THE EFFECTS OF DIRECTIONAL EPISTASIS ON MOLECULAR EVOLUTION

BY

Qin Qin Gong

B. S., Shan Dong University, 1999

M. S., Graduate School of Academia Sinica (Beijing), 2003

Submitted to the Graduate Degree Program in Ecology and Evolutionary Biology
and the Faculty of the Graduate School of the University of Kansas
in Partial Fulfillment of the Requirements for
the Degree of Master of Arts

John K. Kelly

Co-Chair

Maria E. Orive

Co-Chair

Edward O. Wiley

Date approved 24 April 2009

The Thesis Committee for Qin Qin Gong certifies
that this is the approved version of the thesis:

THE EFFECTS OF DIRECTIONAL EPISTASIS
ON MOLECULAR EVOLUTION

Committee:

John K. Kelly

Co-Chair

Maria E. Orive

Co-Chair

Edward O. Wiley

Date approved 24 April 2009

ABSTRACT

Forward population genetic simulations are used to explore the evolution of a sequence of nucleotide sites subject to reversible mutation under selection, mutation, and drift. Three selection schemes are studied: synergistic, antagonistic, and multiplicative interactions among sites. Their respective effects on the level of nucleotide diversity, the pattern of linkage disequilibrium, and the allele frequency spectrum are determined. Surprisingly, none of these aspects are affected by directional epistasis when the overall strength of selection is held constant (where the equilibrium allele frequencies are equal). The equilibrium mean fitness does differ with selection regime, and is relatively higher with synergistic interactions while lower with antagonistic epistasis. These findings legitimate the application of many population genetic models assuming multiplicative selection when there are actually epistatic interactions among sites, and have important implications on the evolution of recombination.

ACKNOWLEDGEMENTS

I would like to first thank Dr. John K. Kelly and Dr. Maria E. Orive for their kindness and patience in mentoring my study at the University of Kansas, and their guidance in the research and comments on this thesis. I would also like to thank Professor Edward O. Wiley for his guidance as a member of my committee, Dr. Stuart Macdonald, Dr. Jennifer Gleason, Tara Marriage, and Dr. Liza Holeski for their comments on some parts of this thesis, and Jaime R. Keeler for her help as the graduate coordinator. Lastly I want to thank my family and friends for their support and love.

INTRODUCTION

Epistasis

In general, epistasis refers to the interaction between genes, but the exact meaning of it varies in different contexts. The term "epistasis" was first introduced into genetics by Bateson (1909) to describe the interaction between genes that distorts the standard Mendelian segregation ratios. Mendelian and molecular geneticists follow this strict classical sense of epistasis (compositional epistasis, Philips 2008) and measure the effects of allele substitution against a fixed, exactly defined genetic background. On the other hand, evolutionary and quantitative geneticists generally use epistasis in the sense of Fisher's "epistascy" to describe the statistical deviation from the additive combination of single-locus genotypes in their effects on a phenotype (Fisher 1918), where the average effect of allele substitution is measured against the population average genetic background. Compositional (classical) and statistical epistasis are in fact two different ways to manifest the molecular interactions between the products of genes (functional epistasis, Philips 2008) at the level of observable traits.

The underlying mechanism varies for different levels of epistatic interaction. The first level of genetic interaction is between nucleotide sites or codons within a protein-coding gene, which affect the structure, stability, and function of the RNA or protein molecule (Ortlund et al. 2007; Wilke and Adami 2001). The second

level of epistasis involves genetic regulatory sequences which initiate, promote, reduce, or inhibit the expression of other genetic components (Ringo 2004). The third level of epistatic interaction occurs between the enzymes and/or signal molecules controlling various biochemical pathways in metabolic and other physiological activities (Keightley 1996). Finally, epistatic interaction also occurs between phenotypes with regard to their fitness effects (Brodie 2000). Phillips (1998) reviewed various definitions of epistasis and encouraged molecular biologists to become more quantitative in their measures of genetic outcomes and evolutionary geneticists to become more mechanistic in their interpretations of evolutionary change to complete the unification of the classical and statistical views of epistasis "through the metaphor of the quantitative flow across a genetic network" (Philips 2008).

For mutations affecting quantitative traits, there are two different types of epistasis based on how the mutation effects are modified by the interactions between them. Directional epistasis, also called magnitude epistasis, refers to antagonistic (also called positive when the mutation effect is negative) or synergistic (also called negative when the mutation effect is negative) interactions, depending on whether mutational effects at different loci diminish or reinforce each other, respectively. More detailed classification of directional epistasis was discussed by Phillips et al. (2000). Another type of epistasis is called sign epistasis, where the direction (sign) of the fitness effect of a mutation depends on the

genotype at other loci, such that the mutation is beneficial on some genetic backgrounds and deleterious on others (Weinreich 2005).

Epistasis is a nearly universal component of the genetic architecture of most complex traits (Carlborg 2004; Holland 2007; Wade 2001) and plays critical roles in many important evolutionary processes. Epistasis causes a particular allele to have different fitness effects on different genetic backgrounds, which enables selection for different combinations of alleles or phenotypes. Such selection leads to the evolution of supergenes, linkage groups, and coadapted gene complex and promotes reproductive isolation and biodiversity (Wolf et al. 2000). Epistasis also affects mutation load (Butcher 1995; Kondrashow 1994), linkage disequilibrium (Barton 1995) and the fixation of mutations (Charlesworth et al. 1993; Kondrashov 1994). The existence or particular form of epistasis is proposed to be necessary in theories about many evolutionary processes, such as speciation (Carson and Templeton 1984; Wolf et al. 2000), the origin and maintenance of sex, recombination, and diploidy (reviewed by de Visser and Elena 2007, and Kouyos et al. 2007), and genetic canalization (Burch and Chao 2004; Rice 2000).

Molecular evolution

Although epistasis appears to be very common in nature and has profound evolutionary implications, it is ignored in most population genetic theories of molecular evolution. The neutral theory has been used as the null model of

molecular evolution (Kimura 1983). The effects of linked selective loci on patterns of neutral evolution and variation have been intensely investigated. Single weakly selected mutations have only negligible effect on the variation at linked neutral loci (Golding 1997; Neuhauser and Krone 1997; Przeworski et al. 1999) whereas selective sweeps of strongly beneficial mutations (Smith and Haigh 1974) and "back-ground selection" against deleterious alleles (Charlesworth et al. 1993) can reduce neutral variation in linked regions. On the other hand, the sequence variability is increased near a single locus under balancing selection (Kaplan et al. 1988; Strobeck 1983), and is elevated at sites that are physically between two sites under balancing selection (Kelly and Wade 2000).

The neutral assumption of mutations at the molecular level is not always the case in reality. There is evidence that the majority of sites in genes, or in the whole genome, are not completely neutral but actually under weak selection (Ohta 1992). For example, the unequal usage of synonymous codons observed in many organisms (Chiapello et al. 1998; Duret and Mouchiroud 1999; Sharp and Li 1986; Sharp et al. 1986; Shields et al. 1988; Stenico 1994) suggests weak selection at synonymous sites. Indirect evidence also shows that transcribed but untranslated regions of genes can experience selection even stronger than synonymous sites (Bauer and Aquadro 1997; Li and Graur 1991).

The dynamics of molecular evolution under the nearly neutral assumption has also been examined by many authors. The extent of codon bias and patterns of

molecular diversity have been analyzed with diffusion theory by Bulmer (1991) and McVean and Charlesworth (1999). When linkage among selected sites is tight, the associations between selected alleles and the genetic backgrounds on which they are found can reduce the efficacy of selection arises, which is known as the Hill-Robertson effect (Felsenstein 1974; Hill and Robertson 1966). McVean and Charlesworth (2000) examined the effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation in a simulation study. They showed that the reduced selection efficacy due to Hill-Robertson interference results in lower fixation probabilities for beneficial alleles and higher fixation probabilities for deleterious alleles, thus decreasing the extent of codon bias; the interference builds up negative linkage disequilibrium (LD) and significantly reduces nucleotide polymorphism and diversity; the interference slightly decreases the contribution to heterozygosity and time to fixation for beneficial alleles and has almost no effect on either property for deleterious mutations.

McVean and Charlesworth (2000) assumed multiplicative selection in their models. However, with advancements in molecular biology, data suggesting epistasis among mutations at the molecular level have accumulated substantially since the last decade. First, compensatory mutations have been widely identified both in fitness assay experiments (Crawford et al. 2007; Poon and Chao 2005) and from analyses of linkage disequilibrium among polymorphic nucleotide sites

(Kirby et al. 1995; Schaeffer and Miller 1993). On the other hand, by comparing the fitness of constructed mutants with known number of mutations, epistatic interactions between deleterious mutations have been directly assayed in various organisms (Table 1; also see review by Burch et al. 2003, and Sanjuan and Elena 2006). Finally, synergistic epistasis is believed to exist among a large number of mutations with small fitness effects in highly conserved non-coding regions in animals (Kryukov et al. 2005) as well as among weakly selected synonymous codons (Akashi 1995, 1996). Amidst all the new information available, the question arises as to how epistatic interaction among selected sites affects the dynamics of molecular evolution relative to that expected under the assumption of independence among selection at multiple sites.

Purpose

The main purpose of this study is to investigate how directional epistasis affects the pattern of molecular evolution. Specifically, I examine if synergistic or antagonistic epistasis between deleterious mutations generates a different allele frequency spectrum, nucleotide diversity, and pattern of linkage disequilibrium from what is predicted by a multiplicative model when the system is at mutation-selection-drift equilibrium. I also examine the interactions between epistasis and Hill-Robertson effects between weakly selected mutations.

There are several different models developed by other authors for the study of

epistasis. In an additive model or multiplicative model, the effect of a multi-locus genotype is the sum or the product of the contributions of each component locus respectively. Fisher described epistais as a deviation from the additive model of allelic effects (Fisher 1918). When fitness effects are considered, modeling epistasis as the deviation from the multiplicative model is preferred. This is because the absence of epistasis in this model guarantees that selection does not build up correlations among alleles at different loci as long as there is no such correlation in the initial population (Karlin 1975). A recently developed multi-linear model based on genetic measurement theory (Hansen and Wagner 2001a; Wagner et al. 1998) has been used to study the epistatic effects on mutation load (Hansen and Wagner 2001b) and the evolution of evolvability (Carter et al. 2005) and genetic architecture (Hansen 2006; Hermisson et al. 2003). Specific models with well defined interactions also have been constructed to study the epistasis related to particular traits whose underlying gene networks or physiological interactions are clearly understood (Clark 1991; Gibson 1996; Kacser and Burns 1981; Lenski et al. 1999; Mestl et al. 1995; Nijhout and Paulsen 1997; Szathmary 1993; Wagner 1994). However, these more realistic models can not be used to study general evolutionary consequences of epistasis because the underlying mechanism varies widely for different traits and different organisms.

In this study, I describe epistasis with a power function (e.g. Lenski et al. 1999) as the deviation from the multiplicative selection. I compare the results of

simulations with different parameter sets by holding the intensity of selection constant. The conclusions from this study have important implications on previous population genetic models about the behavior of weakly selected mutations and theories about many biological processes such as the evolution of codon bias and recombination.

METHOD AND MODELS

I constructed a forward population genetic simulation program written in the C language to examine the effects of epistasis on patterns of molecular evolution. I consider a constant size population of N haploid nucleotide sequences, each consisting of L sites. At each site, there are two alleles; and s is the selection coefficient against the deleterious allele. The overall fitness of individuals with m deleterious alleles was calculated using the power function

$$W_m = e^{-s \cdot m^{\beta}} . \quad (1)$$

Here, β describes the form and amount of epistasis: if $\beta = 1$, fitness effects are multiplicative over sites; there is synergistic epistasis among sites if $\beta > 1$ whereas $\beta < 1$ means antagonistic epistasis (Figure 1). Many authors have considered a quadratic function (Table 1)

$$W_m = e^{-\alpha m - b m^2} \quad (2)$$

where a is the selection coefficient and b measures epistasis ($b > 0$ for synergistic and $b < 0$ for antagonistic epistasis). However, equation (1) provides a better fit to empirical data when the number of mutation under concern is large (Lenski et al. 1999; Maisnier-Patin et al. 2005; You and Yin 2002).

Each simulation run starts with a population in which all sites are fixed for the beneficial allele. Generations are discrete and the sequence of events is selection, mutation and then recombination (if applicable). The probability of an individual

contributing to the next generation is proportional to its relative fitness as calculated from equation (1). Mutation occurs in both directions at the same frequency of μ per site per generation and can occur at currently segregating sites. Recombination occurs between adjacent sites at frequency r per generation with a maximum of one crossover event per sequence.

A series of statistics were calculated for samples taken from the population at mutation-selection-drift equilibrium. Each statistic is the average over samples from 4 runs. For each run, after an initial period of $4/\mu$ generations for the system to reach equilibrium, one sample of 25 alleles was taken without replacement every $2/\mu$ generations until 50 samples were collected. This sampling strategy follows that of McVean and Charlesworth (2000) to minimize evolutionary non-independence between sequential samples. In these simulations, the number of sites (L) was set at 1000; the population size was $N = 200$. Larger population sizes (250, 1000 and 2000) were also studied, but no differences were found in any case. All results appear to depend on the scaled parameters $N\mu$, Ns, and Nr, as expected from diffusion theory (Evens, 1979). The scaled mutation rate per site (both forward and backward) $2N\mu = 0.04$ and $2N\mu = 0.2$ were used because 0.04 is close to the estimated $4N_e\mu$ value for synonymous sites in *Drosophila. melanogaster* (Moriyama and Powell 1996) and 0.2 is close to the estimated $2N_e\mu$ value for synonymous sites in *E. coli* (Hartl et al. 1994).

RESULTS

Comparing results from different selection schemes

For a given value of s, the intensity of selection varies with different forms and magnitudes of epistasis (the value of β). Figure 2 shows how the equilibrium average frequency of deleterious alleles is affected by epistatic selection and recombination. As the selection coefficient increases, the equilibrium average frequency of deleterious alleles decreases, as expected. Due to the Hill-Robertson effect (Hill and Robertson 1968), higher recombination rates correspond to lower deleterious allele frequencies, which is consistent with McVean and Charlesworth's result (2000) in a multiplicative selection model (β =1). For the entire range of recombination rates, antagonistic epistasis retards the strength of selection while synergistic epistasis reinforces it. This effect of epistasis is the same in populations with $2N\mu = 0.04$ and those with $2N\mu = 0.2$. When making comparisons between the multiplicative and epistatic selection models, I need to hold the overall intensity of selection constant. To do so, for all subsequent comparisons between selection schemes, I contrast parameter sets that produce the same average deleterious allele frequency at equilibrium (Charlesworth et al. 1990). In this section, I will present comparisons by plotting the statistics against the average equilibrium frequency of deleterious alleles for runs with different values of β.

Nucleotide diversity and nucleotide polymorphism

The relationship between nucleotide diversity and the intensity of selection is complex. However, without mutation bias, nucleotide diversity decreases monotonically with increasing selection intensity (McVean and Charlesworth, 1999). The increase in selection intensity also results in a lower equilibrium frequency of deleterious alleles. Therefore, a lower equilibrium frequency of deleterious alleles corresponds to a lower level of nucleotide diversity, as shown in Figure 3. McVean and Charlesworth (2000) presented that for a given selection coefficient s, Hill-Robertson interference reduces both nucleotide diversity and selection intensity. Figure 3 shows that when the later is held the same, stronger Hill-Robertson interference ($2Nr = 0$) still corresponds to lower nucleotide diversity than weaker one ($2Nr = 1$). Figure 3 also shows that both synergistic and antagonistic selection have the same pattern as multiplicative selection in terms of the relationship between nucleotide diversity and equilibrium allele frequency. That is, regardless of the form and magnitude of epistasis, populations with a certain frequency of deleterious alleles at equilibrium have the same level of nucleotide diversity. This holds for all levels of recombination that I considered, implying that epistasis does not change the way Hill-Robertson interference affects nucleotide diversity. The same conclusion can be drawn for the effect of epistasis on nucleotide polymorphism (figure 4), and in populations with $2N\mu = 0.04$ and those with $2N\mu = 0.2$.

Patterns of linkage disequilibrium

The expected linkage disequilibrium (LD) between favored alleles at pairs of segregating sites is determined by the joint effect of mutation, selection, drift, and recombination. In finite populations, random drift creates both positive and negative LD at equal rates. If there is no fitness difference between the two alleles at every site, the average absolute magnitude of LD is not zero and is reduced by recombination, but the average of pairwise scaled LDs in close proximity is not significantly different from zero because positive and negative values cancel out on average. On the other hand, selection tends to erase the LD created by random drift. Since positive LD increases whereas negative LD decreases the efficiency of selection, selection reduces positive LD faster than it does negative LD. Thus, drift and selection together generate a tendency towards negative LD and stronger selection results in more negative LD (Hill and Robertson 1966). Negative LD is caused by excess repulsion between preferred codons, which means the ratio of observed variance ($\hat{\sigma}^2$) in the number of beneficial alleles among individuals to the sum of the variance predicted by allele frequency (x_i) at every site, RV= $\hat{\sigma}^2 / \sum_i x_i(1-x_i)$, should be less than 1 and the more negative LD, the more dearth in RV relative to 1 (McVean and Charlesworth 2000).

The average D′ (measures scaled LD, Lewontin 1964) and average r^2 (measures magnitude of LD, Hill and Robertson 1968) for pairs of segregating sites in close proximity are calculated for simulations with a series of

combinations of parameters. For example, the results in r^2 as the function of distance between segregating sites along the sequence for $2N\mu = 0.04$, $2Ns = 0$, $\beta = 1$ and a range of $2Nr$ values are presented in figure 5a. Figure 5b also shows the effect of selection on r^2. For a certain value of $2Nr$ and a series of values of $2Ns$, the average pairwise r^2 for segregating sites within each distance category is plotted against equilibrium frequency of deleterious alleles; and I put such plotting for different values of β together to compare the effect of epistasis on the average value of r^2 for each distance category. The same method is used to compare the effect of epistasis on D′. Figure 6 presents examples of such comparisons for two distance categories with $2N\mu = 0.04$, $2Nr = 0$, which shows that, given the same equilibrium allele frequency, no difference in either of the two measurement of LD are caused by different selection schemes. Similar patterns are observed in other distance categories and all other parameter combinations.

The comparison on RV is less complicated but the plots for different values of β separate slightly from each other (figure 7). However, since the distribution of the statistic RV has high variance (McVean and Charlesworth 2000), it is still safe to say that no significant difference in RV value is caused by different epistatic section as long as the equilibrium allele frequencies are held the same.

Overall, these results indicate that epistasis has no effect on how selection, drift and recombination determine the linkage disequilibrium between preferred alleles.

The effect of epistasis on the equilibrium mean fitness

The mean fitness of a population at selection-mutation-drift equilibrium is determined by the proportion of deleterious alleles kept in the population combined with the fitness effect of each allele as well as the way they interact with each other (epistasis). Stronger selection results in a lower equilibrium frequency of deleterious alleles (Figure 2). When the positive effect of the reduction in the frequency of the deleterious allele can not compensate for the negative effect of the increase in selection coefficient in determining population mean fitness, increasing s leads to the drop of equilibrium mean fitness. Stronger selection can eliminate deleterious alleles more efficiently, so the population mean fitness at equilibrium could become larger as s increases. Because tight linkage reduces the efficiency of selection, resulting in a higher frequency of deleterious alleles at equilibrium, a larger recombination rate leads to a higher equilibrium mean fitness for a given s (McVean and Charlesworth 2000). Figure 8 shows that, for a given equilibrium frequency of deleterious alleles, the population with frequent recombination ($2Nr = 1$) still has a higher mean fitness than that with no recombination ($2Nr = 0$). With everything else the same, synergistic epistasis ($\beta = 1.2$) leads to a higher mean fitness than multiplicative selection, while antagonistic epistasis ($\beta = 0.8$) leads to a lower one. Such pattern of the effect of epistasis on the population mean fitness at equilibrium can be observed both in populations with $2N\mu = 0.04$ and those with $2N\mu = 0.2$.

Estimating the strength of selection

If selection acts independently between all sites, for small s and small $2N_e\mu$, the expected proportion (x) of sites fixed for the beneficial allele in the population at mutation-selection-drift equilibrium is

$$E(x) = \frac{1}{1 + e^{-2N_e s}} \tag{3}$$

(Li 1987; Bulmer 1991), where N_e is the effective population size. Thus I can estimate the value of $2N_e s$ from the proportion of sites fixed for the beneficial allele in a sample taken from a population at equilibrium by rearranging equation (3) to obtain

$$2N_e s = \ln \frac{x}{1 - x} \tag{4}$$

(Bulmer 1991; McVean and Charlesworth 2000). On the other hand, the proportion of segregating sites is usually so low that they have little effect on the overall frequency of alternative alleles at equilibrium; and the relationship between the expected overall frequency of the preferred alleles and the selection coefficient is similar for fixed and segregating sites. Thus, it is safe to say that equation (3) is also an accurate approximation for the expected overall frequency of beneficial alleles in a single sequence picked at random from the population (McVean and Charlesworth 1999). In other words, $2N_e s$ can also be estimated from the overall frequency of beneficial alleles in a sample of sequences in the same way as shown in equation (4) after reassigning x to be the overall allele frequency of beneficial alleles.

For weak selection ($2N_es << 1$), the estimates of $2N_es$ from the overall average allele frequency and those from fixed sites alone are very close to each other as well as to the true value of *2Ns*. As the selection coefficient increases, due to the Hill-Robertson effect, the true value of *2Ns* becomes more and more underestimated by both methods for tight linked sites; and recombination reduces both underestimates. Increasing the selection coefficient also enlarges the discrepancy between the allele proportions among fixed sites and the overall allele frequencies for the whole sequence (McVean and Charlesworth 1999), resulting in more and more difference between the two estimates of $2N_es$; and recombination magnifies this difference (McVean and Charlesworth 2000).

When selection is epistatic, an equation similar to equation (4)

$$2N_e s_e = \ln \frac{x}{1-x} \tag{5}$$

can be used to estimate $2N_es_e$, where s_e is defined as the effective selection coefficient for the epistatic model such that a multiplicative model with selection coefficient s_e generates the same overall allele frequencies with this epistatic model. When epistasis is zero ($\beta = 1$), $2N_es_e = 2N_es$.

Figure 9 shows that epistasis has no effect on the difference between the two estimates, which means, for a given overall equilibrium allele frequency, the allele frequency among fixed sites is the same for different epistatic selection schemes. Since epistasis does not change the relative proportion of polymorphic and fixed sites either (Figure 3, 4), the allele frequencies for segregating sites at equilibrium

must be independent of selection schemes too. This result holds for populations with $2N\mu = 0.04$ and those with $2N\mu = 0.2$.

DISCUSSION

In a finite population with reversible mutation, directional epistasis modifies the average strength of selection. After adjusting for this effect however, standard molecular population genetic statistics are unaffected by epistasis. When a population reaches statistical equilibrium, the nucleotide diversity and polymorphism, the pattern of linkage disequilibrium, the variance in the number of deleterious mutations, and the allele frequency among fixed and segregating sites, are all the same between epistatic and multiplicative selection models. The only difference caused by different selection schemes is that synergistic selection leads to a higher while antagonistic selection results in a lower population mean fitness than multiplicative selection.

With the identical patterns of the statistics between populations at equilibrium with different selection schemes shown in this study, it may be appropriate to release the assumption of independence of selection at multiple sites for many population genetic models. Consider a hypothetical population that is at mutation-selection-drift equilibrium under epistatic selection with selection coefficient s and has an average frequency p for deleterious alleles. Another population with multiplicative selection governed by s_e (effective selection coefficient) has the same equilibrium frequency of deleterious alleles. All the population statistics except equilibrium mean fitness are identical between these

two populations. Because epistatic selection models are much more difficult for mathematical analysis, it is easier to study a multiplicative selection model with selection coefficient s_e and obtain valid implications to the epistatic selection models with selection coefficient s. For example, both the maximum likelihood method with Poisson random field model employed to estimate $2N_e\mu$ and $2N_es$ using the pattern of nucleotide polymorphisms (Hartl et al. 1994) and the selection-mutation-drift model developed to study the selection intensity on codon usage ignored epistatic selection in the evolutionary process. However if there is no difference between multiplicative selection and epistatic selection in terms of the standard molecular population genetic statistics, it is appropriate to apply these models in analyzing empirical data even when selections are synergistic in reality.

Codon bias, the unequal usage of synonymous codons, is one of the most intensely examined genetic phenomena involving weak selection at multiple sites. The degree of codon bias can be interpreted as the result of the combined effect of selection, mutation bias and drift. A number of population genetic models have been developed to study the effect of these factors in determining the level of codon bias (Akashi 1994; Bulmer 1991; Duret 2002; Li 1987; McVean and Charlesworth 1999). Synergistic selection at synonymous sites is plausible for several reasons. One difficulty for the non-epistatic weak selection model in explaining high proportions of preferred codons in genes is the high mutation load (Gillespie 1994; Li 1987; Tachida 1990). This difficulty can be overcome by

adopting synergistic selection to reduce the mutation load (Kimura and Maruyama 1966; Kondrashov 1995). Figure 8 confirms this point. Also empirical estimates of selection coefficient for optimal synonymous codons show a much wider range than that required by multiplicative and additive models (Akashi 1995, 1996); and synergistic epistasis is suggested to account for the observed codon bias range with a reasonable level of selection coefficient variation (Akashi 1995; Kondrashov et al. 2006). Despite the important roles epistasis plays in interpreting codon bias patterns, due to the mathematical difficulty, it is rarely included in the population genetic models of synonymous codon usage. Li (1987) studied the predicted patterns of codon usage in an additive, a multiplicative and an additive with a threshold model and found that the three models produce identical results, based on which he proposed that synergistic selection would generate similar results. The results from our simulations support Li's generalization and legitimate the application of previous genetic models in interpreting empirical data of codon usage even when synergistic selection is the case.

Epistasis plays an important role in the theoretical explanation of the origin and evolutionary benefit of sex and recombination (reviewed by de Visser and Elena 2007, and Kouyos et al. 2007). The most widely studied theories on the benefit of recombination are variation-and-selection models (reviewed by Kondrashov 1993). These models assume that recombination facilitates the population's response to directional selection by increasing useful genetic

variation, thus speeding up adaptation and promoting the elimination of deleterious mutations. For this mechanism to work, a key prerequisite is that the LD between preferred alleles among sites should be negative. There are two possible sources of negative LD; one is directional selection following the effects of drift (Hill-Robertson effect); and the other is negative epistasis. If negative epistasis is a major contributor of the benefit of recombination, it should be common in the genome of organisms where recombination occurs. However, recent experimental data shows that this is not the case (Burch et al. 2003; Sanjuan and Elena 2006). The direct assays of epistasis by fitness measurement experiments in various organisms show a mixed picture: some show prevailing synergism, some are overall antagonistic, and several show no or variable epistasis. On the other hand, although it has been analytically shown that, in an infinite population, positive epistasis generates positive LD and negative epistasis generates negative LD (Eshel and Feldman 1970), for finite populations, the effect of directional epistasis in generating negative LD is proposed to be negligible relative to that of Hill-Robertson interference according to several simulation studies comparing the relative importance of Hill-Robertson effect and negative epistasis in the evolution of the recombination modifier (Iles et al. 2003; Keightley and Otto 2006; Otto and Barton 2001). The results from our simulations show no detectable extra negative LD caused by synergistic epistasis relative to the Hill-Robertson effect. This adds another challenge to theories of the evolution

of recombination based on the effect of negative epistasis in generating negative LD.

SUMMARY AND FUTURE DIRECTIONS

I studied the effects of directional epistasis on patterns of various population molecular genetic statistics in finite populations with reversible mutation. Their implications to population genetic models of codon usage and the evolution of recombination are discussed.

Although the constant directional epistatic selection model I employ here is a better approximation of the real interaction among weakly selected sites such as codon usage and some non-coding sequences than the multiplicative model, it is not applicable to explaining the behavior of mutations with variable mutational effects and epistatic interactions. The distribution of fitness effects of mutations has been studied by many authors in various organisms (reviewed by Eyre-Walker and Keightley 2007). The general pattern for all species and genomic regions is: advantageous mutations are rare; strongly beneficial mutations are exponentially distributed; and the effect distribution of deleterious mutations is complex and multi-modal. Previous studies also show that the sign and multitude of epistasis vary largely among different combinations of mutations (Sanjuan and Elena 2006). No specific distribution of the strength of epistasis or its form has been recognized yet. However, it was reported that the strength of directional epistasis is correlated with the average deleterious effect of a single mutation (Wilke and Adami 2001).

To provide more useful implications in explaining the behavior of molecular

evolution, more elaborated simulations using models with non-uniform distribution of selection effects and variable epistasis coefficients are necessary. A different mathematic equation need to be devised to describe epistatic interactions among selections at multiple sites when a non-uniform distribution of selection effects is adopted in the model. Other types of models in which both pairwise and higher orders of epistasis are specifically defined throughout the sequence are more desirable when the variation in the strength and form of epistasis is considered (Hansen and Wagner 2001a).

Table 1. The magnitude of fitness effect and epistasis of mutations in previous studied organisms.

Organism	*m*	Best fit quadratic model (a, b)	Best fit power model (s, β)
Bacteriophage Φ6 (Burch and Chao 2004)	5	$a = 0.0562$, $b = -0.0038$	$s = 0.0562$, $\beta = 0.7436$
FMDV (Elena 1999)	6	$a = 0.0950 \pm 0.0317$, $b = 0.0109 \pm 0.0298$	$s = 0.0950$, $\beta = 1.2923$
Salmonella typhimurium (Maisnier-Patin *et al.* 2005)	> 50		$s = 0.041 \pm 0.0054$, $\beta = 0.46 \pm 0.038$
E. coli (Elena and Lenski 1997)	3	$a = 0.0273 \pm 0.0059$, $b = 0.0037 \pm 0.0070$	$s = 0.0273$, $\beta = 1.3106$
Aspergillus niger (De Visser *et al.* 1997)	6	$a = 0.18045$, $b = -0.01440$	$s = 0.18045$, $\beta = 0.63632$
D.melanogaster (Kitagawa 1967; Mukai 1969; Seager *et al*, 1982)	8.5	$a = 0.009813$, $b = 0.005550$	$s = 009813$, $\beta = 1.8219992$

The best fit quadratic models are originally reported by the authors; the best fit power models are approximately converted from the quadratic models except that the result for *Salmonella typhimurium* is original. See equation (1) and (2) in the Modeling section of the text for the detail of the quadratic and power models.

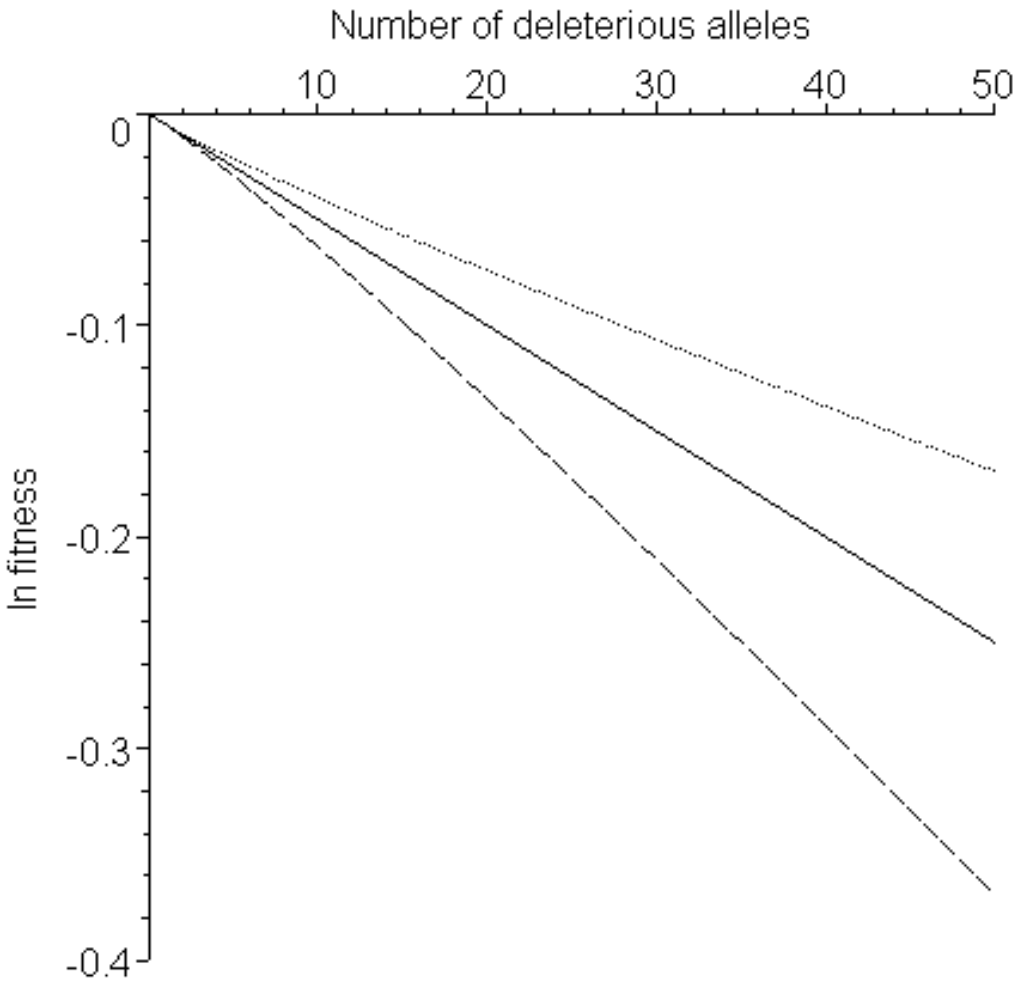


Figure 1. Hypothetical effects of increasing numbers of deleterious alleles on ln fitness in different selection models. The solid line illustrates multiplicative effects; the dotted curve shows antagonistic epistasis; and the dashed curve represents synergistic epistasis.

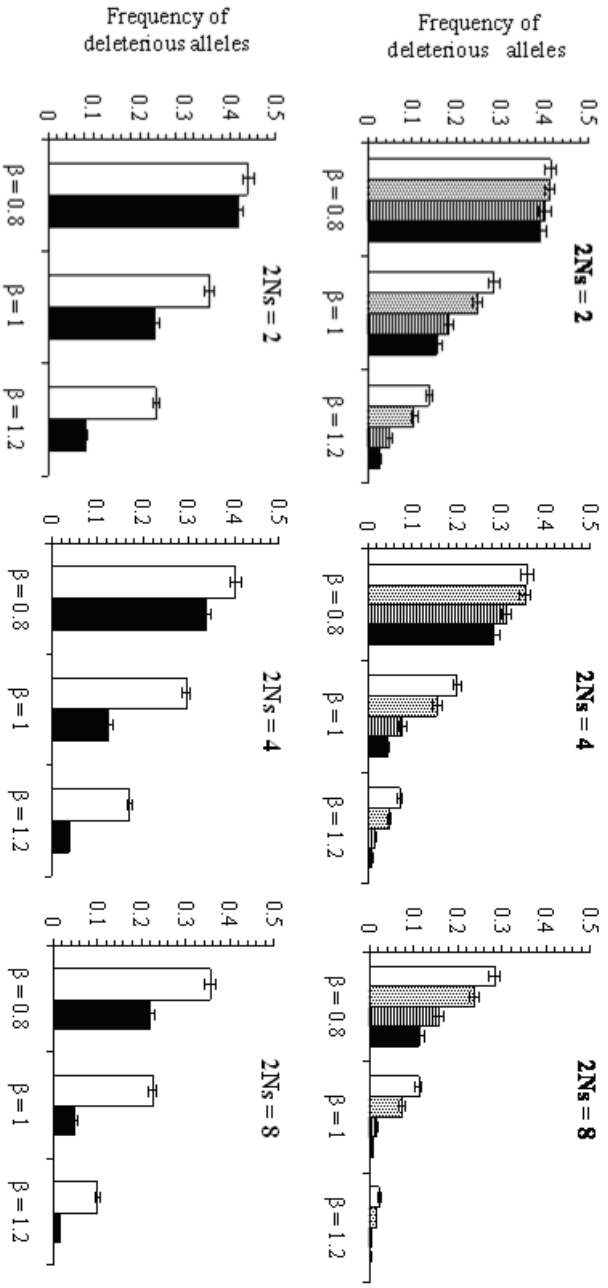


Figure 2. The effect of epistatic selection and recombination on the average frequency of deleterious alleles at equilibrium (▫: $2Nr = 0$; ▫: $2Nr = 0.01$; ▫: $2Nr = 0.1$; ▪: $2Nr = 1$). Above: $2N\mu = 0.04$; below: $2N\mu = 0.2$.

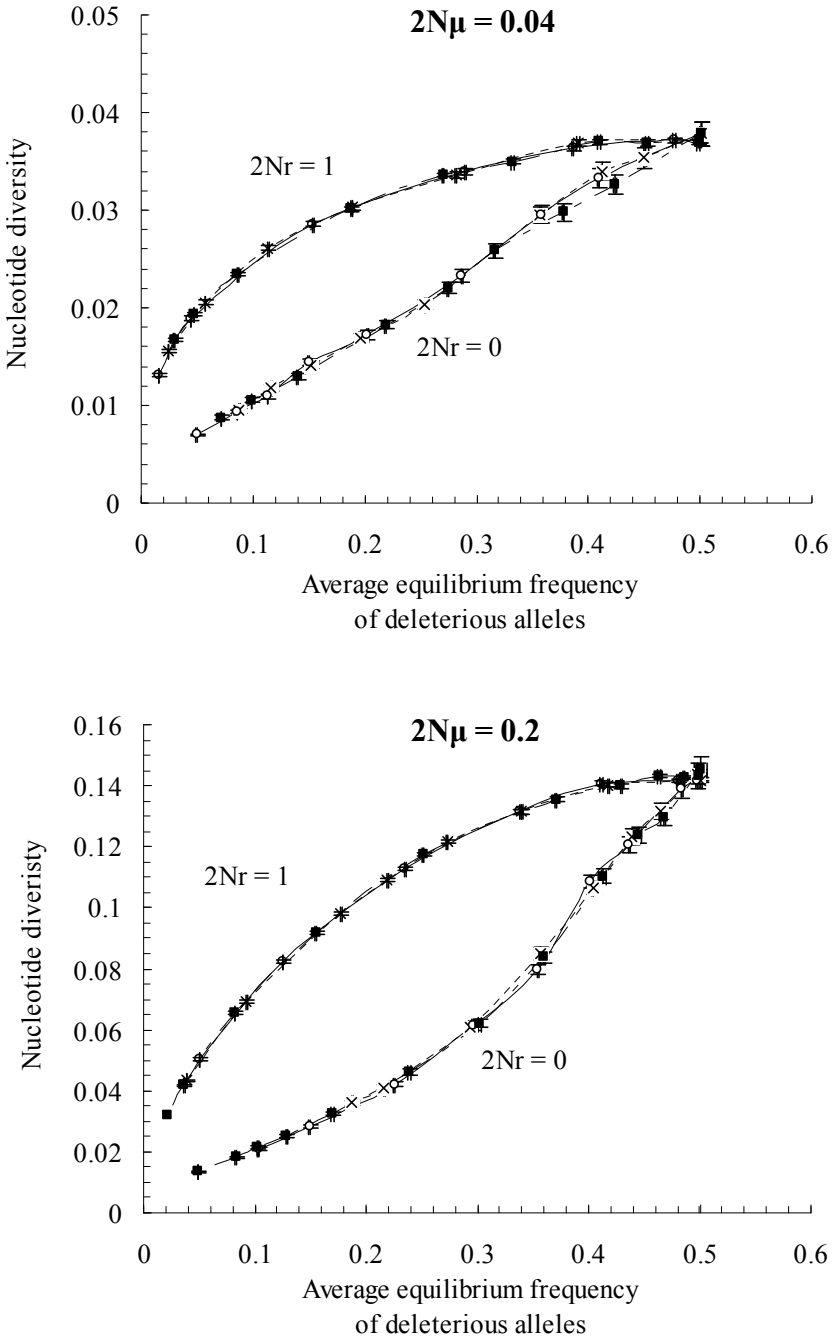


Figure 3. Nucleotide diversity under different selection scenarios. In each graph, the two groups of plots are for the cases with no recombination and with $2Nr = 1$ respectively. Within each group, different symbols represent different forms and magnitudes of epistasis (– ■ –: $\beta = 1.2$; —○—: $\beta = 1$; - - ×· - -: $\beta = 0.8$).

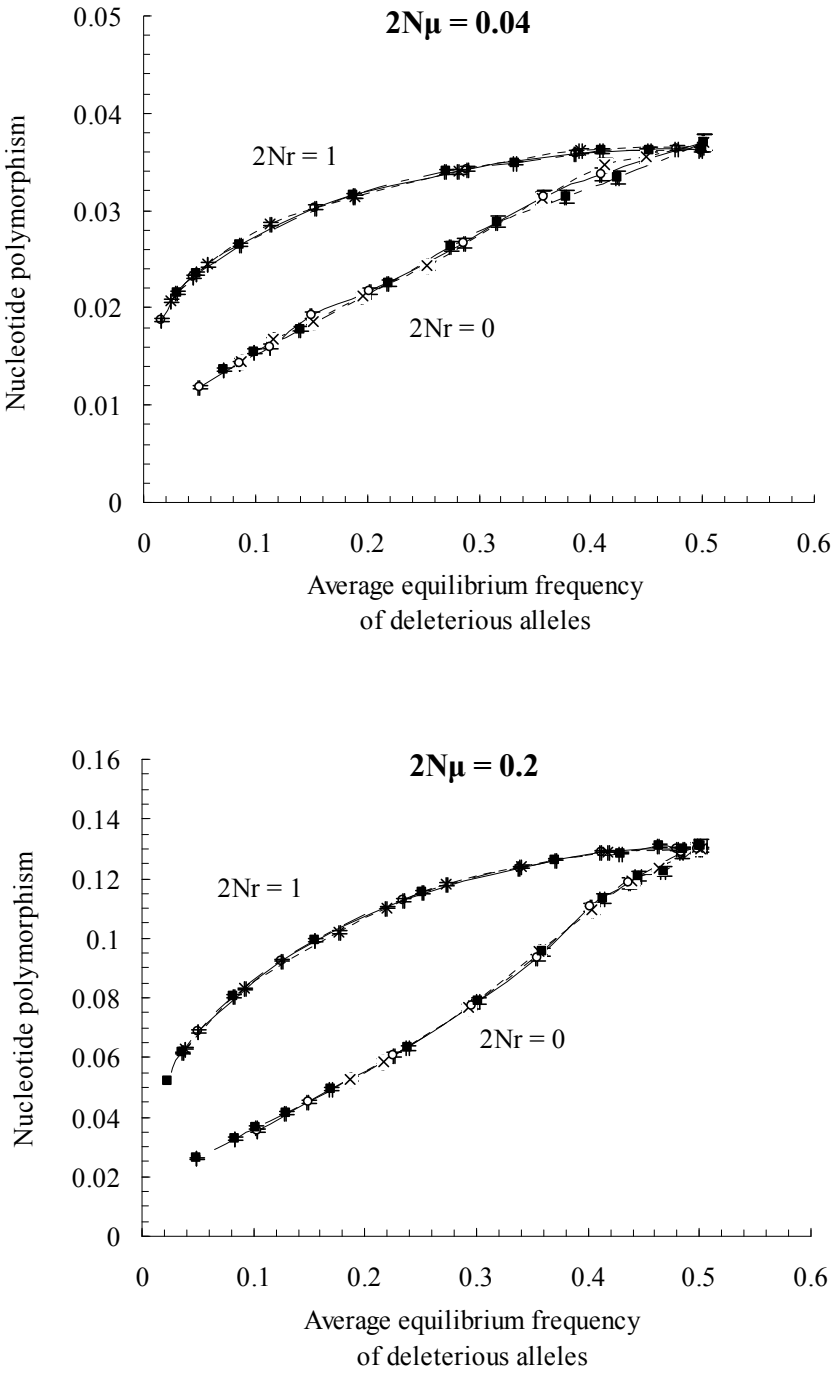


Figure 4. Nucleotide polymorphism under different selection scenarios. In each graph, the two groups of plots are for the cases with no recombination and with $2Nr = 1$ respectively. Within each group, different symbols represent different forms and magnitudes of epistasis (– ■ –: $\beta = 1.2$; —○—: $\beta = 1$; - - -×- - -: $\beta = 0.8$).

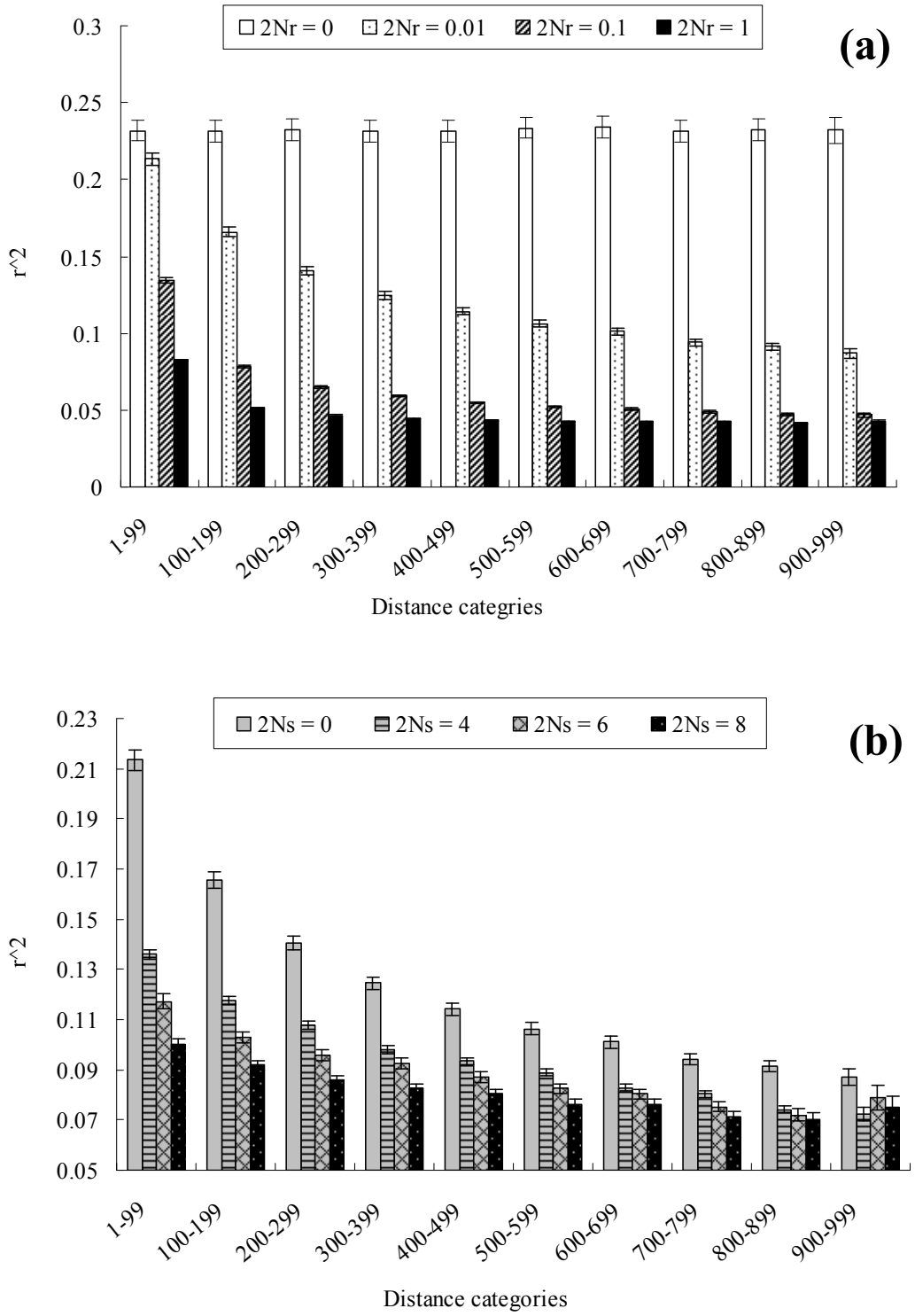


Figure 5. Average pairwise r^2 in a sample of 25 alleles as a function of the distance between segregating sites for 1000 sites, showing the effect of recombination ((a), $2Ns = 0$) and selection intensity ((b), $2Nr = 0.01$). $2N\mu = 0.04$, $\beta = 1$.

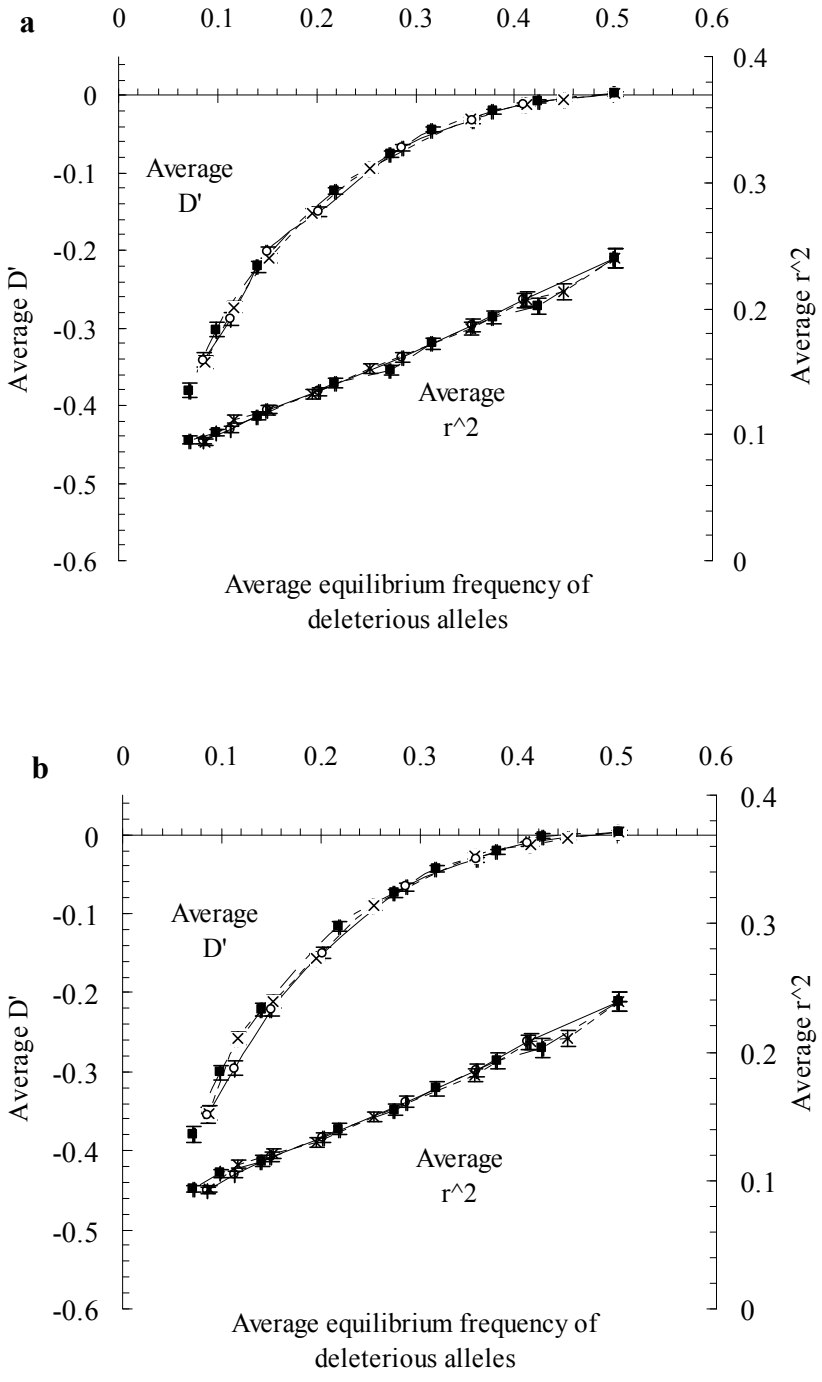
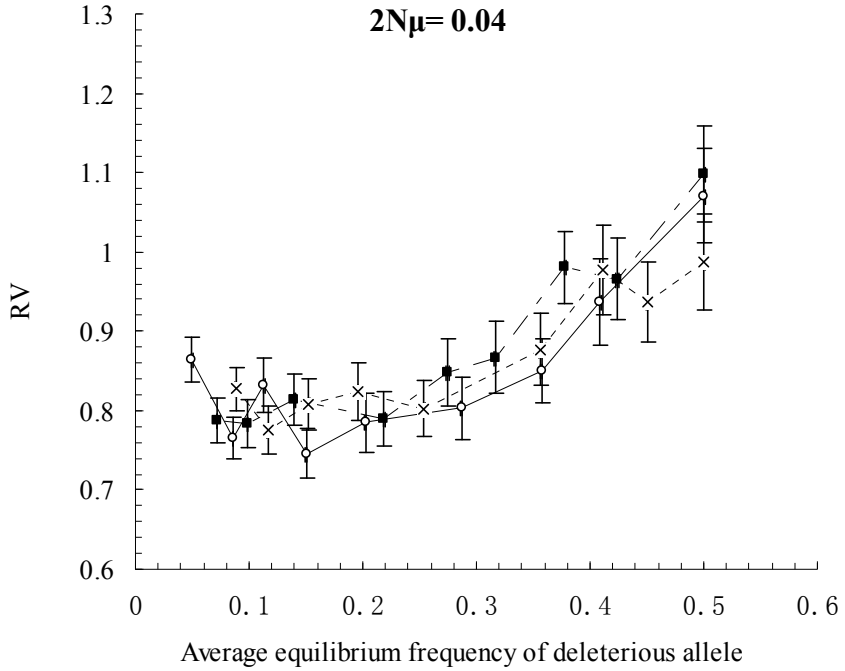


Figure 6. The effect of epistasis (– ■ –: $\beta = 1.2$; —○—: $\beta = 1$; - - -×- - -: $\beta = 0.8$) on two measures of LD, showing average measurements for pairs of segregating sites within the distance of 0-99 (a) and 500-599 sites (b). $2N\mu = 0.04$; $2Nr = 0$.



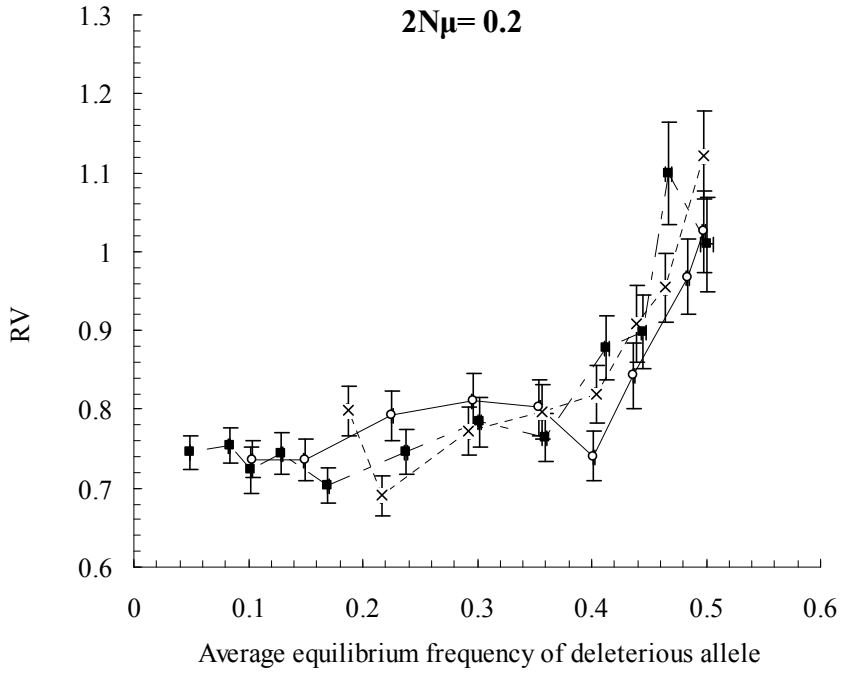


Figure 7. The influence of epistasis (– ■ –: β = 1.2; —○—: β = 1; - - -×- - -: β = 0.8) on the ratio of observed-to-expected variance in the number of preferred alleles among individuals (RV). $2Nr$ = 0.

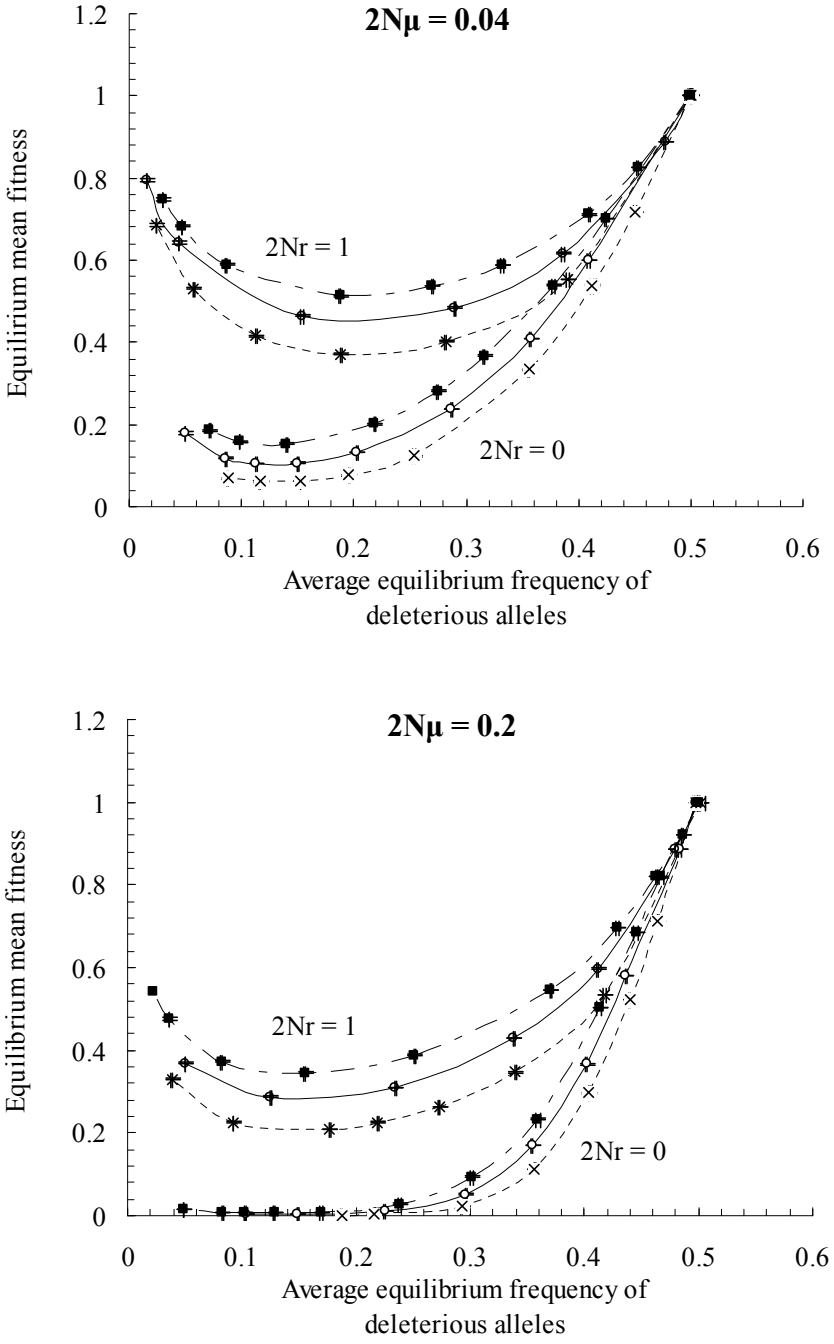


Figure 8. The effect of epistasis (– ■ –: $\beta = 1.2$; —○—: $\beta = 1$; - - -×- - -: $\beta = 0.8$) and recombination on the population mean fitness at equilibrium.

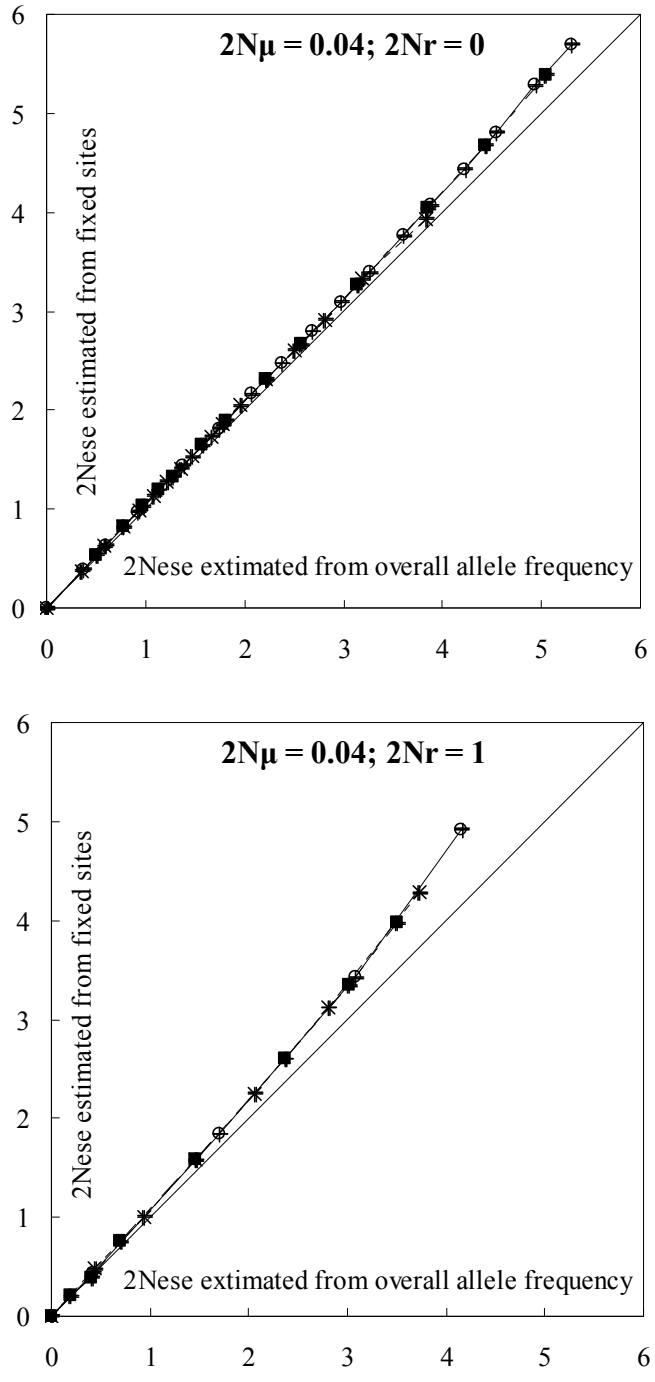


Figure 9. (to be continued)

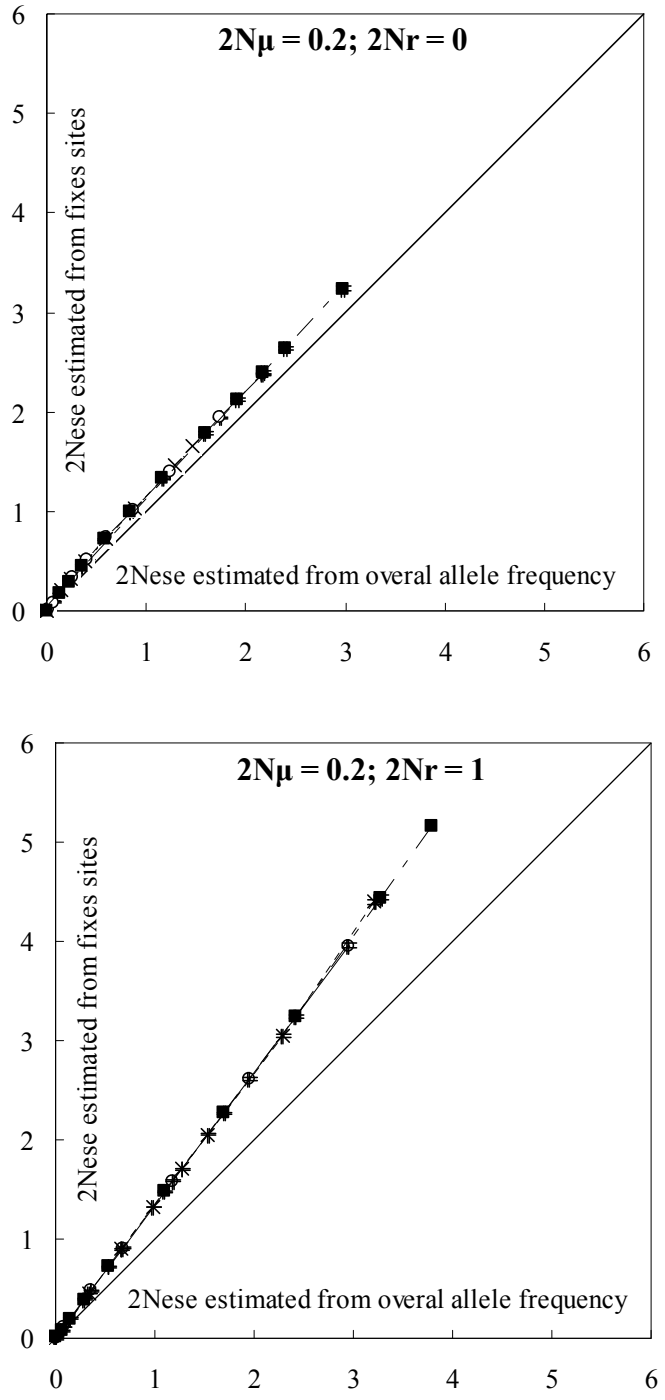


Figure 9. The comparison of two estimated $2N_e s_e$,

showing the effect of recombination and epistasis (– ■ –: $\beta = 1.2$; —○—: $\beta = 1$;

- - -×- - -: $\beta = 0.8$). Every point represents two estimates for a same value of s.

REFERENCES

Akashi, H. 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* 136: 927-935.

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.

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

Bateson, W. 1909. *Mendel's Principles of Heredity.* Cambridge Univ. Cambridge, MA.

Bauer, V. L., and Aquadro, C. F. 1997. Rates of DNA sequence evolution are not sex-biased in *Drosophila melanogaster* and *D. simulans*. *Mol. Biol. Evol*. 14: 1252-1257.

Brodie, E. D. III 2000. Why evolutionary genetics does not always add up. Pages 3-19 in Wolf, J. B., et al., eds., *Epistasis and the evolutionary process*. Oxford University Press, Oxford.

Bulmer, M. G. 1991. The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129: 897-907.

Burch, C. L., and Chao, L. 2004. Epistasis and its relationship to canalization in

the RNA virus Φ 6. *Genetics* 167: 559-567.

Burch, C. L., Hanley, K. A., and Turner, P. E. 2003. Patterns of epistasis in RNA viruses: a review of the evidence from vaccine design. *J. Evol. Biol.* 16: 1223-1235.

Butcher, D. 1995. Muller's ratchet, epistasis and mutation effects. Genetics 141:431-437.

Carlborg, O., and Haley, C. S. 2004. Epistasis: too often neglected in complex trait studies? *Nature Rev. Genet.* 5:618-625.

Carson, H. L., and Templeton, A. R. 1984. Gnetic revolutions in relation to speciation phenomenon: the founding of new populations. *Annual Review of Ecology and Systematics* 15: 97-131

Carter, A. J., Hermisson, J., and Hansen, T. F. 2005. The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theor. Popul. Biol.* 68: 179–196.

Charlesworth, D., Morgan, M. T., and Charlesworth, B. 1990. Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* 44: 1469-1489.

Charlesworth, B., Morgan, M. T., and Charlesworth, D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134: 1289-1303.

Chiapello, H., Lisacek, F., Caboche, M., and Henaut, A. 1998. Codon usage and

gene function are related in sequences of *Arabidopsis thaliana*. *Gene* 209: GC1-GC38.

Clark, A. G. 1991. Mutation-selection balance and metabolic control theory. *Genetics* 129: 909-923.

Crawford, H., et al. 2007. Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *Virology* 81(15): 8346-51.

De Visser, J. A. G. M., and Elena, S. F. 2007. The evolution of sex: empirical insights into the roles of epistasis and drift. *Genetics* 8: 139-149.

De Visser, J. A. G. M., Hoekstra, R. F., and van den Ende, H. 1997. Test of interaction between genetic markers that affect fitness in *Aspergillus niger*. *Evolution* 51: 1499-1505.

Duret, L. 2002. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* 12: 640-649.

Duret, L., And Mouchiroud, D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis, Drosophila,* and *Arabidopsis*. *Proc. Natl. Acad. Sci.* USA 96: 4482-4487.

Elena, S. F., and Lenski, R. E. 1997. Test of synergistic interactions among deleterious mutations in bacteria. *Nature* 390: 395-398.

Elena, S. F. 1999. Little evidence for synergism among deleterious mutations in a

nonsegmented RNA virus. *Mol. Evol.* 49: 703-707.

Eshel, I., and Feldman, M. W. 1970 On the evolutionary effect of recombination. *Theor. Popul. Biol*. 1: 88-100.

Evens, W. J. 1979. *Mathematical population genetics. Biomathematics*, Vol. 9. Springer-Verlag, Berlin.

Eyre-Walker, A., and Keightley, P. D. 2007. The distribution of fitness effects of new mutations. *Nature Reviews Genetics* 8, 610-618.

Felsenstein, J. 1974. The evolutionary advantage of recombination. *Genetics* 78: 737-756.

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Philosophical Transactions of the Royal Society of Edinburgh* 52: 399–433.

Gibson, G. 1996. Epistasis and pleiotropy as natural properties of transcriptional regulation. *Theor. Popul. Biol.* 49:58-59.

Gillespie, J. H. 1994. Substitutional processes in molecular evolution. III. Deleterious alleles. *Genetics* 138: 943-952.

Golding, G. B. 1997. The effect of purifying selection on genealogies. pp. 271–285 in *Progress in Population Genetics and Human Evolution*, edited by P. Donnelly and S. Tavare. Springer-Verlag, NewYork.

Hansen, T. F., 2006. The evolution of genetic architecture. *Annu. Rev. Ecol. Evol. Syst*. 37: 123–157

Hansen, T. F., and Wagner, G. P. 2001a. Modeling genetic architecture: a multilinear theory of gene interaction. *Theor. Popul. Boil.* 59: 61-86.

Hansen, T. F., and Wagner, G. P. 2001b. Epistasis and the mutation load: a measurement theory approach. *Genetics* 158: 477-485

Hartl, D. L., Moriyama, E. N., and Sawyer, S. A. 1994. Selection Intensity for codon bias. *Genetics* 138: 227-234.

Hermisson, J., T. F. Hansen and Wagner, G. P. 2003. Epistasis in polygenic traits and the evolution of genetic architecture under stabilizing selection. *Am. Nat*. 161: 708–734

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

Hill, W. G., and Robertson, A. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38: 226-231.

Holland, J. B. 2007. Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.* 10:156-161.

Iles, M. M., Walters, K., and Cannings, C. 2003. Recombination can evolve in large finite populations given selection on sufficient loci. *Genetics* 165: 333-337.

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

Kaplan, N. L., Darden, T., and Hudson, R. R. 1988. The coalescent process in

models with selection. *Genetics* 120: 819-829.

Karlin, S. 1975.General two locus selection models: some objectives, results and interpretations. *Theoret. Popul. Boil.* 7:364398

Keightley, P. D. 1996. Nature of deleterious mutation load in *Drosophila. Genetics* 144: 1993-1999.

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

Kelly, J. K., and Wade, M. J. 2000. Molecular evolution near a two-locus balanced polymorphism. *Theor. Biol.* 204: 83-101.

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

Kimura, M., and Maruyama, T. 1966. the mutational load with epistatic gene interactions in fitness. *Genetics* 54: 1337-1351.

Kirby, D. A., Muse, S. V., and Stephan, W. 1995. Maintenance of pre-mRNA secondary structure by epistatic selection. *Proc. Nat. Acad. Sci.* 92: 9047-9051.

Kitagawa, O. 1967. Interaction in fitness between lethal genes in heterozygous condition in *Drosophila melanogaster. Genetics* 57: 809-820.

Kondrashov, A. F., Ogurtsov, A. Y., and Kondrashov, A. S. 2006. Selection in favor of nucleotides G and C diversifies evolution rates and levels of polymorphism at mammalian synonymous sites. *Theor. Boil.* 240: 616-626.

Kondrashov, A. S. 1993. Classification of hypotheses on the advantage of amphimixis. *Heredity* 84: 372-387.

Kondrashov, A. S. 1994. Muller's ratchet under epistatic selection. *Genetics* 136: 1469-1473.

Kondrashov, A. S. 1995. Contamination of the genome by very slightly deleterious mutations: why have I not died 100 times over? *Theor. Biol.* 175: 583-594.

Kouyous, R. D., Silander, O. K., and Bonhoeffer, S. 2007. Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol. Evol.* 22(6): 308-315.

Kryukov, G. V., Schmidt, S., and Sunyaev, S. 2005. Small fitness effect of mutations in highly conserved non-coding regions. *Hum. Mol. Genet.* 14: 2221-2229.

Lenski, R. E. et al. 1999. Genome complexity, robustness and genetic interactions in digital organisms. *Nature* 400: 66-664.

Lewontin, R.C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 49:49–67.

Li, W. -H. 1987. Models of nearly neutral mutations with particular implications for non random usage of synonymous codons. *Mol. Evol.* 24: 337-345.

Li, W.-H., and Grauer, D. 1991. *Fundamentals of Molecular Evolution.* Sinauer Associates, Inc., Sunderland MA.

Maisnier-Patin, S., Roth, J. R., Fredriksson, A., Nystrom, T., Berg, O. G., and Andersson, D. I. 2005. Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nat. Genet.* 37: 1376-1379.

McVean, G. A. T., and Charlesworth ,B. 1999. A population genetic model for the evolution of synonymous codon usage: patterns and predictions. *Genet. Res*. 74: 145-158.

McVean, G. A. T., and Charlesworth, B. 2000. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* 155: 929-944.

Mestl, T., et al. 1995. A mathematical framework for describing and analysing gene regulatory networks. *J. Theor. Biol.* 176: 291-300.

Moriyama, E. N., and Powell, J. R. 1996. Intraspecific nuclear DNA variation in Drosophila. *Mol. Biol. Evol.* 13: 261-277.

Mukai, T. 1969. The genetic structure of natural populations of *Drosophila melanogaster*. VII. Synergistic interaction of spontaneous mutant polygenes controlling viability. *Genetics* 61:749-761.

Neuhauser, C., and Krone, S. M. 1997. The genealogy of samples in models with s election. *Genetics* 145: 519–534.

Nijhout, H. F., and Paulsen, S. M. 1997. Developmental models and polygenic characters. *Am. Nat.* 149:394-405.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol.*

Syst. 23: 263-86.

Ortlund, et al. 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317 (5844): 1544-1548.

Otto, S. P., and Barton, N. H. 2001. Selection for recombination in small populations. *Evolution* 55: 1921-1931.

Phillips, P. C. 1998. The language of gene interaction. *Genetics* 149:1167-1171.

Phillips, P. C. 2008. Epistasis: the essential role of gene interactions in the structure and evolution of genetic systerms. *Genetics* 9: 855-867.

Phillips, P. C., et al. 2000. Beyond the average: the evolutionary importance of gene interactions and variability of epistasis effects. Pages 20-40 in wolf, J. B., et al. eds., *Epistasis and the evolutionary process*. Oxford University Press, Oxford, Cambridge.

Poon, A., and Chao, L. 2005. The rate of compensatory mutation in the DNA bacteriophage ΦX174. *Genetics* 170: 989-999.

Przeworski, M., Charlesworth, B., and Wall, J. D. 1999. Genealogies and weak purifying selection. *Mol. Biol. Evol.* 16: 246-252.

Rice, S. H. 2000. The evolution of developmental interactions: epistasis, canalization, and integration. Pages 82-98 in Wolf, J. B., et al. eds., *Epistasis and the evolutionary process*. Oxford University Press, Oxford, Cambridge.

Ringo, J. 2004. *Fundamental genetics*. Cambridge University Press, Cambridge.

Sanjuan, R., and Elena, S. F. 2006. Epistasis correlates to genome complexity.

PNAS 103: 14402-14405.

Schaeffer, S. W., and Miller, E. L. 1993. Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. *Genetics* 135: 541-552.

Seager, R. D., Ayala, F. J., and Marks, R. W. 1982. Chromosome interactions in *Drosophila melanogaster*. II. Total fitness. *Genetics* 102: 485-502.

Sharp, P. M., and Li, W. -H. 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24(1-2): 28-38.

Sharp, P. M., Tuohy, T. M. F., and Mosurski, K.R. 1986. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 14: 5125-5143.

Shields, D. C., Sharp, P. M., Higgins, D. G. and Wright, F. 1988. "Silent" sites in Drosophila Genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.* 5: 704-716.

Smith, J. M., and Haigh, J. 1974. The hitchhiking effect of a favorable gene. *Genet. Res.* 23: 23-35.

Stenico, M., Lloyd, A. T. and Sharp, P. M. 1994. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Res*. 22: 2437-2446.

Strobeck, C. 1983. Expected linkage disequilibrium for a neutral locus linked to a

chromosomal arrangement. *Genetics* 103: 545-555.

Szathmary, E. 1993. Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. *Genetics* 133:127-132.

Tachida, H. 1990. A population genetic model of selection that maintains specific trinucleotides at a specific location. *Mol. Evol.* 31: 10-17.

Wade, M. J. 2001. Epistasis, complex trait, and mapping genes. *Genetica* 112-113: 59-69.

Wagner, A. 1994. Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proc. Natl. Acad. Sci. USA* 91: 4387-4391.

Wagner, G. P., et al. 1998. Genetic measurement theory of epistatic effects. *Genetica* 102/103: 569-580.

Weinreich, D. M., et al. 2005. Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59(6): 1165-1174.

Wilke, C. O., and Adami, C. 2001. Interaction between directional epistasis and average mutational effects. *Proc. R. Soc. London* B 268:1469-1474.

Wolf, J. B., Brodie, E. D. III, and Wade, M. J. 2000. *Epistasis and the evolutionary process.* Oxford University Press, Oxford.

You, L., and Yin, J. 2002. Dependence of epistasis on environment and mutation severity as revealed by *in silico* mutagenesis of phage T7. *Genetics* 160: 1273-1281.