. *Genetics* 151: 921–927.
- Bustamante, C. D., J. Wakeley, S. Sawyer, and D. L. Hartl, 2001 Directional selection and the site-frequency spectrum. *Genetics* 159: 1779–1788.
- Bustamante, C. D., R. Nielsen, S. A. Sawyer, K. M. Olsen, M. D. Purugganan *et al.*, 2002 The cost of inbreeding in *Arabidopsis*. *Nature* 416: 531–534.
- Campos, J. L., B. Charlesworth, and P. R. Haddrill, 2012 Molecular evolution in non-recombining regions of the *Drosophila melanogaster* genome. *Genome Biol. Evol.* 4: 278–288.
- Cargill, M., D. Altshuler, J. Ireland, P. Sklar, K. Ardlie *et al.*, 1999 Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22: 231–238.
- Carneiro, M., F. W. Albert, J. Melo-Ferreira, N. Galtier, P. Gayral *et al.*, 2012 Evidence for widespread positive and purifying selection across the European rabbit (*Oryctolagus cuniculus*) genome. *Mol. Biol. Evol.* 29: 1837–1849.
- Chakraborty, R., P. A. Fuerst, and M. Nei, 1980 Statistical studies on protein polymorphism in natural populations. III. Distribution of allele frequencies and the number of alleles per locus. *Genetics* 94: 1039–1063.
- Chamary, J. V., J. L. Parnley, and L. D. Hurst, 2006 Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* 7: 98–108.
- Chao, L., and D. E. Carr, 1993 The molecular clock and the relationship between population size and generation time. *Evolution* 47: 688–690.
- Charlesworth, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* 63: 213–227.
- Charlesworth, B., 2009 Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10: 195–205.
- Charlesworth, D., 2006 Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* 2: e64.
- Charlesworth, D., B. Charlesworth, and M. T. Morgan, 1995 The pattern of neutral molecular variation under the background selection model. *Genetics* 141: 1619–1632.
- Charlesworth, J., and A. Eyre-Walker, 2006 The rate of adaptive evolution in enteric bacteria. *Mol. Biol. Evol.* 23: 1348–1356.
- Charlesworth, J., and A. Eyre-Walker, 2007 The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. *Proc. Natl. Acad. Sci. USA* 104: 16992–16997.
- Chen, Y., D. B. Carlini, J. F. Baines, J. Parsch, J. M. Braverman *et al.*, 1999 RNA secondary structure and compensatory evolution. *Genes Genet. Syst.* 74: 271–286.
- Cherry, J. L., 1998 Should we expect substitution rate to depend on population size? *Genetics* 150: 911–919.
- Chimpanzee Sequencing and Analysis Consortium, 2005 Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437: 69–87.
- Connallon, T., and L. L. Knowles, 2007 Recombination rate and protein evolution in yeast. *BMC Evol. Biol.* 7: 235.
- Damuth, J., 1981 Population density and body size in mammals. *Nature* 290: 699–700.
- Darwin, C., 1859 *The Origin of Species*. J. Murray, London.
- DePristo, M. A., D. M. Weinreich, and D. L. Hartl, 2005 Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* 6: 678–687.
- de Visser, J. A., T. F. Cooper, and S. F. Elena, 2011 The causes of epistasis. *Proc. Biol. Sci.* 278: 3617–3624.
- Dimmic, M. W., M. J. Hubisz, C. D. Bustamante, and R. Nielsen, 2005 Detecting coevolving amino acid sites using Bayesian mutational mapping. *Bioinformatics* 21(Suppl 1): i126–i135.
- Doniger, S. W., H. S. Kim, D. Swain, D. Corcuera, M. Williams *et al.*, 2008 A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* 4: e1000183.
- Dunn, K. A., J. P. Bielawski, and Z. Yang, 2001 Substitution rates in *Drosophila* nuclear genes: implications for translational selection. *Genetics* 157: 295–305.
- Duret, L., and N. Galtier, 2009 Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* 10: 285–311.
- Dutheil, J., T. Pupko, A. Jean-Marie, and N. Galtier, 2005 A model-based approach for detecting coevolving positions in a molecule. *Mol. Biol. Evol.* 22: 1919–1928.
- Dykhuizen, D. E., and D. L. Hartl, 1983 Selection in chemostats. *Microbiol. Rev.* 47: 150–168.
- Edwards, C. T. T., E. C. Holmes, O. G. Pybus, D. J. Wilson, R. P. Viscidi *et al.*, 2006 Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection. *Genetics* 174: 1441–1453.
- Ellegren, H., 2009 A selection model of molecular evolution incorporating the effective population size. *Evolution* 63: 301–305.
- Elyashiv, E., K. Bullaughey, S. Sattath, Y. Rinott, M. Przeworski *et al.*, 2010 Shifts in the intensity of purifying selection: an analysis of genome-wide polymorphism data from two closely related yeast species. *Genome Res.* 20: 1558–1573.
- Eyre-Walker, A., 2002 Changing effective population size and the McDonald-Kreitman test. *Genetics* 162: 2017–2024.
- Eyre-Walker, A., and P. D. Keightley, 2009 Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* 26: 2097–2108.
- Eyre-Walker, A., P. D. Keightley, N. G. C. Smith, and D. Gaffney, 2002 Quantifying the slightly deleterious mutation model of molecular evolution. *Mol. Biol. Evol.* 19: 2142–2149.

- Fares, M. A., M. X. Ruiz-González, A. Moya, S. F. Elena, and E. Barrio, 2002 Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature* 417: 398.
- Fares, M. A., A. Moya, and E. Barrio, 2004 GroEL and the maintenance of bacterial endosymbiosis. *Trends Genet.* 20: 413–416.
- Fay, J. C., 2011 Weighing the evidence for adaptation at the molecular level. *Trends Genet.* 27: 343–349.
- Fay, J. C., G. J. Wyckoff, and C. I. Wu, 2001 Positive and negative selection on the human genome. *Genetics* 158: 1227–1234.
- Fay, J. C., G. J. Wyckoff, and C.-I. Wu, 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.
- Foxe, J. P., V.-U.-N. Dar, H. Zheng, M. Nordborg, B. S. Gaut *et al.*, 2008 Selection on amino acid substitutions in Arabidopsis. *Mol. Biol. Evol.* 25: 1375–1383.
- Fujimoto, A., T. Kado, H. Yoshimaru, Y. Tsumura, and H. Tachida, 2008 Adaptive and slightly deleterious evolution in a conifer, *Cryptomeria japonica*. *J. Mol. Evol.* 67: 201–210.
- Fukami-Kobayashi, K., D. R. Schreiber, and S. A. Benner, 2002 Detecting compensatory covariation signals in protein evolution using reconstructed ancestral sequences. *J. Mol. Biol.* 319: 729–743.
- Funk, D. J., J. J. Wernegreen, and N. A. Moran, 2001 Intraspecific variation in symbiont genomes: bottlenecks and the aphid-buchnera association. *Genetics* 157: 477–489.
- Gerber, A. S., R. Loggins, S. Kumar, and T. E. Dowling, 2001 Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? *Annu. Rev. Genet.* 35: 539–566.
- Gillespie J. H., 1987 Molecular evolution and the neutral allele theory. *Oxf. Surv. Evol. Biol.*, 4: 10–37.
- Gillespie, J., 1994a Alternatives to the neutral theory, pp. 1–17 in *Non-neutral Evolution: Theories and Molecular Data*, edited by B. Golding. Chapman and Hall, New York.
- Gillespie, J. H., 1994b Substitution processes in molecular evolution. III. Deleterious alleles. *Genetics* 138: 943–952.
- Gillespie, J. H., 2000 Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* 155: 909–919.
- Gillespie, J. H., 2001 Is the population size of a species relevant to its evolution? *Evolution* 55: 2161–2169.
- Glémin, S., and N. Galtier, 2012 Genome evolution in outcrossing versus selfing versus asexual species. *Methods Mol. Biol.* 855: 311–335.
- Goldstein, R. A., 2011 The evolution and evolutionary consequences of marginal thermostability in proteins. *Proteins* 79: 1396–1407.
- Gossmann, T. I., B. H. Song, A. J. Windsor, T. Mitchell-Olds, C. J. Dixon *et al.*, 2010 Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol. Biol. Evol.* 27: 1822–1832.
- Gossmann, T. I., P. D. Keightley, and A. Eyre-Walker, 2012 The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol. Evol.* 4: 658–667.
- Greenberg, R., and J. F. Crow, 1960 A comparison of the effect of lethal and detrimental chromosomes from *Drosophila* populations. *Genetics* 45: 1153–1168.
- Grunstein, M., P. Schedl, and L. Kedes, 1976 Sequence analysis and evolution of sea urchin (*Lytechinus pictus* and *Strongylocentrotus purpuratus*) histone H4 messenger RNAs. *J. Mol. Biol.* 104: 351–369.
- Haag, E. S., 2007 Compensatory vs. pseudocompensatory evolution in molecular and developmental interactions. *Genetica* 129: 45–55.
- Haddrill, P. R., D. Bachtrog, and P. Andolfatto, 2008 Positive and negative selection on noncoding DNA in *Drosophila simulans*. *Mol. Biol. Evol.* 25: 1825–1834.
- Haddrill, P. R., L. Loewe, and B. Charlesworth, 2010 Estimating the parameters of selection on nonsynonymous mutations in *Drosophila pseudoobscura* and *D. miranda*. *Genetics* 185: 1381–1396.
- Halligan, D. L., and P. D. Keightley, 2009 Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* 40: 151–172.
- Halligan, D. L., F. Oliver, A. Eyre-Walker, B. Harr, and P. D. Keightley, 2010 Evidence for pervasive adaptive protein evolution in wild mice. *PLoS Genet.* 6: e1000825.
- Hartl, D. L., and C. H. Taubes, 1998 Towards a theory of evolutionary adaptation. *Genetica* 102–103: 525–533.
- Hartl, D. L., D. E. Dykhuizen, and A. M. Dean, 1985 Limits of adaptation: the evolution of selective neutrality. *Genetics* 111: 655–674.
- Hasegawa, M., Y. Cao, and Z. Yang, 1998 Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. *Mol. Biol. Evol.* 15: 1499–1505.
- Hellmann, I., K. Prüfer, H. Ji, M. C. Zody, S. Pääbo *et al.*, 2005 Why do human diversity levels vary at a megabase scale? *Genome Res.* 15: 1222–1231.
- Herbeck, J. T., D. J. Funk, P. H. Degnan, and J. J. Wernegreen, 2003 A conservative test of genetic drift in the endosymbiotic bacterium *Buchnera*: slightly deleterious mutations in the chaperonin groEL. *Genetics* 165: 1651–1660.
- Hershberg, R., and D. A. Petrov, 2008 Selection on codon bias. *Annu. Rev. Genet.* 42: 287–299.
- Hietpas, R. T., J. D. Jensen, and D. N. A. Bolon, 2011 Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. USA* 108: 7896–7901.
- Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. *Genet. Res.* 8: 269–294.
- Hughes, A. L., 2005 Evidence for abundant slightly deleterious polymorphisms in bacterial populations. *Genetics* 169: 533–538.
- Hughes, A. L., 2008 Near neutrality: leading edge of the neutral theory of molecular evolution. *Ann. N. Y. Acad. Sci.* 1133: 162–179.
- Hughes, A. L., 2009 Small effective population sizes and rare nonsynonymous variants in potyviruses. *Virology* 393: 127–134.
- Hughes, A. L., and R. Friedman, 2009 More radical amino acid replacements in primates than in rodents: support for the evolutionary role of effective population size. *Gene* 440: 50–56.
- Hughes, A. L., B. Packer, R. Welch, A. W. Bergen, S. J. Chanock *et al.*, 2003 Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proc. Natl. Acad. Sci. USA* 100: 15754–15757.
- Hughes, A. L., R. Friedman, P. Rivaille, and J. O. French, 2008 Synonymous and nonsynonymous polymorphisms vs. divergences in bacterial genomes. *Mol. Biol. Evol.* 25: 2199–2209.
- Hvilsom, C., Y. Qian, T. Bataillon, Y. Li, T. Mailund *et al.*, 2012 Extensive X-linked adaptive evolution in central chimpanzees. *Proc. Natl. Acad. Sci. USA* 109: 2054–2059.
- Johnson, K. P., and J. Seger, 2001 Elevated rates of nonsynonymous substitution in island birds. *Mol. Biol. Evol.* 18: 874–881.
- Johnson, S. G., and R. S. Howard, 2007 Contrasting patterns of synonymous and nonsynonymous sequence evolution in asexual and sexual freshwater snail lineages. *Evolution* 61: 2728–2735.
- Kacser, H., and J. A. Burns, 1981 The molecular basis of dominance. *Genetics* 97: 639–666.
- Kafatos, F. C., A. Efstratiadis, B. G. Forget, and S. M. Weissman, 1977 Molecular evolution of human and rabbit beta-globin mRNAs. *Proc. Natl. Acad. Sci. USA* 74: 5618–5622.
- Ke, S., X. H. F. Zhang, and L. A. Chasin, 2008 Positive selection acting on splicing motifs reflects compensatory evolution. *Genome Res.* 18: 533–543.

- Keightley, P. D., and A. Eyre-Walker, 2007 Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177: 2251–2261.
- Keightley, P. D., and D. L. Halligan, 2011 Inference of site frequency spectra from high-throughput sequence data: quantification of selection on nonsynonymous and synonymous sites in humans. *Genetics* 188: 931–940.
- Kenigsberg, E., A. Bar, E. Segal, and A. Tanay, 2010 Widespread compensatory evolution conserves DNA-encoded nucleosome organization in yeast. *PLoS Comput. Biol.* 6: e1001039.
- Kimura, M., 1962 On the probability of fixation of mutant genes in a population. *Genetics* 47: 713–719.
- Kimura, M., 1968 Evolutionary rate at the molecular level. *Nature* 217: 624–626.
- Kimura, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61: 893–903.
- Kimura, M., 1985 *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge; London; New York.
- Kimura, M., 1985 The role of compensatory neutral mutations in molecular evolution. *J. Genet.* 64: 7–19.
- Kimura, M., and J. F. Crow, 1964 The number of alleles that can be maintained in a finite population. *Genetics* 49: 725–738.
- Kimura, M., and T. Ohta, 1971a Protein polymorphism as a phase of molecular evolution. *Nature* 229: 467–469.
- Kimura, M., and T. Ohta, 1971b On the rate of molecular evolution. *J. Mol. Evol.* 1: 1–17.
- Kimura, M., and N. Takahata, 1995 *Population Genetics, Molecular Evolution, and the Neutral Theory*. University of Chicago Press, Chicago.
- King, J. L., and T. H. Jukes, 1969 Non-Darwinian evolution. *Science* 164: 788–798.
- Kohne, D. E., 1970 Evolution of higher-organism DNA. *Q. Rev. Biophys.* 3: 327–375.
- Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov, 2002 Dobzhansky-Muller incompatibilities in protein evolution. *Proc. Natl. Acad. Sci. USA* 99: 14878–14883.
- Kondrashov, F. A., A. Y. Ogurtsov, and A. S. Kondrashov, 2006 Selection in favor of nucleotides G and C diversifies evolution rates and levels of polymorphism at mammalian synonymous sites. *J. Theor. Biol.* 240: 616–626.
- Korber, B. T., R. M. Farber, D. H. Wolpert, and A. S. Lapedes, 1993 Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proc. Natl. Acad. Sci. USA* 90: 7176–7180.
- Kosiol, C., T. Vinar, R. R. da Fonseca, M. J. Hubisz, C. D. Bustamante *et al.*, 2008 Patterns of positive selection in six mammalian genomes. *PLoS Genet.* 4: e1000144.
- Kryazhimskiy, S., J. Dushoff, G. A. Bazykin, and J. B. Plotkin, 2011 Prevalence of epistasis in the evolution of influenza A surface proteins. *PLoS Genet.* 7: e1001301.
- Kulathinal, R. J., B. R. Bettencourt, and D. L. Hartl, 2004 Compensated deleterious mutations in insect genomes. *Science* 306: 1553–1554.
- Laird, C. D., and B. J. McCarthy, 1968 Magnitude of interspecific nucleotide sequence variability in *Drosophila*. *Genetics* 60: 303–322.
- Larracuent, A., T. Sackton, A. Greenberg, A. Wong, N. Singh *et al.*, 2008 Evolution of protein-coding genes in *Drosophila*. *Trends Genet.* 24: 114–123.
- Latter, B., 1975 Enzyme polymorphisms: gene frequency distributions with mutation and selection for optimal activity. *Genetics* 79: 325.
- Lewontin, R., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.
- Lewontin, R. C., 1991 Twenty-five years ago in Genetics: electrophoresis in the development of evolutionary genetics: Milestone or millstone? *Genetics* 128: 657–662.
- Li, W. H., 1987 Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. *J. Mol. Evol.* 24: 337–345.
- Li, W. H., M. Tanimura, and P. M. Sharp, 1987 An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* 25: 330–342.
- Lind, P. A., O. G. Berg, and D. I. Andersson, 2010 Mutational robustness of ribosomal protein genes. *Science* 330: 825–827.
- Lindblad-Toh, K., C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe *et al.*, 2005 Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438: 803–819.
- Liti, G., D. M. Carter, A. M. Moses, J. Warringer, L. Parts *et al.*, 2009 Population genomics of domestic and wild yeasts. *Nature* 458: 337–341.
- Loewe, L., B. Charlesworth, C. Bartolomé, and V. Nöel, 2006 Estimating selection on nonsynonymous mutations. *Genetics* 172: 1079–1092.
- Ludwig, M. Z., C. Bergman, N. H. Patel, and M. Kreitman, 2000 Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403: 564–567.
- Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482: 173–178.
- Marais, G., 2003 Biased gene conversion: implications for genome and sex evolution. *Trends Genet.* 19: 330–338.
- Maside, X., and B. Charlesworth, 2007 Patterns of molecular variation and evolution in *Drosophila americana* and its relatives. *Genetics* 176: 2293–2305.
- Maynard Smith, J., 1970a The causes of polymorphism. *Symp. Zool. Soc. Lond.* 26: 371–383.
- Maynard Smith, J., 1970b Population size, polymorphism, and the rate of non-Darwinian evolution. *Am. Nat.* 104: 231–237.
- Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. *Genet. Res.* 89: 391–403.
- McCutcheon, J. P., and N. A. Moran, 2011 Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10: 13–26.
- McDonald, J., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351: 652–654.
- Meer, M. V., A. S. Kondrashov, Y. Artzy-Randrup, and F. A. Kondrashov, 2010 Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature* 464: 279–282.
- Moore, F. B. G., D. E. Rozen, and R. E. Lenski, 2000 Pervasive compensatory adaptation in *Escherichia coli*. *Proc. Biol. Sci.* 267: 515–522.
- Moran, N., 1996 Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. USA* 93: 2873–2878.
- Mulligan, C. J., A. Kitchen, and M. M. Miyamoto, 2006 Comment on “Population size does not influence mitochondrial genetic diversity in animals.” *Science* 314: 1390.
- Nabholz, B., J. F. Mauffrey, E. Bazin, N. Galtier, and S. Glémin, 2008 Determination of mitochondrial genetic diversity in mammals. *Genetics* 178: 351–361.
- Nachman, M. W., W. M. Brown, M. Stoneking, and C. F. Aquadro, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* 142: 953–963.
- Neher, E., 1994 How frequent are correlated changes in families of protein sequences? *Proc. Natl. Acad. Sci. USA* 91: 98–102.
- Nei, M., and D. Graur, 1984 Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* 17: 73–118.
- Neiman, M., G. Hehman, J. T. Miller, J. M. Logsdon, and D. R. Taylor, 2010 Accelerated mutation accumulation in asexual lineages of a freshwater snail. *Mol. Biol. Evol.* 27: 954–963.
- Nordborg, M., T. T. Hu, Y. Ishino, J. Jhaveri, C. Toomajian *et al.*, 2005 The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* 3: e196.
- Ohta, T., 1972a Evolutionary rate of cistrons and DNA divergence. *J. Mol. Evol.* 1: 150–157.

- Ohta, T., 1972b Population size and rate of evolution. *J. Mol. Evol.* 1: 305–314.
- Ohta, T., 1973 Slightly deleterious mutant substitutions in evolution. *Nature* 246: 96–98.
- Ohta, T., 1974 Mutational pressure as the main cause of molecular evolution and polymorphism. *Nature* 252: 351–354.
- Ohta, T., 1975 Statistical analyses of *Drosophila* and human protein polymorphisms. *Proc. Natl. Acad. Sci. USA* 72: 3194–3196.
- Ohta, T., 1993 Amino acid substitution at the *Adh* locus of *Drosophila* is facilitated by small population size. *Proc. Natl. Acad. Sci. USA* 90: 4548–4551.
- Ohta, T., 1995 Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* 40: 56–63.
- Ohta, T., and M. Kimura, 1971 On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* 1: 18–25.
- Ohta, T., and M. Kimura, 1975 Theoretical analysis of electrophoretically detectable polymorphisms: models of very slightly deleterious mutations. *Am. Nat.* 109: 137–145.
- Ohta, T., and H. Tachida, 1990 Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics* 126: 219–229.
- Omholt, S. W., E. Plahte, L. Oyehaug, and K. Xiang, 2000 Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. *Genetics* 155: 969–980.
- Orr, H. A., 2005 The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* 6: 119–127.
- Ortlund, E. A., J. T. Bridgham, M. R. Redinbo, and J. W. Thornton, 2007 Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317: 1544–1548.
- Osada, N., and H. Akashi, 2012 Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. *Mol. Biol. Evol.* 29: 337–346.
- Paland, S., and M. Lynch, 2006 Transitions to asexuality result in excess amino acid substitutions. *Science* 311: 990–992.
- Petit, N., and A. Barbadilla, 2009 The efficiency of purifying selection in mammals vs. *Drosophila* for metabolic genes. *J. Evol. Biol.* 22: 2118–2124.
- Phadnis, N., 2005 Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. *Genetics* 171: 385–392.
- Piganeau, G., and A. Eyre-Walker, 2009 Evidence for variation in the effective population size of animal mitochondrial DNA. *PLoS ONE* 4: e4396.
- Plotkin, J. B., and G. Kudla, 2010 Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12: 32–42.
- Pollock, D. D., W. R. Taylor, and N. Goldman, 1999 Coevolving protein residues: maximum likelihood identification and relationship to structure. *J. Mol. Biol.* 287: 187–198.
- Poon, A., 2005 The rate of compensatory mutation in the DNA bacteriophage X174. *Genetics* 170: 989–999.
- Poon, A., B. H. Davis, and L. Chao, 2005 The coupon collector and the suppressor mutation: estimating the number of compensatory mutations by maximum likelihood. *Genetics* 170: 1323–1332.
- Popadin, K., L. V. Polishchuk, L. Mamirova, D. Knorre, and K. Gunbin, 2007 Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large vs. small mammals. *Proc. Natl. Acad. Sci. USA* 104: 13390–13395.
- Presgraves, D. C., 2005 Recombination enhances protein adaptation in *Drosophila melanogaster*. *Curr. Biol.* 15: 1651–1656.
- Pybus, O. G., A. Rambaut, R. Belshaw, R. P. Freckleton, A. J. Drummond *et al.*, 2006 Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol. Biol. Evol.* 24: 845–852.
- Rand, D. M., and L. M. Kann, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* 13: 735–748.
- Rhesus Macaque Genome Sequencing and Analysis Consortium *et al.*, 2007 Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316: 222–234.
- Robertson, G., 1968 The spectrum of genetic variation, pp. 5–16 in *Population Biology and Evolution*, edited by R. Lewontin. Syracuse University Press, Syracuse, NY.
- Rocha, E. P. C., J. M. Smith, L. D. Hurst, M. T. G. Holden, J. E. Cooper *et al.*, 2006 Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J. Theor. Biol.* 239: 226–235.
- Satta, Y., 2001 Comparison of DNA and protein polymorphisms between humans and chimpanzees. *Genes Genet. Syst.* 76: 159–168.
- Savageau, M. A., 1992 Dominance according to metabolic control analysis: Major achievement or house of cards? *J. Theor. Biol.* 154: 131–136.
- Sawyer, S. A., and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. *Genetics* 132: 1161–1176.
- Sawyer, S. A., D. E. Dykhuizen, and D. L. Hartl, 1987 Confidence interval for the number of selectively neutral amino acid polymorphisms. *Proc. Natl. Acad. Sci. USA* 84: 6225–6228.
- Sawyer, S. A., R. J. Kulathinal, C. D. Bustamante, and D. L. Hartl, 2003 Bayesian analysis suggests that most amino acid replacements in *Drosophila* are driven by positive selection. *J. Mol. Evol.* 57(Suppl 1): S154–S164.
- Sawyer, S. A., J. Parsch, Z. Zhang, and D. L. Hartl, 2007 Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 104: 6504–6510.
- Schneider, A., B. Charlesworth, A. Eyre-Walker, and P. D. Keightley, 2011 A method for inferring the rate of occurrence and fitness effects of advantageous mutations. *Genetics* 189: 1427–1437.
- Shapiro, J. A., W. Huang, C. Zhang, M. J. Hubisz, J. Lu *et al.*, 2007 Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl. Acad. Sci. USA* 104: 2271–2276.
- Simmons, M. J., and J. F. Crow, 1977 Mutations affecting fitness in *Drosophila* populations. *Annu. Rev. Genet.* 11: 49–78.
- Slotte, T., J. P. Foxe, K. M. Hazzouri, and S. I. Wright, 2010 Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Mol. Biol. Evol.* 27: 1813–1821.
- Smith, N. G. C., and A. Eyre-Walker, 2002 Adaptive protein evolution in *Drosophila*. *Nature* 415: 1022–1024.
- Strasburg, J. L., N. C. Kane, A. R. Raduski, A. Bonin, R. Michelmore *et al.*, 2011 Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. *Mol. Biol. Evol.* 28: 1569–1580.
- Subramanian, S., 2012 The abundance of deleterious polymorphisms in humans. *Genetics* 190: 1579–1583.
- Subramanian, S., D. R. Denver, C. D. Millar, T. Heupink, A. Aschrafi *et al.*, 2009 High mitogenomic evolutionary rates and time dependency. *Trends Genet.* 25: 482–486.
- Sunyaev, S. R., W. C. Lathe, V. E. Ramensky, and P. Bork, 2000 SNP frequencies in human genes an excess of rare alleles and differing modes of selection. *Trends Genet.* 16: 335–337.
- Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437–460.
- Tajima, F., 1989 The effect of change in population size on DNA polymorphism. *Genetics* 123: 597–601.
- Takahata, N., 1993 Relaxed natural selection in human populations during the Pleistocene. *Jpn. J. Genet.* 68: 539–547.
- Tokuriki, N., and D. S. Tawfik, 2009a Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* 459: 668–673.



- Tokuriki, N., and D. S. Tawfik, 2009b Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* 19: 596–604.
- Tokuriki, N., F. Stricher, L. Serrano, D. S. Tawfik, and D. Eisenberg, 2008 How protein stability and new functions trade off. *PLoS Comput. Biol.* 4: e1000002.
- van Ham, R. C., J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal *et al.*, 2003 Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl. Acad. Sci. USA* 100: 581–586.
- Wang, X., G. Minasov, and B. K. Shoichet, 2002 Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* 320: 85–95.
- Warnecke, T., and E. P. C. Rocha, 2011 Function-specific accelerations in rates of sequence evolution suggest predictable epistatic responses to reduced effective population size. *Mol. Biol. Evol.* 28: 2339–2349.
- Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7: 256–276.
- Watterson, G. A., 1977 Heterosis or neutrality? *Genetics* 85: 789–814.
- Weber, C. C., and L. D. Hurst, 2009 Protein rates of evolution are predicted by double-strand break events, independent of crossing-over rates. *Genome Biol. Evol.* 1: 340–349.
- Weinreich, D. M., and D. M. Rand, 2000 Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. *Genetics* 156: 385–399.
- Weirauch, M. T., and T. R. Hughes, 2010 Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends Genet.* 26: 66–74.
- Welch, J. J., 2006 Estimating the genomewide rate of adaptive protein evolution in *Drosophila*. *Genetics* 173: 821–837.
- Wernegreen, J. J., 2011 Reduced selective constraint in endosymbionts: elevation in radical amino acid replacements occurs genome-wide. *PLoS ONE* 6: e28905.
- Wernegreen, J. J., and N. A. Moran, 1999 Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes. *Mol. Biol. Evol.* 16: 83–97.
- Wilke, C. O., and D. A. Drummond, 2010 Signatures of protein biophysics in coding sequence evolution. *Curr. Opin. Struct. Biol.* 20: 385–389.
- Wilkie, A. O., 1994 The molecular basis of genetic dominance. *J. Med. Genet.* 31: 89–98.
- Williamson, S. H., R. Hernandez, A. Fledel-Alon, L. Zhu, R. Nielsen *et al.*, 2005 Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl. Acad. Sci. USA* 102: 7882–7887.
- Wilson, D. J., R. D. Hernandez, P. Andolfatto, and M. Przeworski, 2011 A population genetics-phylogenetics approach to inferring natural selection in coding sequences. *PLoS Genet.* 7: e1002395.
- Wise, C. A., M. Sraml, and S. Easteal, 1998 Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. *Genetics* 148: 409–421.
- Woolfit, M., and L. Bromham, 2005 Population size and molecular evolution on islands. *Proc. Biol. Sci.* 272: 2277–2282.
- Wright, S., 1929 Fisher's theory of dominance. *Am. Nat.* 63: 274–279.
- Wright, S., 1934 Physiological and evolutionary theories of dominance. *Am. Nat.* 68: 24–53.
- Wright, S., 1938 The distribution of gene frequencies under irreversible mutation. *Proc. Natl. Acad. Sci. USA* 24: 253–259.
- Wright, S. I., and P. Andolfatto, 2008 The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. *Annu. Rev. Ecol. Evol. Syst.* 39: 193–213.
- Wylie, C. S., and E. I. Shakhnovich, 2011 A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc. Natl. Acad. Sci. USA* 108: 9916.
- Yeang, C.-H., and D. Haussler, 2007 Detecting coevolution in and among protein domains. *PLoS Comput. Biol.* 3: e211.
- Zhang, J., 2000 Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J. Mol. Evol.* 50: 56–68.
- Zhang, Z., and J. Parsch, 2005 Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression. *Mol. Biol. Evol.* 22: 1945–1947.
- Zuckermandl, E., and L. Pauling, 1965 Evolutionary divergence and convergence in proteins, pp. 97–166 in *Evolving Genes and Proteins*, edited by V. Bryson, and H. Vogel. Academic Press, New York.

Communicating editor: M. Turelli

# GENETICS

**Supporting Information**

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.140178/-/DC1>

## **Weak Selection and Protein Evolution**

Hiroshi Akashi, Naoki Osada, and Tomoko Ohta

**Table S1** Data used for protein/DNA polymorphisms

Species (population) name	$\pi_N$	$\pi_S$	# of gene	Reference
<b><i>Drosophila</i></b>				
<i>Drosophila melanogaster</i>	$1.9 \times 10^{-3}$	$2.3 \times 10^{-2}$	419	(SHAPIRO <i>et al.</i> 2007)
<i>D. simulans</i>	$2.6 \times 10^{-3}$	$3.5 \times 10^{-2}$	11,466	(BEGUN <i>et al.</i> 2007)
<i>D. ananassae</i>	$1.4 \times 10^{-3}$	$2.4 \times 10^{-2}$	29	(GRATH <i>et al.</i> 2009)
<i>D. americana</i>	$4.0 \times 10^{-4}$	$2.5 \times 10^{-2}$	10	(MASIDE and CHARLESWORTH 2007)
<i>D. miranda</i>	$7.2 \times 10^{-4}$	$6.6 \times 10^{-3}$	37	(HADRILL <i>et al.</i> 2010)
<i>D. pseudoobscura</i>	$6.6 \times 10^{-4}$	$2.3 \times 10^{-2}$	37	(HADRILL <i>et al.</i> 2010)
<i>D. mauritiana</i>	$1.1 \times 10^{-3}$	$1.9 \times 10^{-2}$	17	(MCDERMOTT and KLIMAN 2008)
<b>Yeasts</b>				
<i>Saccaromyces cerevisiae, European</i>	$5.2 \times 10^{-4}$	$1.4 \times 10^{-3}$	3132	(ELYASHIV <i>et al.</i> 2010)
<i>S. cerevisiae, Malaysian</i>	$1.6 \times 10^{-4}$	$4.4 \times 10^{-4}$	3132	(ELYASHIV <i>et al.</i> 2010)
<i>S. cerevisiae, sake</i>	$7.3 \times 10^{-4}$	$2.7 \times 10^{-3}$	3132	(ELYASHIV <i>et al.</i> 2010)
<i>S. paradoxus, European</i>	$5.0 \times 10^{-4}$	$2.1 \times 10^{-3}$	3132	(ELYASHIV <i>et al.</i> 2010)
<i>S. paradoxus, Far Eastern</i>	$3.8 \times 10^{-4}$	$1.5 \times 10^{-3}$	3132	(ELYASHIV <i>et al.</i> 2010)
<i>S. paradoxus, North American</i>	$7.4 \times 10^{-4}$	$4.6 \times 10^{-3}$	3132	(ELYASHIV <i>et al.</i> 2010)
<b>Plants</b>				
<i>Populus tremula</i>	$1.7 \times 10^{-3}$	$1.2 \times 10^{-2}$	77	(INGVARSSON 2010)
<i>Capsella grandiflora</i>	$1.8 \times 10^{-3}$	$2.3 \times 10^{-2}$	257	(SLOTTE <i>et al.</i> 2010)
<i>Arabidopsis thaliana</i>	$1.7 \times 10^{-3}$	$7.2 \times 10^{-3}$	483	(SLOTTE <i>et al.</i> 2010)
<i>A. halleri</i>	$2.6 \times 10^{-3}$	$1.7 \times 10^{-2}$	24	(HEIDEL <i>et al.</i> 2010)
<i>Helianthus petiolaris</i>	$3.4 \times 10^{-3}$	$3.1 \times 10^{-2}$	35	(STRASBURG <i>et al.</i> 2011)
<i>H. paradoxus</i>	$1.4 \times 10^{-3}$	$4.6 \times 10^{-3}$	35	(STRASBURG <i>et al.</i> 2011)
<i>H. exilis</i>	$3.1 \times 10^{-3}$	$2.0 \times 10^{-2}$	35	(STRASBURG <i>et al.</i> 2011)
<i>H. tuberosus</i>	$4.0 \times 10^{-3}$	$2.4 \times 10^{-2}$	35	(STRASBURG <i>et al.</i> 2011)

<i>H. argophyllus</i>	$1.8 \times 10^{-3}$	$1.5 \times 10^{-2}$	35 (STRASBURG <i>et al.</i> 2011)
<i>H. annuus</i>	$3.1 \times 10^{-3}$	$2.4 \times 10^{-2}$	35 (STRASBURG <i>et al.</i> 2011)
<i>Glycine max</i>	$4.1 \times 10^{-4}$	$7.9 \times 10^{-4}$	116 (ZHU <i>et al.</i> 2003)

Estimates of nucleotide diversity at nonsynonymous ( $\pi_N$ ) and synonymous ( $\pi_S$ ) sites from the previous studies. The number of analyzed genes is shown in the table.

#### References for Table S1

- BEGUN, D. J., A. K. HOLLOWAY, K. STEVENS, L. W. HILLIER, Y.-P. POH *et al.*, 2007 Population Genomics: Whole-Genome Analysis of Polymorphism and Divergence in *Drosophila simulans*. PLoS Biol. **5**: e310.
- ELYASHIV, E., K. BULLAUGHEY, S. SATTATH, Y. RINOTT, M. PRZEORSKI *et al.*, 2010 Shifts in the intensity of purifying selection: An analysis of genome-wide polymorphism data from two closely related yeast species. Genome Res. **20**: 1558-1573.
- GRATH, S., J. BAINES and J. PARSCH, 2009 Molecular evolution of sex-biased genes in the *Drosophila ananassae* subgroup. BMC Evol Biol **9**: 291.
- HADDRILL, P. R., L. LOEWE and B. CHARLESWORTH, 2010 Estimating the Parameters of Selection on Nonsynonymous Mutations in *Drosophila pseudoobscura* and *D. miranda*. Genetics **185**: 1381-1396.
- HEIDEL, A. J., S. E. RAMOS-ONSINS, W.-K. WANG, T.-Y. CHIANG and T. MITCHELL-OLDS, 2010 Population history in *Arabidopsis halleri* using multilocus analysis. Mol. Ecol. **19**: 3364-3379.
- INGVARSSON, P. K., 2010 Natural Selection on Synonymous and Nonsynonymous Mutations Shapes Patterns of Polymorphism in *Populus tremula*. Mol. Biol. Evol. **27**: 650-660.
- MASIDE, X., and B. CHARLESWORTH, 2007 Patterns of Molecular Variation and Evolution in *Drosophila americana* and Its Relatives. Genetics **176**: 2293-2305.
- MCDERMOTT, S. R., and R. M. KLIMAN, 2008 Estimation of Isolation Times of the Island Species in the *Drosophila simulans* Complex from Multilocus DNA Sequence Data. PLoS ONE **3**: e2442.
- SHAPIRO, J. A., W. HUANG, C. ZHANG, M. J. HUBISZ, J. LU *et al.*, 2007 Adaptive genic evolution in the *Drosophila* genomes. Proc. Natl. Acad. Sci. USA **104**: 2271-2276.
- SLOTTE, T., J. P. FOXE, K. M. HAZZOURI and S. I. WRIGHT, 2010 Genome-Wide Evidence for Efficient Positive and Purifying Selection in *Capsella grandiflora*, a Plant Species with a Large Effective Population Size. Mol. Biol. Evol. **27**: 1813-1821.
- STRASBURG, J. L., N. C. KANE, A. R. RADUSKI, A. BONIN, R. MICHELMORE *et al.*, 2011 Effective Population Size Is Positively Correlated with Levels of Adaptive Divergence among Annual Sunflowers. Mol. Biol. Evol. **28**: 1569-1580.
- ZHU, Y. L., Q. J. SONG, D. L. HYTEN, C. P. VAN TASSELL, L. K. MATUKUMALLI *et al.*, 2003 Single-Nucleotide Polymorphisms in Soybean. Genetics **163**: 1123-1134.