

# Impact of Genetic Architecture on the Relative Rates of X versus Autosomal Adaptive Substitution

Tim Connallon,<sup>\*1</sup> Nadia D. Singh,<sup>2</sup> and Andrew G. Clark<sup>1</sup>

<sup>1</sup>Department of Molecular Biology and Genetics, Cornell University

<sup>2</sup>Department of Genetics, North Carolina State University

**\*Corresponding author:** E-mail: tmc233@cornell.edu.

**Associate editor:** Hideki Innan

## Abstract

Molecular evolutionary theory predicts that the ratio of autosomal to X-linked adaptive substitution ( $K_A/K_X$ ) is primarily determined by the average dominance coefficient of beneficial mutations. Although this theory has profoundly influenced analysis and interpretation of comparative genomic data, its predictions are based upon two unverified assumptions about the genetic basis of adaptation. The theory assumes that 1) the rate of adaptively driven molecular evolution is limited by the availability of beneficial mutations, and 2) the scaling of evolutionary parameters between the X and the autosomes (e.g., the beneficial mutation rate, and the fitness effect distribution of beneficial alleles, per X-linked versus autosomal locus) is constant across molecular evolutionary timescales. Here, we show that the genetic architecture underlying bouts of adaptive substitution can influence both assumptions, and consequently, the theoretical relationship between  $K_A/K_X$  and mean dominance. Quantitative predictions of prior theory apply when 1) many genomically dispersed genes potentially contribute beneficial substitutions during individual steps of adaptive walks, and 2) the population beneficial mutation rate, summed across the set of potentially contributing genes, is sufficiently small to ensure that adaptive substitutions are drawn from new mutations rather than standing genetic variation. Current research into the genetic basis of adaptation suggests that both assumptions are plausibly violated. We find that the qualitative positive relationship between mean dominance and  $K_A/K_X$  is relatively robust to the specific conditions underlying adaptive substitution, yet the quantitative relationship between dominance and  $K_A/K_X$  is quite flexible and context dependent. This flexibility may partially account for the puzzlingly variable X versus autosome substitution patterns reported in the empirical evolutionary genomics literature. The new theory unites the previously separate analysis of adaptation using new mutations versus standing genetic variation and makes several useful predictions about the interaction between genetic architecture, evolutionary genetic constraints, and effective population size in determining the ratio of adaptive substitution between autosomal and X-linked genes.

**Key words:** dominance, epistasis, genetics of adaptation, soft sweeps, molecular evolution.

## Introduction

Molecular evolutionary contrasts between the X chromosome and the autosomes occupy a prominent position within population and comparative genomics (Vicoso and Charlesworth 2006; Mank 2009; Ellegren 2011). This prominence reflects one of the primary weaknesses of molecular evolutionary data: although sequence diversity and divergence patterns provide a partial record of natural selection in nature, the absence of accompanying phenotypic or fitness data severely limits opportunities to infer process from observed sequence patterns. This limitation is somewhat alleviated by a well-developed population genetics theory, which describes the specific conditions that generate evolutionary differences between the X and the autosomes. Analysis of interspecific divergence of molecular sequences, when explicitly informed and motivated by these theoretical models, facilitates inferences of the population genetic details of adaptation.

X-autosome contrasts are particularly useful for elucidating the genetic factors constraining adaptation. In a classic paper that has become hugely influential within the field

of comparative genomics, Charlesworth et al. (1987) showed that the relative rate of adaptive substitution between X-linked and autosomal genes may be determined by a single population genetic parameter: the mean dominance coefficient of beneficial mutations (traditionally depicted as  $h$ , with partially recessive effects corresponding to  $0 < h < 1/2$ ; partially dominant effects correspond to  $1/2 < h < 1$ ). The underlying model follows from the pioneering work of Kimura and Ohta (1971a, 1971b, 1971c) (see also Kimura 1971, 1979), which considered the long-term rate of adaptive substitution per gene to be a product of two constraints: the rate at which new beneficial mutations are produced in each generation and the probability that each mutant allele successfully becomes fixed (the latter constraint being a positive function of  $h$ ). By assuming a beneficial mutation rate per gene of  $\nu$  ( $1 \gg \nu > 0$ ), selection and dominance coefficients per beneficial mutation of  $s$  and  $h$  (respectively;  $1 \gg s > 0$ ), and an effective population size per autosomal and X-linked gene of  $2N_e$  and  $1.5N_e$  (respectively;  $sh \gg 1/N_e$ ), Charlesworth et al. (1987) showed that the rate of adaptive substitution is  $K_A \approx 4N_e\nu sh$  per

autosomal gene and  $K_x \approx N_e v s (1 + 2h)$  per X-linked gene. The ratio of these individual rates,

$$\frac{K_A}{K_x} = \frac{4h}{1 + 2h}, \quad (1)$$

is therefore independent of all parameters except the dominance coefficient. This result is empirically useful for two related reasons: first, it provides a process-based explanation for observed molecular evolutionary differences between the X and the autosomes (e.g., faster-X substitution is expected when beneficial alleles tend to have recessive fitness effects:  $h < 0.5$ ) and second, it establishes the necessary theoretical framework for estimating mean dominance of beneficial mutations using sequence data (Orr 2010). Subsequent generalizations of the model quantitatively alter the above result, yet similarly predict a high sensitivity of  $K_A/K_x$  to dominance (Charlesworth et al. 1987; Kirkpatrick and Hall 2004; Vicoso and Charlesworth 2009).

Although this classic theory has been used extensively within comparative genomics (reviewed in Vicoso and Charlesworth 2006; Presgraves 2008; Mank et al. 2010b; Orr 2010; Ellegren 2011), its relevance to molecular divergence data remains ambiguous, as it relies upon two critical, yet unverified, assumptions about the genetic basis of adaptation. The model assumes that 1) new beneficial mutations rather than standing genetic variation contribute to adaptation (Orr and Betancourt 2001; also see p. 123 of Charlesworth et al. 1987) and 2) X-linked and autosomal sites mutate to beneficial alleles with comparable fitness effects and at comparable rates during the span of molecular evolution (e.g., the terms  $v$ ,  $s$ , and  $h$  are similar between the X and the autosomes). Although violation of either assumption may alter the relationship between  $K_A/K_x$  and dominance (e.g., Orr and Betancourt 2001), it is currently unclear how much this should pose a concern for the analysis of sequence data. More importantly, it is unclear how these assumptions relate to the biology of adaptation. Under what conditions of adaptation are these assumptions likely to be valid? To the extent that these assumptions will sometimes be violated, how severely will the theoretical relationship between  $K_A/K_x$  and dominance be altered?

With these unresolved issues in mind, we present a new model for the relative rate of X versus autosome adaptive substitution. Following previous biologically inspired models of adaptation—such as Fisher’s “geometric model” (Fisher 1958; Orr 1998) and Gillespie’s “mutational landscape model” (Gillespie 1984, 1991; Orr 2002, 2005a, 2005b)—we conceptualize adaptively driven molecular evolution as an iterated series of discrete selective sweeps, with each sweep (or “step” of adaptation) leading to the fixation of a beneficial allele. We show that the relationship between  $K_A/K_x$  and dominance is heavily influenced by the genetic architecture underlying the individual steps of adaptation. Predictions of the classic theory (e.g., eq. 1) emerge when the following three conditions are met: 1) many genes “compete” to fix substitutions during individual steps of adaptation, 2) these genes are uniformly dispersed across the genome, and 3) the population’s total beneficial mutation rate is sufficiently small to ensure that substitutions are drawn from new beneficial mutations

rather than standing genetic variation. Violation of any of these three conditions can dampen or even eliminate the sensitivity of  $K_A/K_x$  to dominance. Moreover, the likelihood that a lineage evolves within the parameter space described by the classic theory depends upon its historical effective population size. Substitution dynamics under the classic model are more relevant to lineages with small effective population sizes than to lineages with large population sizes. These new theoretical results are discussed within the context of emerging data on the genetic basis of adaptation (e.g., parallel evolution and repeatability; see Discussion) and species-specific patterns of substitution.

## Materials and Methods, and Results

### The Basic Model: Adaptation Using New Mutations

Our model is dynamically similar to Gillespie’s “mutational landscape” model of DNA sequence adaptation (Gillespie 1984, 1991; Orr 2002, 2005a, 2005b), in that the long-term process of molecular evolution can be divided into a series of steps of adaptive substitution. Prior analyses of the mutational landscape model consider DNA sequence evolution of single genes (or very small genomes; e.g., Orr 2002, 2005a, 2005b; Unckless and Orr 2009), yet the model’s basic framework easily accommodates multiple genes under arbitrary patterns of linkage, with each gene essentially competing to fix a beneficial substitution during individual steps of adaptive walks. During an arbitrary step of adaptation, we suppose that a beneficial substitution is drawn from specified set of  $n$  relevant genes ( $n \geq 1$ ) that compete to contribute the next substitution in the series. The probability that a given gene, within a set of  $n$ , contributes the next substitution depends on each gene’s mutation rate to beneficial alleles, the distribution of selection and dominance coefficients for those beneficial alleles, and each gene’s location within the genome.

Because evolutionary steps are discrete, the model accommodates a broad range of evolving epistatic relationships between loci within the context of individual adaptive walks. At one extreme, genotype-fitness landscapes may be “rugged” (or highly epistatic; for discussion, see Unckless and Orr 2009), so that positive selection at a set of  $n > 1$  competing genes (during a single iteration of the model) is resolved by the first substitution to fix among the  $n$  genes. Competition between entirely new sets of genes may describe subsequent steps of adaptation. At the opposite extreme, where fitness landscapes are “smooth” (or nonepistatic), beneficial mutation and selection at each locus remain independent of substitution events occurring elsewhere within the genome (Unckless and Orr 2009). This latter case, within the context of our model, corresponds to a scenario of beneficial substitution without competition; individual steps of adaptation are resolved by substitutions at single genes (i.e.,  $n = 1$ ). The nature of epistasis can vary across different contexts of adaptation, with the underlying genetic properties of  $n$  likely to vary among individual bouts of adaptive substitution.

For ease of presentation and comparison with prior theory, we initially make four simplifying assumptions (and subsequently relax each):

- Selection and dominance parameters are the same in each sex, and the population size of the X is three-quarters that of the autosomes ( $2N_e = N_{eA} = 4N_{ex}/3$ , where the subscripts refer to autosomal and X-linkage, respectively).
- For a given substitutional step involving  $n$  genes, the location of each (X-linked or autosomal) is randomly assigned and independent of the other  $n - 1$  genes, such that, if the total genomic fraction of X-linked genes in a genome is  $p$ , then the probability that a single gene is X-linked is  $p$  (the probability of autosomal linkage is  $1 - p$ ). Under such conditions, genomic locations will follow a binomial distribution, with  $x$  X-linked copies and  $a$  autosomal copies [ $x \sim \text{Binomial}(n, p)$ ;  $a = n - x$ ].
- Each of  $n$  genes is a mutationally equivalent genetic unit. That is, for the particular time point in question (i.e., step  $i$  during a given adaptive walk), each of the genes mutates to a beneficial allele at rate  $v_i$ , and the selection parameters for each beneficial mutation are drawn from common distributions. Thus,  $E(s_i)$  and  $E(s_i h_i)$  will be the same for each gene.
- “Strong selection/weak mutation” (SSWM; Gillespie 1984, 1991; Orr 2002) conditions apply across each set of competing genes: that is,  $nv_i < 1/(2N) \ll s_i h_i \ll 1$ .

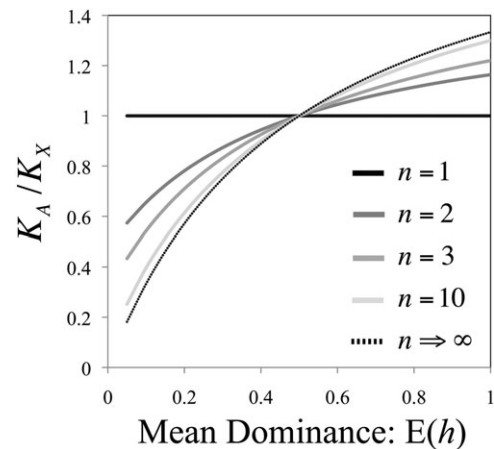
For each of the genes on the autosomes (and given the specified assumptions), the probability that a beneficial mutation arises within a single generation and becomes established in the population will be  $\sim 4N_e v_i E(s_i h_i)$ . For each X-linked gene, this probability will be  $\sim N_e v_i E[s_i(1 + 2h_i)]$ . For an arbitrary step  $i$  during adaptation, which involves  $x$  X-linked and  $a$  autosomal genes, the probability of an X-linked substitution will be

$$\begin{aligned} \Pi_X &= \frac{N_e v_i E[s_i(1 + 2h_i)]x}{N_e v_i E[s_i(1 + 2h_i)]x + 4N_e v_i E(s_i h_i)a} \quad (2) \\ &= \frac{\psi}{\beta + \psi(1 - \beta)}, \end{aligned}$$

where  $\psi = x/n$  and  $\beta = 4E(s_i h_i)/E[s_i(1 + 2h_i)]$ . Assuming  $\beta$  is constant, and averaging across all combinations of  $a$  and  $x$  (for a given  $n$ ), the proportion of adaptive substitutions that are X-linked during substitution steps involving  $n$  genes will be

$$\begin{aligned} E(\Pi_X) &= \sum_{x=0}^n \Pi_X \binom{n}{x} p^x (1-p)^{n-x} \\ &= \begin{cases} p & \text{for } n = 1 \\ \frac{2p(1-p)}{1+\beta} + p^2 & \text{for } n = 2 \\ 3p(1-p) \left( \frac{2+\beta(1+3p)}{(1+2\beta)(2+\beta)} \right) + p^3 & \text{for } n = 3 \\ \frac{p}{\beta+p(1-\beta)} - \frac{\beta(1-\beta)p(1-p)}{n[\beta+p(1-\beta)]^2} & \text{for } n \gg 1 \end{cases} \quad (3) \end{aligned}$$

The first three results ( $n = 1, 2, 3$ ) are exact, whereas the last is a good approximation for  $n \gg 1$ . Controlling for the proportion of the genome that is autosomal or



**FIG. 1.** The relative rate of X versus autosome adaptive substitution depends on the number of genes ( $n$ ) involved in adaptive walks. Equation (3) with  $\beta = 4E(h)/[1 + 2E(h)]$ ,  $p = 0.1$ , and  $n$  specified (see figure legend) were used to calculate the autosome versus X-linked adaptive substitution rate,  $K_A/K_X = \{[1 - E(\Pi_X)]/(1 - p)\}/[E(\Pi_X)/p]$ . For  $n \rightarrow \infty$ ,  $K_A/K_X = 4E(h)/[1 + 2E(h)]$ , which parallels the result of Charlesworth et al. (1987).

X-linked provides the relative substitution rate per locus:  $K_A/K_X = \{[1 - E(\Pi_X)]/(1 - p)\}/[E(\Pi_X)/p]$ . This is constrained to fall within the range of  $\min\{\beta, 1/\beta\} < K_A/K_X < \max\{\beta, 1/\beta\}$ .  $K_A/K_X = 1$  when  $n = 1$ , and  $K_A/K_X \approx 4E(h)/[1 + 2E(h)]$  when  $n \gg 1$ , assuming  $E(s)E(h) \gg \sqrt{r_{sh}^2 \text{var}(s)\text{var}(h)}$ , where  $r_{sh}$  represents the correlation coefficient between selection and dominance parameters (this assumption requires that parameters  $s$  and  $h$  are not strongly correlated, per beneficial allele;  $r_{sh} > 0$  will dampen and  $r_{sh} < 0$  will exacerbate faster-X substitution patterns). For substitution steps that may only involve one gene ( $n = 1$ ), no substitution rate differences are expected between X-linked and autosomal loci, whereas for adaptive walks involving multiple genes,  $K_A/K_X$  is sensitive to dominance, and this degree of sensitivity ultimately converges (for large- $n$ ) to the theoretical predictions of Charlesworth et al. (1987) (eq. 1; see fig. 1).

A simple biological reason accounts for this pattern transition between  $n = 1$  to  $n > 1$ . When single (or very few) genes can respond to a specific environmental challenge by fixing beneficial alleles, substitutions during the adaptive walk are constrained to fall within these genes, wherever they happen to reside within a genome. When many genes (spread throughout the genome) mutate to beneficial alleles relevant to a particular context of adaptation, those that have more favorable patterns of linkage (i.e., that are X-linked when  $E(h) < 1/2$  or autosomal when  $E(h) > 1/2$ ) tend to “outcompete” those with less favorable linkage patterns by contributing a disproportionate share of substitutions during each adaptive walk. This effect can be dampened when  $n$  is small because of the increased probability that all  $n$  have the same pattern of linkage (i.e., all  $n$  are autosomal or all are X-linked, in which case the entire series of beneficial substitutions will necessarily

become fixed within a single linkage group), whereas when  $n$  is large, competition between X-linked and autosomal genes occurs with probability approaching one, yielding a result similar to that of Charlesworth et al. (1987).

### Generalized Model Behavior under SSWM Conditions

Excluding the SSWM condition (which we return to further below), violation of the remaining assumptions will trivially modify the above results, as follows:

Assumption 1: Selection parameters are the same in each sex. Equation (3) remains applicable, though sex-specific fitness effects will alter the ratio  $\beta$ . For example, under female-limited selection,  $\beta_f = 1$  and thus  $K_A/K_X = 1$  for all  $n$ . Under male-limited selection,  $\beta_m = 2E(s_i h_i)/E(s_i)$  and  $K_A/K_X$  becomes slightly more sensitive to dominance as long as  $n > 1$  and remains insensitive to dominance when  $n = 1$ .

Assumption 2: Effective population size on the X is three-quarters that of the autosomes ( $4N_{ex}/3 = N_{eA}$ ). Although it is often assumed that the effective population size for an X-linked locus is three-quarters that of an autosomal locus, this may not always be the case (Charlesworth 2009; Vicoso and Charlesworth 2009). Violation of this assumption can affect the relative rate of adaptive substitution between the X and the autosomes by altering the relative population mutation rate to beneficial alleles on the two chromosomes (e.g.,  $N_{ex}\nu/N_{eA}\nu$ ). For  $4N_{ex}/3 \neq N_{eA}$ , the results above remain unchanged when  $n = 1$ . For  $n > 1$ , the curves in figure 1 are shifted downwards when  $4N_{ex}/3 > N_{eA}$  and shifted upwards when  $4N_{ex}/3 < N_{eA}$  (thereby increasing and decreasing tendency to faster-X substitution, respectively). Despite these shifts, the degree of sensitivity of  $K_A/K_X$  to dominance will nevertheless remain unaltered.

Assumption 3: Genomic locations of each member of a set of  $n$  genes are independent (i.e., their distribution with respect to the X is binomial). An alternative possibility is that genes that comprise each set of  $n$  spatially cluster within a genome, as sometimes occurs for coexpressed or functionally related genes in eukaryotes (e.g., Lee and Sonnhammer 2003; Hurst et al. 2004; Michalak 2008). Such clustering, by inflating the variance of  $x$  above the binomial expectation, i.e.,  $\text{var}(x) > np(1 - p)$ , will reduce the probability that X-linked and autosomal genes compete to fix substitutions. This will dampen sensitivity of  $K_A/K_X$  to dominance. As the distribution's variance becomes progressively greater and ultimately approaches  $\text{var}(x) = n^2p(1 - p)$ ,  $E(\Pi_X)$  converges to  $p$  and  $K_A/K_X$  converges to one, for all values of  $n$ .

Assumption 4: Genes within each set of  $n$  are mutationally equivalent (that is,  $\nu_i$ ,  $E(s_i)$ , and  $E(s_i h_i)$  are similar among genes). Genetic equivalence is an important assumption of many models that contrast the X and au-

tosomes, but this may be violated if genes vary in their beneficial mutation rates (e.g., the number of sites mutating to a beneficial allele may vary) or selection coefficients, with respect to particular contexts of adaptation. When a large number of genes contribute beneficial alleles ( $n$  is large), and there is no inherent bias between the X and the autosomes, such variability between genes will tend to average out across sets of X-linked and autosomal genes. If  $n$  is relatively small, a gene with particularly high mutation rate or large beneficial selection coefficients will tend to dominate during the step of an adaptive walk. When one gene dominates a system of  $n$  genes, the models behave increasingly like adaptive walks involving a single gene (i.e.,  $n = 1$ ), leading to  $K_A/K_X \approx 1$ .

### Adaptation Using New Mutations and Standing Genetic Variation (Moderate SSWM Violation)

The predictions of faster-X theory critically depend on the validity of SSWM conditions, which ensure that adaptation involves the fixation of new beneficial mutations (see p. 123 of Charlesworth et al. 1987; Orr and Betancourt 2001; Orr 2010). Because SSWM conditions require a sufficiently low total beneficial mutation rate, SSWM violation becomes increasingly likely as the number of genes (and hence, the number of mutable sites) involved in an adaptive walk increases. The probability of adaptation using standing genetic variation increases with a population's total beneficial mutation rate ( $2N_e\nu$ ). As  $2N_e\nu$  becomes increasingly large, simple adaptive substitution models will ultimately break down because 1) adaptation is no longer limited by the availability of beneficial genetic variation, and 2) adaptation involves allele frequency changes at many sites throughout a genome (e.g., "polygenic adaptation"; Pritchard and Di Rienzo 2010; Pritchard et al. 2010) rather than a series of selective sweeps. Theoretical predictions regarding X versus autosome substitution rates are also likely to break down under polygenic adaptation, involving a very large number of genes within a genome (we return to this topic in the Discussion).

Much more progress can be made under moderate SSWM violations, where the individual steps of adaptation remain at least somewhat mutation limited. In such cases, simple models of adaptive substitution remain reasonable, despite a proportion of substitutions now being derived from standing genetic variation. Prior theory has focused on the probability of adaptation from standing genetic variation at single genes (e.g., Orr and Betancourt 2001; Hermisson and Pennings 2005), yet the general approach of these single-gene models can be extended to a multigene framework involving a combination of X-linked and autosomal loci. To simplify the following calculations, we assume a low probability of adaptation using standing genetic variation per locus, which will be reasonable if the beneficial mutation rate per gene is small ( $2N_e\nu \ll 1$ , as seems likely) and each array of beneficial alleles (at a given substitutional step) was deleterious prior to the bout of

positive selection (though this latter assumption is not critical; see Appendix).

When adaptation uses a combination of new mutations and standing genetic variation, we can characterize the relative rate of substitution on the autosomes and X as a function of the conditional ratios:

$$\frac{K_A}{K_X} = [1 - \text{Pr}(\text{SGV})] \left( \frac{K_A}{K_X} \right)_{\text{new}} + \text{Pr}(\text{SGV}) \left( \frac{K_A}{K_X} \right)_{\text{SGV}}, \quad (4)$$

where  $(K_A/K_X)_{\text{new}}$  is given by SSWM results (e.g., for  $n \gg 1$ ,  $(K_A/K_X)_{\text{new}} \approx 4E(h)/[1 + 2E(h)]$ , from above),  $(K_A/K_X)_{\text{SGV}}$  is the autosome to X substitution rate ratio, conditional on adaptation using the standing genetic variation, and  $\text{Pr}(\text{SGV})$  is the probability of adaptation from standing genetic variation during a step of an adaptive walk.

To calculate the probability of adaptation from standing genetic variation, suppose that, from a large pool of  $n$  genes,  $\sim n(1-p)$  are on the autosomes and  $\sim np$  are X-linked. The probability that at least one segregating autosomal allele becomes established in the population is

$$\begin{aligned} \text{Pr}(\text{SGV}|\text{A total}) &= 1 - \prod_{i=1}^{n(1-p)} [1 - \text{Pr}(\text{SGV}|\text{locus } i)] \\ &\approx 1 - \exp(-n(1-p)E[\text{Pr}(\text{SGV}|\text{A locus})]), \end{aligned} \quad (5)$$

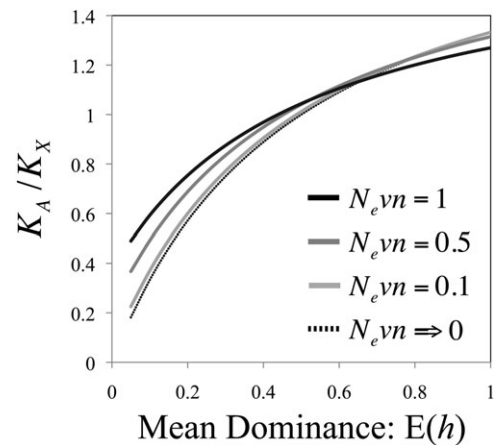
with the approximation appropriate for  $E[\text{Pr}(\text{SGV}|\text{A locus})] \approx \frac{\sum_{i=1}^{n(1-p)} \text{Pr}(\text{SGV}|\text{locus } i)}{n(1-p)} \ll 1$ . Similarly, the probability that at least one X-linked allele invades will be

$$\text{Pr}(\text{SGV}|\text{X total}) \approx 1 - \exp(-npE[\text{Pr}(\text{SGV}|\text{X locus})]), \quad (6)$$

where  $E[\text{Pr}(\text{SGV}|\text{X locus})] \approx \frac{\sum_{j=1}^{np} \text{Pr}(\text{SGV}|\text{locus } j)}{np} \ll 1$ . Given equations (5) and (6), the probability of adaptation from standing genetic variation will be  $\text{Pr}(\text{SGV}) = 1 - [1 - \text{Pr}(\text{SGV}|\text{A total})][1 - \text{Pr}(\text{SGV}|\text{X total})]$ . If dominance does not differ between beneficial and deleterious alleles and purifying selection is much greater than positive selection ( $h = h_d$ ;  $s/s_d \ll 1$ ; where  $s_d$  and  $h_d$  are the selection and dominance coefficients for a deleterious allele), then the mean probability of adaptation from the standing genetic variation is  $E[\text{Pr}(\text{SGV}|\text{X locus})] \approx 3N_e\nu E[s/s_d]$  and  $E[\text{Pr}(\text{SGV}|\text{A locus})] \approx 4N_e\nu E[s/s_d]$  (see Orr and Betancourt 2001; Hermisson and Pennings 2005), which we use for subsequent calculations, noting that our qualitative results remain applicable when these conditions do not hold (see Appendix). By incorporating these expectations into equations (5) and (6), the probability of adaptation from standing genetic variation simplifies to

$$\text{Pr}(\text{SGV}) \approx 1 - \exp(-N_e\nu n(4-p)E[s/s_d]). \quad (7)$$

For adaptation from standing variation, there are two cases of interest. When X-linked or autosomal alleles invade from SGV (but not invasion of both X-linked and autosomal alleles), then there will be no direct competition between intermediate-frequency alleles on the two chromosome types. Consequently, the chromosome type associated with invasion will ultimately contribute the next substitution during the adaptive walk. If instead alleles simultaneously invade



**Fig. 2.** The relative rate of X versus autosome adaptive substitution when a proportion of substitutions are drawn from standing genetic variation. Results are based on equations (4), (7), and (9), with  $(K_A/K_X)_{\text{new}} = 4E(h)/[1 + 2E(h)]$ ,  $p = 0.1$ , and  $E[s/s_d] = 0.1$ . Given the parameters, the corresponding total probabilities of adaptation using standing genetic variation are as follows:  $\text{Pr}(\text{SGV}) = 0.038$  for  $N_e\nu n = 0.1$ ;  $\text{Pr}(\text{SGV}) = 0.177$  for  $N_e\nu n = 0.5$ ; and  $\text{Pr}(\text{SGV}) = 0.323$  for  $N_e\nu n = 1$ . For  $N_e\nu n \rightarrow 0$ ,  $K_A/K_X = 4E(h)/[1 + 2E(h)]$ , which corresponds to the high- $n$ /low- $N_e\nu n$  result under SSWM conditions and parallels the result of Charlesworth et al. (1987).

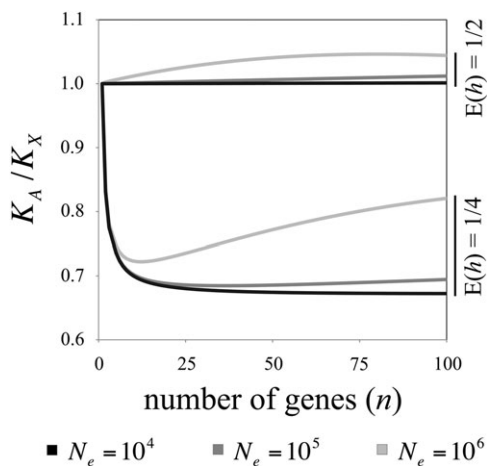
on the X and autosomes, then the next substitution to occur depends on the outcome of a direct race to fixation between those X-linked and autosomal alleles that have reached intermediate frequency. The probability that both X-linked and autosomal alleles invade simultaneously, given adaptation from standing genetic variation, is

$$\begin{aligned} \text{Pr}(\text{X and A}|\text{SGV}) &\approx \\ &\frac{[1 - \exp(-4N_e\nu n(1-p)E[s/s_d])][1 - \exp(-3N_e\nu npE[s/s_d])]}{1 - \exp(-N_e\nu n(4-p)E[s/s_d])}. \end{aligned} \quad (8)$$

Numerical evaluation of equation (8) shows that  $\text{Pr}(\text{X and A}|\text{SGV})$  is very small when the overall probability of adaptation from the SGV is less than one half and the genomic proportion of X-linked genes is within reasonable limits (e.g.,  $p < 0.2$ , as is typical for sex chromosomes; White 1973; Bull 1983) (see supplementary fig. S1, Supplementary Material online). Thus, for moderate violations of SSWM, which we consider here, cases of direct competition between intermediate frequency alleles on the X and autosomes can be safely ignored. This greatly simplifies the calculation of  $(K_A/K_X)_{\text{SGV}}$ , which now only depends on invasion probabilities on the two chromosomes:

$$\begin{aligned} \left( \frac{K_A}{K_X} \right)_{\text{SGV}} &\approx \frac{\text{Pr}(\text{SGV}|\text{A total})p}{\text{Pr}(\text{SGV}|\text{X total})(1-p)} \\ &\approx \frac{p}{1-p} \cdot \frac{1 - \exp(-4N_e\nu n(1-p)E[s/s_d])}{1 - \exp(-3N_e\nu npE[s/s_d])}. \end{aligned} \quad (9)$$

Note that, within the limit of  $N_e\nu nE[s/s_d] \rightarrow 0$ , equation (9) reduces to  $(K_A/K_X)_{\text{SGV}} \approx 4/3$ , which is the same as the ratio of X versus autosome invasion probabilities, per locus, first derived by Orr and Betancourt (2001).



**Fig. 3.** Effective population size and the relative rate of X versus autosome substitution. Results are based on equations (4), (7), and (9), with  $(K_A/K_X)_{\text{new}} = 4E(h)/[1 + 2E(h)]$ ,  $p = 0.1$ ,  $E[s/s_d] = 0.1$ , and  $\nu = 10^{-8}$ .

The relative rate of substitution on the autosomes versus X, given moderate violation of SSWM conditions, can be evaluated using equations (4), (7), and (9) (see fig. 2). Adaptation from standing genetic variation blunts substitution rate differences between X-linked and autosomal loci, making  $K_A/K_X$  less sensitive to the average dominance of beneficial alleles. This general effect is not limited to the region of parameter space evaluated analytically; similar results emerge under adaptation from segregating alleles that were previously neutral or mildly deleterious (Appendix). Moreover, because the probability of adaptation from standing genetic variation increases with effective population size,  $K_A/K_X$  ratios will differ between lineages with different historical  $N_e$ . For example, if beneficial alleles are (on average) partially recessive, “faster-X” patterns are accentuated in lineages with small effective size and dampened in lineages with large  $N_e$  (fig. 3). If beneficial alleles have additive to dominant fitness effects, population size effects on  $K_A/K_X$  are relatively modest (though “faster-autosome” effects may be slightly amplified in lineages with large  $N_e$ ; e.g., fig. 3).

## Discussion

Evolutionary models of X versus autosome adaptive substitution are typically based upon two critical assumptions about the biology of adaptation. During the course of molecular evolution, if 1) X-linked and autosomal genes consistently exhibit similar properties of beneficial mutation (i.e., the proportion of sites mutable to beneficial alleles, and the distribution of selection coefficients among beneficial alleles, does not systematically differ between X-linked and autosomal loci) and 2) adaptation uses new mutations rather than standing genetic variation, then  $K_A/K_X$  is predicted to be a simple function of the mean dominance coefficient among beneficial alleles (Charlesworth et al. 1987; Kirkpatrick and Hall 2004; Vicoso and Charlesworth 2009; Orr 2010).

We show that the validity of these assumptions should depend on population size and the genetic architecture underlying the individual steps of adaptive substitution. Assumptions of prior theory break down when 1) beneficial substitutions are constrained to involve few contributing genes during individual steps of adaptive substitution or 2) the cumulative beneficial mutation rate, summed across a set of potentially contributing genes, is sufficiently large (relative to  $N_e$ ) that standing genetic variation contributes to adaptive substitution. The impact of dominance on the relative rate of adaptive substitution ultimately depends on the interaction between the distribution of  $n$  among the individual steps of adaptation and the historical effective population size in which adaptive substitution occurs. Prior theory applies when  $n \gg 1 \gg N_e \nu n$  (on average), in which case, empirical estimates of  $K_A/K_X$  are particularly useful for inferring properties of the distribution of dominance coefficients among beneficial mutations—critical, yet poorly understood evolutionary parameters (see Orr 2010).

The relative rate of X versus autosome adaptive substitution depends on the parameter range in which natural populations typically evolve. Below, we consider the predictions of the theory within the context of current empirical observations of the genetic basis of adaptation. This growing body of research suggests that bouts of adaptation are often constrained to involve few genes (and thus, small- $n$  parameter scenarios are generally plausible). Given the sensitivity of  $K_A/K_X$  to both genetic architecture and effective population size, we outline aspects of the biology of adaptation that may be useful for informing predictions about X versus autosome molecular evolution. In general, the degree to which comparative genomic data reflect properties of dominance will likely be context specific, yet this context specificity may be predictable.

## Models of Sequential Substitution versus Polygenic Adaptation

The theory developed here might be useful for interpreting empirical patterns of X versus autosome substitution rates, yet our discussion of the theoretical predictions (see below) comes with an important caveat. This theory is based upon an adaptive walk model in which most “competition” between the X and the autosomes (if any) is indirect. Each bout of competition involves a race between alleles at X-linked versus autosomal loci to become established within the population (i.e., to “invade” from low frequency by avoiding stochastic loss). The race to fixation is reasonably predicted by rates of establishment as long as there is a low probability of both X-linked and autosomal alleles invading at the same time. Generally, the probability of concurrent invasion will be low as long as most adaptation involves the fixation of new beneficial mutations rather than standing genetic variation (see eq. 8 and surrounding text; supplementary fig. S1, Supplementary Material online).

Simple substitution models ultimately break down as  $N_e \nu n$  becomes sufficiently large that population adaptation is completely unconstrained by the availability of beneficial

genetic variation (i.e., when adaptation is not mutation limited). With increasing  $N_e\mu n$ , the genetic basis of adaptation transitions from a model of sequential selective sweeps (which we consider here) to one of polygenic adaptation involving allele frequency changes at many loci (see Pritchard and Di Rienzo 2010; Pritchard et al. 2010). Translating polygenic models of adaptation into empirical predictions for the rate of adaptive substitution (particularly within the context of X versus autosome contrasts) is likely to be a considerable challenge for future theory. Because alleles at contributing loci directly compete under polygenic adaptation, the details of epistasis between each pair of beneficial alleles must be specified in order to characterize the evolutionary fate of each variant within an entire array of segregating beneficial alleles. The basic population genetic dynamics of adaptation are also likely to change in polygenic models 1) because the substitution process involves an important stochastic filtering phase, following bouts of positive selection (e.g., Barton 1989; Pritchard et al. 2010), and 2) because the distribution of fitness effects among beneficial mutations and substitutions becomes increasingly sensitive to environmental properties rather than to intrinsic genetic factors (e.g., Charlesworth et al. 1987, p. 127–128; Kopp and Hermisson 2007, 2009). Explicitly developing models of DNA substitution by polygenic adaptation represents a fertile area for future theory.

Despite this caveat, species with abundant evidence for positively selected molecular substitutions (such as *Drosophila*) also exhibit patterns of polymorphism consistent with sequential sweep models (Sattath et al. 2011, though incidence of hard sweeps may be lower than expected based on long-term estimates of adaptive molecular evolution; see also Hernandez et al. 2011). Likewise, many adaptive phenotypes have a simple genetic basis involving few genes or substitutions of large effect (see below). These two lines of empirical evidence suggest that sequential sweep models of adaptation are indeed plausible, given what we currently know about the genetics of adaptation. The relative importance of sequential sweep models obviously remains an open question that should be vigorously pursued.

### Parallel Evolution and the Number of Genes Involved during Bouts of Adaptation

It is far from clear whether competition between loci should generally be important during individual bouts of adaptive substitution. For example, beneficial substitution to improve the stability or function of single proteins may involve competition among sites within a protein sequence rather than competition between sites within different genes. Positive selection to improve the performance or stability of single genes, independent of other genes in a genome, should generate noncompetitive ( $n = 1$ ) bouts of adaptive substitution. Under such scenarios, each gene's adaptive substitution rate, as calculated with molecular divergence data, may be relatively independent of intrinsic genetic factors, such as the mutation rate per site, selection and dominance coefficients, or local effective population size (e.g., Gillespie 2004). Adaptation of genetically complex traits would seem more likely

to involve allele frequency changes and substitutions at many genes. To the extent that DNA substitutions are driven in response to directional selection on complex traits, our results involving widespread competition between genes (large- $n$ ) may be particularly applicable.

When many genes mutate to beneficial alleles and each potentially contributes to adaptive substitution, there may be many distinct evolutionary genetic pathways (or solutions) to a favored phenotype. Repeated bouts of adaptation to the same environmental condition should therefore generate nonrepeatable patterns of substitution, involving different sets of contributing genes. In contrast, for those adaptive phenotypes that have a simple genetic basis—where substitutions in very few genes facilitate evolution of the favored trait—replicated bouts of adaptation should often involve parallel (repeated) substitution trajectories, involving the same underlying genes (for a formal theory relating parallelism to the number of competing sites during adaptive substitution, see Orr 2005c). A large experimental literature explicitly deals with this issue of genetic parallelism and repeatability during adaptation, by using a combination of experimental evolution and genetic mapping of convergently evolved adaptive phenotypes (experimental evolution: e.g., Weinreich et al. 2006; Bollback and Huelsenbeck 2009; Rokyta et al. 2009; lactose digestion: Tishkoff et al. 2007; ACE pesticide resistance in *Drosophila*: Karasov et al. 2010; trichome evolution in *Drosophila*: Sucena et al. 2003; butterfly mimicry: Reed et al. 2011; color vision in vertebrates: Yokoyama and Radlwimmer 2001; C-4 photosynthesis: Christin et al. 2008, 2009; cellulose digestion in primates: Zhang 2006; morphology of freshwater sticklebacks: Cresko et al. 2004; Shapiro et al. 2004; Colosimo et al. 2005; pigmentation: Protas et al. 2006; Arendt and Reznick 2008 and citations within; Manceau et al. 2010; reviews: Wood et al. 2005; Gompel and Prud'Homme 2009; Stern and Orgogozo 2009; Christin et al. 2010; Elmer and Meyer 2011; Losos 2011). Parallelism (evolutionary repeatability) is common within these case studies, which suggests that small- $n$  models of adaptation are generally plausible, even with respect to the evolution of genetically complex traits. Although additional work is surely required to explicitly link these observations to patterns of molecular evolution, the results of these studies tentatively suggest that  $K_A/K_X$  could be less sensitive to dominance than predicted by prior theory, even when most substitutions are drawn from new beneficial mutations.

### Variation of $N_e\mu n$ among Lineages with Different Effective Population Size

Although sensitivity to dominance may be dampened if individual bouts of substitution involve few competing genes (e.g., fig. 1), a positive correlation between  $K_A/K_X$  and  $E(h)$  is still expected as long as  $n > 1$ , on average. The qualitative predictions of prior theory (e.g., that beneficial substitutions are fixed at a higher rate on the X when  $E(h) < 1/2$  and more frequently on the autosomes when  $E(h) > 1/2$ ; Charlesworth et al. 1987) are therefore likely to remain applicable across much of the parameter space of  $n$ .

Nevertheless, the degree of sensitivity to dominance is also mediated by effective population size, which should generate some predictable lineage-specific patterns (see fig. 3).

Due to the increasing contribution of standing genetic variation to adaptation, sensitivity of  $K_A/K_X$  to dominance should partially break down with increasing population size (figs. 2 and 3). Empirical  $K_A/K_X$  ratios should therefore be particularly useful for estimating mean dominance in lineages of small size, assuming that the proportion of substitutions driven by positive selection can be accurately estimated (e.g., by using McDonald-Kreitman–based tests to estimate the rate of beneficial substitution; Eyre-Walker 2006). Similarly, X versus autosome substitution rate differences, where they exist, should be more pronounced in lineages with small historical  $N_e$  relative to lineages with large  $N_e$ . These population size effects may account (at least in part) for the unexplained variation of X versus autosome substitution rates, observed in different species. Faster-X evolution, which is more common in vertebrate compared with insect species (e.g., Vicoso and Charlesworth 2006; Pregraves 2008; Mank et al. 2010b; Orr 2010), might arise because  $N_e\nu$  is, on average, much lower in vertebrates than insects (due to the larger effective population size of the latter). Since effective population size can also play an important role in mediating the relative rate of slightly deleterious substitutions on the X and autosomes (Mank et al. 2010a, 2010b), population size variation may greatly impact total rates of substitution (i.e., those fixed by genetic drift or selection) between sex chromosomes and autosomes.

## Conclusion

The rapidly expanding body of population genomic data, coupled with statistical methods for characterizing adaptively driven molecular divergence (e.g., Fay et al. 2001; Smith and Eyre-Walker 2002; Sawyer et al. 2003; Eyre-Walker 2006; Welch 2006; Bazykin and Kondrashov 2011), should greatly improve current estimates of the relative rates of X versus autosome adaptive substitution across different lineages and gene functional categories. The theory presented here suggests that, even with accurate estimates of adaptive substitution, interpreting patterns of X-linked versus autosomal molecular adaptation should prove challenging, as these can be much more context dependent than has been emphasized by prior evolutionary theory. At the same time, the predicted variability for  $K_A/K_X$  ratios in gene- and lineage-specific contexts permits a much greater theoretical flexibility and may provide an improved match between models of molecular evolution and empirical substitution patterns.

## Supplementary Material

Supplementary figure S1 is available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We are grateful to members of the Clark lab for discussion, to two anonymous reviewers for very thoughtful comments and suggestions, and especially to Brian Charlesworth for

comments on an early version of the manuscript and for valuable correspondence about the conceptualization of DNA sequence adaptation. This work was supported by National Institutes of Health (grant GM64590 to A.G.C. and A.B.C.).

## Appendix

Relaxing assumptions on the relative probability of adaptation from SGV between X-linked and autosomal genes. In the main paper, it is assumed that, prior to positive selection, segregating variation evolved under strong purifying selection, such that  $s/s_d \ll 1$ . Under such conditions, and given a low mutation rate per locus, the probability of adaptation using standing variation simplifies to  $\sim 3N_e\nu/s_d$  and  $\sim 4N_e\nu/s_d$  for an X-linked and autosomal locus, respectively. The ratio of probabilities simplifies to  $\sim (3N_e\nu/s_d)/(4N_e\nu/s_d) = 3/4$ , as previously shown by Orr and Betancourt (2001). How might violation of these specified conditions alter these relative probabilities of adaptation from SGV? The most extreme violation of assumptions involves the probability of fixation for previously neutral alleles. Using the general approach of Hermisson and Pennings (2005), the probability of adaptation from previously neutral variation will be  $\Pr(\text{SGV}|\text{X locus}) \approx 1 - [1 + N_e s(1 + 2h)]^{-3N_e\nu}$  and  $\Pr(\text{SGV}|\text{autosome locus}) \approx 1 - [1 + 4N_e sh]^{-4N_e\nu}$  per locus. For  $N_e\nu$  small per locus, the ratio of probabilities reduces to  $\frac{\Pr(\text{SGV}|\text{X locus})}{\Pr(\text{SGV}|\text{autosome locus})} \approx \frac{3\ln[1 + N_e s(1 + 2h)]}{4\ln[1 + 4N_e sh]}$ . For strong selection on beneficial alleles (i.e.,  $N_e s > 100$ ), the ratio is generally within the range  $1 > X/A > 7/10$ , which is close the range predicted for alleles previously at mutation-selection balance.

## References

- Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol.* 23:26–32.
- Barton NH. 1989. The divergence of a polygenic system subject to stabilizing selection, mutation and drift. *Genet Res.* 54:59–77.
- Bazykin GA, Kondrashov AS. 2011. Detecting past positive selection through ongoing negative selection. *Genome Biol Evol.* 3:1006–1013.
- Bollback JP, Huelsenbeck JP. 2009. Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. *Genetics* 181:225–234.
- Bull JJ. 1983. Evolution of sex determination mechanisms. Menlo Park (CA): Benjamin/Cummings. p. 316.
- Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet.* 10:195–205.
- Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. *Am Nat.* 130:113–146.
- Christin PA, Salamin N, Muasya AM, Roalson EH, Russier F, Besnard G. 2008. Evolutionary switch and genetic convergence on rbcL following the evolution of C4 photosynthesis. *Mol Biol Evol.* 25:2361–2368.
- Christin PA, Samaritani E, Petitpierre B, Salami N, Besnard G. 2009. Evolutionary insights on C4 photosynthetic subtypes in grasses from genomics and phylogenetics. *Genome Biol Evol.* 1:221–230.



- Christin PA, Weinreich DM, Besnard G. 2010. Causes and evolutionary significance of genetic convergence. *Trends Genet.* 26:400–405.
- Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* 307:1928–1933.
- Cresko WA, Amores A, Wilson C, Murphy J, Currey M, Phillips P, Bell MA, Kimmel CB, Postlethwait JH. 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc Natl Acad Sci U S A.* 101:6050–6055.
- Ellegren H. 2011. Sex chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet.* 12:157–166.
- Elmer KR, Meyer A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol Evol.* 26:298–306.
- Eyre-Walker A. 2006. The genomic rate of adaptive evolution. *Trends Ecol Evol.* 21:569–575.
- Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. *Genetics* 158:1227–1234.
- Fisher RA. 1958. The genetical theory of natural selection. 2nd ed. New York: Dover Publications Inc. p. 291.
- Gillespie JH. 1984. Molecular evolution over the mutational landscape. *Evolution* 38:1116–1129.
- Gillespie JH. 1991. The causes of molecular evolution. New York: Oxford University Press. p. 352.
- Gillespie JH. 2004. Why  $k = 4N\mu$  is silly. In: Singh RS, Uyenoyama MK, editors. The evolution of population biology. Cambridge: Cambridge University Press. p. 178–192.
- Gompel N, Prud'Homme B. 2009. The causes of repeated genetic evolution. *Dev Biol.* 332:36–47.
- Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169:2335–2352.
- Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, Sella G, Przeworski M. 2011. Classic selective sweeps were rare in recent human evolution. *Science* 331:920–924.
- Hurst LD, Pal C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet.* 5:299–310.
- Karasov T, Messer PW, Petrov DA. 2010. Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet.* 6:e1000924.
- Kimura M. 1971. Theoretical foundation of population genetics at the molecular level. *Theor Popul Biol.* 2:174–208.
- Kimura M. 1979. Model of effectively neutral mutations in which selective constraint is incorporated. *Proc Natl Acad Sci U S A.* 76:3440–3444.
- Kimura M, Ohta T. 1971a. On the rate of molecular evolution. *J Mol Evol.* 1:1–17.
- Kimura M, Ohta T. 1971b. Protein polymorphism as a phase of molecular evolution. *Nature* 229:467–469.
- Kimura M, Ohta T. 1971c. Theoretical aspects of population genetics. Princeton (NJ): Princeton University Press. p. 219.
- Kirkpatrick M, Hall DW. 2004. Male-biased mutation, sex-linkage, and the rate of adaptive evolution. *Evolution* 58:437–440.
- Kopp M, Hermisson J. 2007. Adaptation of a quantitative trait to a moving optimum. *Am Nat.* 176:715–719.
- Kopp M, Hermisson J. 2009. The genetic basis of phenotypic adaptation II: the distribution of adaptive substitutions in the moving optimum model. *Genetics* 183:1453–1476.
- Lee JM, Sonnhammer ELL. 2003. Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res.* 13:875–882.
- Losos JB. 2011. Convergence, adaptation, and constraint. *Evolution* 65:1827–1840.
- Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE. 2010. Convergence in pigmentation at multiple levels: mutations, genes and function. *Philos Trans R Soc B.* 365:2439–2450.
- Mank JE. 2009. Sex chromosomes and the evolution of sexual dimorphism: lessons from the genome. *Am Nat.* 73:141–150.
- Mank JE, Nam K, Ellegren H. 2010a. Faster-Z evolution is predominantly due to genetic drift. *Mol Biol Evol.* 27:661–670.
- Mank JE, Vicoso B, Berlin S, Charlesworth B. 2010b. Effective population size and the faster-X effect: empirical results and their interpretation. *Evolution* 64:663–674.
- Michalak P. 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91:243–248.
- Orr HA. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52:935–949.
- Orr HA. 2002. The population genetics of adaptation: the adaptation of DNA sequences. *Evolution* 56:13117–21330.
- Orr HA. 2005a. The genetic theory of adaptation: a brief history. *Nat Rev Genet.* 6:119–127.
- Orr HA. 2005b. Theories of adaptation: what they do and don't say. *Genetica* 123:3–13.
- Orr HA. 2005c. The probability of parallel evolution. *Evolution* 59:216–220.
- Orr HA. 2010. The population genetics of beneficial mutations. *Philos Trans R Soc B.* 365:1195–1201.
- Orr HA, Betancourt AJ. 2001. Haldane's Sieve and adaptation from the standing genetic variation. *Genetics* 157:875–884.
- Presgraves DC. 2008. Sex chromosomes and speciation in *Drosophila*. *Trends Genet.* 24:336–343.
- Pritchard JK, Di Rienzo A. 2010. Adaptation: not by sweeps alone. *Nat Rev Genet.* 11:665–667.
- Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol.* 20:R208–R215.
- Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin CJ. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet.* 38:107–111.
- Reed RD, Papa R, Martin A, et al. (13 co-authors). 2011. *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* 333:1137–1141.
- Rokyta DR, Abdo Z, Wichman HA. 2009. The genetics of adaptation for eight microvirid bacteriophages. *J Mol Evol.* 69:229–239.
- Sattath S, Elyashiv E, Kolodny O, Rinott Y, Sella G. 2011. Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in *Drosophila simulans*. *PLoS Genet.* 7:e1001302.
- Sawyer SA, Kulathinal RJ, Bustamante CD, Hartl DL. 2003. Bayesian analysis suggests that most amino acid replacements in *Drosophila* are driven by positive selection. *J Mol Evol.* 57:S154–S164.
- Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428:717–723.
- Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415:1022–1024.
- Stern DL, Orgogozo V. 2009. Is genetic evolution predictable? *Science* 323:746–751.
- Sucena E, Delon I, Jones I, Payre F, Stern DL. 2003. Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature* 424:935–938.
- Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, et al. (19 co-authors). 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet.* 39:31–40.
- Unckless RL, Orr HA. 2009. The population genetics of adaptation: multiple substitutions on a smooth fitness landscape. *Genetics* 183:1079–1086.

- Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet.* 7:645–653.
- Vicoso B, Charlesworth B. 2009. Effective population size and the faster-X effect: an extended model. *Evolution* 63:2413–2426.
- Weinreich DM, Delaney NF, DePristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111–114.
- Welch JJ. 2006. Estimating the genomewide rate of adaptive protein evolution in *Drosophila*. *Genetics* 173:821–837.
- White MJD. 1973. *Animal cytology and evolution*. London: Cambridge University Press. p. 961.
- Wood TE, Burke JM, Rieseberg LH. 2005. Parallel genotypic adaptation: when evolution repeats itself. *Genetica* 123:157–170.
- Yokoyama S, Radlwimmer FB. 2001. The molecular genetics and evolution of red and green color vision in vertebrates. *Genetics* 158:1697–1710.
- Zhang J. 2006. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat Genet.* 38:819–823.