

# LONG-TERM ADAPTATION OF EPISTATIC GENETIC NETWORKS

Roman Yukilevich,<sup>1,2</sup> Joseph Lachance,<sup>1</sup> Fumio Aoki,<sup>1</sup> and John R. True<sup>1</sup>

<sup>1</sup>*Department of Ecology and Evolution, State University of New York, Stony Brook, New York, 11794*

<sup>2</sup>*E-mail: yukilevi@life.bio.sunysb.edu*

Received January 4, 2008

Accepted June 2, 2008

Gene networks are likely to govern most traits in nature. Mutations at these genes often show functional epistatic interactions that lead to complex genetic architectures and variable fitness effects in different genetic backgrounds. Understanding how epistatic genetic systems evolve in nature remains one of the great challenges in evolutionary biology. Here we combine an analytical framework with individual-based simulations to generate novel predictions about long-term adaptation of epistatic networks. We find that relative to traits governed by independently evolving genes, adaptation with epistatic gene networks is often characterized by longer waiting times to selective sweeps, lower standing genetic variation, and larger fitness effects of adaptive mutations. This may cause epistatic networks to either adapt more slowly or more quickly relative to a nonepistatic system. Interestingly, epistatic networks may adapt faster even when epistatic effects of mutations are on average deleterious. Further, we study the evolution of epistatic properties of adaptive mutations in gene networks. Our results show that adaptive mutations with small fitness effects typically evolve positive synergistic interactions, whereas adaptive mutations with large fitness effects evolve positive synergistic and negative antagonistic interactions at approximately equal frequencies. These results provide testable predictions for adaptation of traits governed by epistatic networks and the evolution of epistasis within networks.

**KEY WORDS:** Adaptive walks, antagonistic epistasis, constraint, evolutionary theory, individual-based simulation, mutation, sign epistasis, synergistic epistasis.

Genes do not function in isolation, but instead are typically part of complex genetic networks. An important consequence of gene networks is that the effects of mutations are unlikely to be independent of the rest of the genome. Alleles at different loci may epistatically interact, such that a genotype may deviate in fitness from its expectation based on the separate allelic effects at different loci (referred to as “functional epistasis” Whitlock et al. 1995; Phillips 1998).

Recent site-directed mutagenesis studies have provided a wealth of evidence that mutations often exhibit such fitness epistasis (see Sanjuan and Elena 2006 and references therein). The average epistatic effect may range from synergistic to antagonistic (Sanjuan and Elena 2006) and may either change the direction (sign) or magnitude of effect across different genetic backgrounds (Weinreich et al. 2005). Epistasis is also virtually ubiquitous in QTL studies (Carlborg and Haley 2004; Malmberg

and Mauricio 2005), and in crosses between distinct populations and species (Whitlock et al. 1995). Epistatic gene networks underlying specific traits are beginning to be characterized (Fedorowicz et al. 1998; Mackay 2001; van Swinderen and Greenspan 2005; Sambandan et al. 2006).

However, theoretical models of adaptation typically ignore epistasis by assuming that genes contribute independently to fitness. This may be due to the mathematical complexity inherent in describing epistatic genetic architectures as well as the historical emphasis on statistical epistatic variance, which is often a very small proportion of the total segregating fitness variation within populations (Cheverud and Routman 1995; Whitlock et al. 1995; Phillips 1998; Brodie 2000; Hansen and Wagner 2001; Weinreich et al. 2005).

Starting with Sewall Wright (1931, 1932, 1949, 1969), some have repeatedly emphasized the importance of incorporating

epistasis into models of adaptation (Mayr 1959, 1963; Lewontin 1978; Phillips et al. 2000; Wade 2000, 2002; Wagner 2002; Hansen et al. 2006). The role of epistasis has been studied in how it affects stabilizing selection on an additive character (e.g., Gimelfarb 1989; Zhivotovsky and Gavrillets 1992; Gavrillets and de Jong 1993). Much attention has also been given to “how” epistasis generates rugged adaptive landscapes, where there are multiple local fitness optima or peaks (i.e., genotypes surrounded by deleterious one-mutant neighbors; Wright 1932; Kauffman and Levin 1987; Kauffman 1993; Whitlock et al. 1995; Gavrillets 2004; Weinreich and Chao 2005). Many models focus on how populations can escape being trapped at local fitness peaks (Wright 1932; Whitlock et al. 1995; Hansen and Houle 2004; Weinreich and Chao 2005). Because double mutations are rare, the probability of peak shift is often very low (Wright 1932; Maynard Smith 1970; Whitlock et al. 1995; Kauffman 1993; Gavrillets 2004; Weinreich 2005; Weinreich and Chao 2005). Hence, it is generally assumed that a major role of epistasis is to prevent populations from exploring the genotypic landscape by selection (but see Gavrillets 2004; Weinreich et al. 2005).

Aside from understanding how populations escape local fitness peaks, it is also important to study how epistatic genetic architecture influences the actual adaptive walk toward fitness optima during long-term adaptation. This determines the nature of the adaptive process and which types of mutations become established in the evolving population. This question has received much less theoretical attention. An important exception is the *NK* model of Kauffman and Levin (1987) and Kauffman (1993), where *N* is the number of loci in a gene network and *K* is the epistatic connectivity among these loci (also see Carter et al. 2005; Hansen et al. 2006). By changing the parameter *K*, the *NK* model can generate different amounts of fitness epistasis between alleles at different loci in the genome and thus affect the ruggedness of adaptive landscapes. Kauffman (1993) showed that increasing *K* constrains adaptive walks by forcing populations to get stuck on local peaks sooner and at lower fitness. In essence, the waiting times to beneficial mutations become eventually prohibitively long and the population gets stuck.

However, the *NK* model rests on several assumptions that limit its utility in understanding how epistatic networks shape long-term adaptation. First, the model has been studied with the assumption of strong selection, weak mutation (SSWM) scenario (Gillespie 1983, 1984), which ignores segregating genetic variation and assumes that sweeping beneficial mutations enjoy independent fates (i.e., clonal interference and recombination among sweeping mutations is ignored). This leads to completely sequential accumulation of beneficial mutations during an adaptive walk.

On one hand, this allows for functional or physiological epistasis between a mutation and established alleles at other loci (Brodie 2000: p. 15). The fate of a mutation is still highly de-

pendent on how well it interacts with alleles that were fixed beforehand (Wright 1949; Gillespie 1984; Phillips 1996; Phillips et al. 2000; Weinreich et al. 2005). In this sense, the *NK* model is qualitatively similar to Fisher’s (1930) geometric target model, except that the latter assumes functional epistasis only implicitly (Brodie 2000: pg. 15; Orr 2002; Martin et al. 2007). On the other hand, because substitutions occur sequentially, with no variation in genetic background, functional epistasis between different segregating mutations does not contribute to the adaptive process. It is still unknown how the architecture of epistatic networks influences segregating genetic variation and in turn how the presence of multiple genetic backgrounds for mutations changes adaptive walks.

Second, the *NK* model assumes that mutant fitnesses are drawn from the same (fixed) distribution throughout the adaptive walk. This is similar to the nonepistatic House of Cards distribution, where the mutant fitnesses are independent of the fitness of the parental allele (Kingman 1978; Gillespie 1983, 1984, 1994; Orr 2002, 2006). Even though Kauffman (1993) argued that epistatic connectivity was responsible for generating uncorrelated/rugged fitness landscapes, it is clear that even additive models that assume the House of Cards mutation distribution also result in uncorrelated fitness landscapes, as has been recently emphasized by Welch and Waxman (2005; also see Gillespie 1994; Orr 2006). Indeed, Welch and Waxman have called into question the importance of epistatic networks in evolutionary dynamics. First, they show that when selection is weak (i.e., nearly neutral evolution: Ohta 1997, 1998), the *NK* and House of Cards models give nearly identical results. Second, they show that when selection is strong, both models result in fitness plateaus over time where populations get stuck on local peaks (see also Gillespie 1984, 1994). They conclude that the role of epistatic genetic architecture in adaptive walks is unclear.

Further, it is also possible to model mutation distributions that dynamically change over the course of adaptation rather than being fixed. These are known as “shift models” (see Ohta 1977; Hartl et al. 1985; Gillespie 1994). Here mutations are sampled from a distribution centered on or near the fitness of the parental allele. Thus the mutation distribution dynamically shifts with the evolving population, resulting in more correlated and hence smoother adaptive landscapes, with fewer local optima (e.g., Weinberger 1990; Orr 2006). Indeed, a nonepistatic or additive genetic architecture gives a perfectly smooth adaptive landscape with only one global peak (e.g., Kauffman 1993; Orr 2006). It is still unclear whether epistatic and nonepistatic genetic architecture differ in their adaptive walks on smoother fitness landscapes (see also Welch and Waxman 2005).

Finally, the *NK* model has generally focused on how epistatic connectivity of gene networks impacts waiting times to

substitutions and the length of the adaptive walks (Kauffman 1993). Much less is known about how epistatic networks determine the fitness effects of mutations, even though this may be an important feature of epistasis. Unlike other population genetic models, the fitness effects of mutations are intimately linked to epistatic connectivity (Welch and Waxman 2005). This occurs because the fitnesses of mutations are modified by alleles at  $K$  other loci within a given genotype (i.e., epistatic effects). Therefore, increasing the epistatic connectivity of a gene network, in principle, should increase the fitness variance of mutational effects (Welch and Waxman 2005). It is still unknown how the fitness effects of substitutions interact with their waiting times during adaptive walks of epistatic gene networks (also see Phillips et al. 2000). We could then understand how the distribution of fitness effects of substitutions differs between epistatic and nonepistatic models (e.g., Gillespie 1984, 1991; Orr 1998, 2002, 2006).

Here we develop a new model of epistatic gene networks that retains the general  $NK$  framework (i.e., a multilocus system with epistatic connectivity), but extends the model in several directions. Our model encompasses: (1) segregating genetic variation during the adaptive walk, which allows for epistatic interactions between different sweeping mutations, (2) evolution in the presence of fixed and dynamic mutation distributions, and (3) the study of how gene networks affect the fitness effects of substitutions in combination with their waiting times. We focus on “strong selection” scenarios in which the adaptive walk is primarily driven by selection as opposed to genetic drift (see Gillespie 1984; Welch and Waxman 2005). Our results provide novel insights into how epistatic gene networks shape the dynamics of adaptive walks and how these in turn impact the evolution of epistatic interactions between sweeping mutations.

## MODEL

### *Allelic and epistatic effects in multilocus networks*

Consider a model in which there are only allelic effects. In this case, mutations at different loci combine independently toward fitness such that no epistasis is possible. On the other hand, when pairwise or higher order epistatic effects are allowed between mutations at different loci, the model can generate magnitude and/or sign epistasis (e.g., Phillips et al. 2000; Weinreich et al. 2005). The simplest epistatic network consists of two loci (Wright 1932; Gavrillets 2004; Weinreich et al. 2005). More generally, mutations may epistatically interact with alleles at multiple loci in the genome (Mayr 1963; Wright 1969; Lewontin 1974; Kauffman 1993; Wade 2000).

To model multilocus epistasis, we start with a haploid network of  $L$  loci (we use  $L$  instead of  $N$  to denote number of loci following standard usage because  $N$  is reserved for population size). First, we assume that each allele at every locus has

an allelic effect ( $W_i$ ), which is independent of alleles at other loci. Second, we assume that each pair of alleles has a unique epistatic effect ( $W_{i-i}$ ), which applies only when both alleles are present in the genotype. Epistatic effects between different pairs of alleles in the genotype are assumed to be independent of each other (e.g., Phillips et al. 2000). For a multilocus genotype, these terms are combined to obtain the overall fitness of the individual (also see Zhivotovsky and Gavrillets 1992; Gavrillets and Jong 1993). This is done by multiplying the allelic and epistatic effects:

$$\{\text{allelic effects}\} \bullet \{\text{epistatic effects}\} = \{w_{A_i} \bullet w_{B_i} \bullet w_{C_i} \dots\} \\ \bullet \{w_{A_i-B_i} \bullet w_{A_i-C_i} \bullet w_{B_i-C_i} \dots\}. \quad (1)$$

We assume there is no environmental effect on genotype fitness. Each fitness effect is nonnegative and is initially assumed to equal 1 (initially alleles have no epistasis). We allow allelic effects in the model because the fitness of an allele may be partially independent of genetic background (Jerry Garcia effect: Brodie 2000; see Threadgill et al. 1995; Chambers et al. 1998; Elena and Lenski 2001 for empirical examples). Multiplying terms across the genotype is a preferred null model because it avoids creating false correlations between alleles at different loci in the absence of epistasis (Felsenstein 1965; Ewens 1979; Kauffman 1993; Phillips et al. 2000; Welch and Waxman 2005, also see empirical epistasis studies: e.g., Sanjuan et al. 2004). It is also convenient that without epistatic effects, our model reduces to a standard multiplicative model, where the rate of evolution increases linearly as a function of the number of loci (Felsenstein 1965; Franklin and Lewontin 1970; Lewontin 1974; Ewens 1979; Gavrillets 2004). This model is used for comparison below. Eliminating allelic effects and adding instead of multiplying terms have no bearing on our results. We also ignore higher order epistasis (but see Analytical Framework).

Because not all pairs of loci in a multilocus network exhibit allelic epistatic interactions (as observed in e.g., Fedorowicz et al. 1998; Sambandan et al. 2006), we incorporate epistatic connectivity ( $K$ ). We define epistatic connectivity slightly differently than in Kauffman (1993). In our model,  $K$  is the proportion of loci in a network that can potentially participate in allelic epistatic interactions. Thus  $K$  is a percentage rather than an integer in our model and may range from 0% (no alleles can interact) to 100% (alleles between all loci can interact). Partially connected gene networks result when alleles between some loci, but not between others, are capable of epistatically interacting. By defining  $K$  in terms of the proportion of loci that are epistatically connected, we can study the effect of  $K$  independent of the number of loci in a network.

## Individual-Based Simulations

### THE LIFE CYCLE

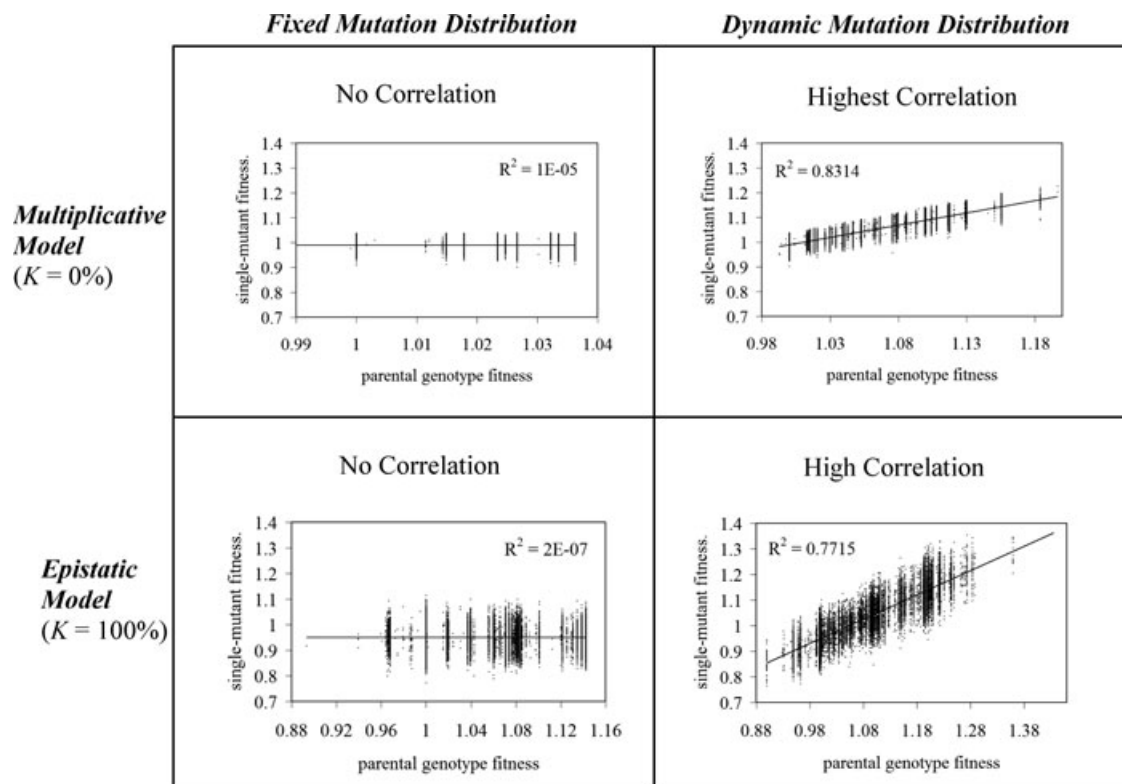
We use Monte Carlo simulations written in C (code available by request). We convert arbitrary genotypic fitness values in equation (1) to probability of zygote to juvenile survivorship ranging from 0 to 1. We assume that the population starts out monomorphic with a single (reference) genotype with a probability of zygote to juvenile survivorship of 55% (initial value 1 in eq. 1). The fitness of a mutation in equation (1) is multiplied by the reference genotype survivorship of 55% to get the new survivorship probability. We use 55% initial viability to ensure that the population replaces itself every generation with our assumed fecundity (see below). The life cycle of the program begins with adults and proceeds through adult mating → zygote production → recombination → meiosis → mutation → zygote survival → juvenile density-dependent mortality.

First, pairs of individuals are randomly sampled with replacement. Each pair mates and produces a standard number of zygote offspring (fecundity = 2). Sampling continues until fecundity ×  $N$  zygotes are accumulated. To study the effect of recombination, we introduce a transient diploid stage in which each zygote undergoes meiosis, allowing for recombination between maternal and paternal genes. The zygotes are then exposed to mutations at their  $L$  loci, which modify their probability of survivorship. We assume

that fitness effects are sampled from a Gaussian mutation distribution (see below for details). Zygotes then experience hard viability selection based on their probability of survival. Then the surviving juvenile genotypes undergo density-dependent mortality to the adult stage by being randomly sampled until the surviving pool reaches the carrying capacity, which is set by the logistic growth equation. The surviving adult pool then initiates the next generation.

### FIXED VERSUS DYNAMIC MUTATION DISTRIBUTIONS

We explore fixed (House of Cards) and dynamic (shift) mutation distributions during the adaptive walk (see below for details). This allows us to study both highly uncorrelated (rugged) and more correlated (smoother) fitness landscapes, respectively (see Introduction). Figure 1 shows actual simulated data describing the correlation between the fitness of parental genotype and its single mutants for a multiplicative model and a 100%  $K$  epistatic model. The fixed mutation distribution generates uncorrelated fitness landscapes. The dynamic mutation distribution generates highly correlated fitness landscapes, but somewhat less so in an epistatic model compared to a multiplicative model. Also, the epistatic model produces a wider range of single mutant fitnesses compared to the multiplicative model under both fixed and dynamic mutation distributions (Fig. 1).



**Figure 1.** An example of simulated data showing the correlation between parental genotype fitness and its single-mutant genotypic fitnesses in a five-locus model (parameters identical to Figs. 4 and 8; see below).

## SIMULATED PARAMETERS

Population sizes ( $N$ ) are maintained at a carrying capacity of 50,000 adults. Genetic networks are composed of 2 to 40 loci. Epistatic connectivity of a gene network ranges from  $K = 0\%$  (multiplicative model) to  $K = 25\%$ ,  $K = 50\%$ , and  $K = 100\%$ . Connections are random for partially connected networks. We also explore power law, ring and star connection topologies. Recombination rates between neighboring loci range from  $R = 0$  (complete linkage),  $R = 0.25$ , and  $R = 0.5$  (independent assortment). Mutation rate per locus per generation =  $10^{-5}$  (Falconer and Mackay 1996). Per locus  $\theta$  ( $4Nu$ ) = 2. Mutations are assumed to be on average deleterious (e.g., Nielsen and Yang 2003; Kassen and Bataillon 2006; Eyre-Walker et al. 2006; Eyre-Walker and Keightley 2007).

### Fixed mutation distribution

In this scenario, a constant mutation distribution is assumed throughout the adaptive walk, with a fixed mean ( $\mu$ ) for allelic and epistatic effects = 99% of initial genotype fitness and a fixed variance ( $\sigma^2$ ) = 0.0004.

### Dynamic mutation distribution

Here the mean allelic and epistatic effects of a mutation ( $\mu_a$  and  $\mu_e$ , respectively) are relative to those of its immediate parental allele. The  $\mu_a$  of a mutation = 99% fitness of the allelic effect of its parental allele throughout the adaptive walk (i.e., deleterious on average). The means of epistatic effects of a mutation ( $\mu_e$ 's) are assumed to be either deleterious or neutral relative to the effects of its parental allele (referred to as Model 1 and Model 2, respectively). In Model 1 all pairwise  $\mu_e$ 's of a mutation = 99% fitness of the epistatic effects of its parental allele. In Model 2, all pairwise  $\mu_e$ 's of a mutation = 100% fitness. Note that in Model 2, mutations are still always on average deleterious because their mean allelic effects are deleterious.

For simplicity, all loci and their epistatic interactions are assumed to contribute equally, on average, to the genotypic fitness (i.e., variance,  $\sigma^2$ , of allelic and epistatic distributions are identical and are assumed to equal 0.0004). Given the population size above, this assumes a "strong selection scenario" because per locus  $N\sigma > 1000$  (see Welch and Waxman 2005). For a fixed mutation distribution scenario, each simulation is run until a population reaches a local genotypic optimum (defined when there are no more substitutions for 50,000 generations and a fitness plateau is evident). For the dynamic mutation distribution scenario, the simulation is stopped when a fitness of 100% zygote to juvenile survivorship is reached by a substitution (i.e., a global peak).

## Results

### ANALYTICAL FRAMEWORK

Here we describe a theoretical framework that sharpens our intuition about adaptation of epistatic networks in the presence of mutations and strong selection. The framework follows Gillespie's (1984) Strong Selection Weak Mutation (SSWM) scenario, which ignores segregating genetic variation within populations. This framework also only applies to a dynamic mutation distribution (shift model), where the mean and variance of mutant genotypes shifts with the evolving population such that, relative to a wild-type allele, the mean and variance remains the same throughout the walk (see Fig. 1). Because all fitness effects are multiplied to obtain the genotypic fitness (see eq. 1), the effect of a mutation on the fitness of a multilocus genotype relative to its parental genotype can be approximated using the log-normal distribution, with a mean and variance:

$$m = [1 + (L - 1) \cdot K] \cdot \ln \left( \frac{\mu_i}{\sqrt{1 + \sigma_i^2/\mu_i^2}} \right) \quad (2)$$

$$v = [1 + (L - 1) \cdot K] \cdot \ln (1 + \sigma_i^2/\mu_i^2) \quad (3)$$

where  $L$  is the number of genes in the network,  $K$  is the epistatic connectivity, and  $\mu_i$  and  $\sigma_i^2$  are the mean and variance, respectively, of each allelic and epistatic effect of a mutation. The expected value and variance of mutant genotypes is given by:

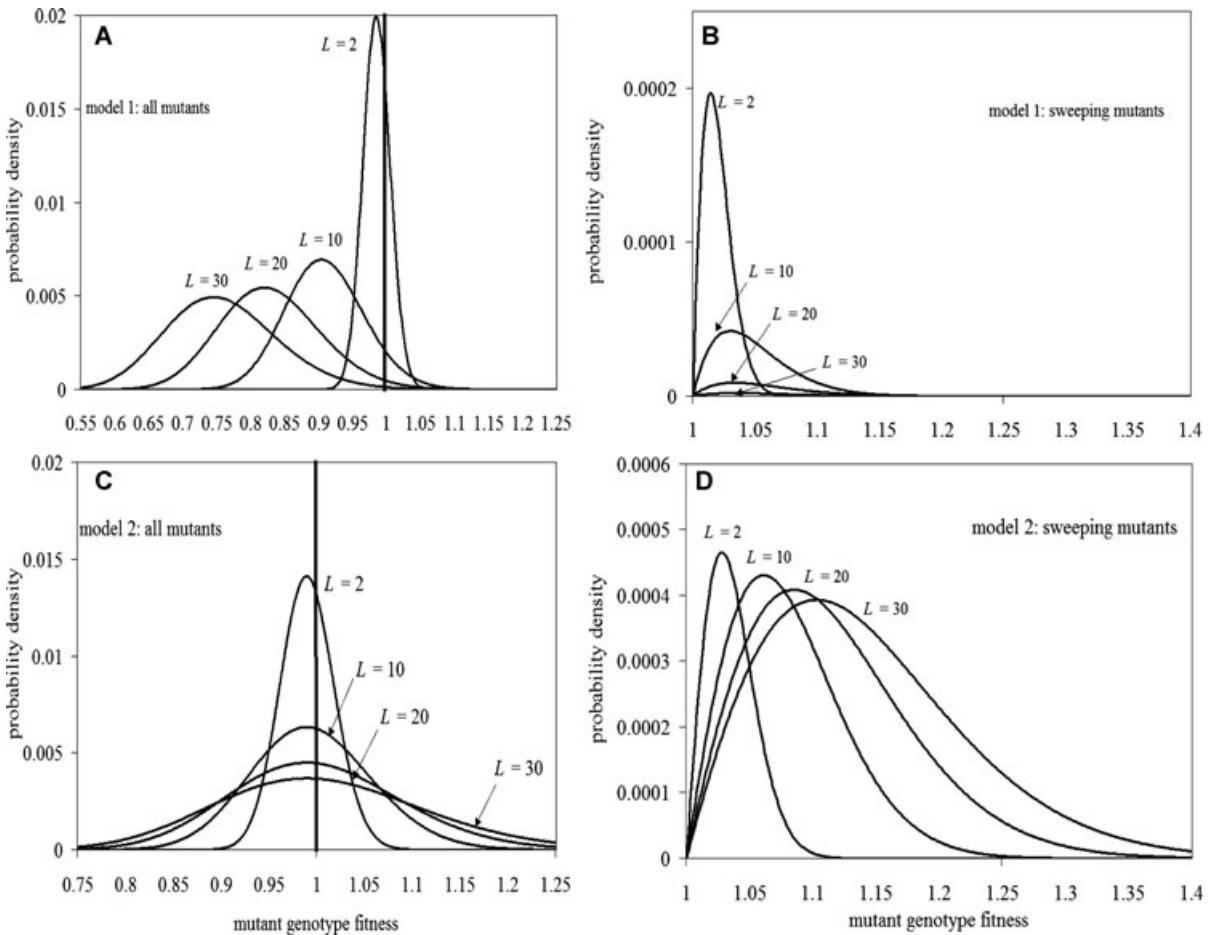
$$M(X) = e^{m+v/2} \quad (4)$$

$$V(X) = (e^v - 1)e^{2m+v} \quad (5)$$

with a probability density function of:

$$pdf = \frac{e(-[\ln(X) - m]^2/2v)}{X\sqrt{2v\pi}} \quad (6)$$

First, consider a model in which  $K = 0$ , such that mutations do not interact with any alleles. In this case, mutations only modify the allelic effect. Thus,  $m$  and  $v$  in equations (2) and (3), respectively, should remain the same, regardless of  $L$ . However when  $K > 0$ , pairwise epistatic effects can modify  $m$  and  $v$  in equations (2) and (3). Because in Model 1 epistatic effects are on average deleterious (all  $\mu_e$ 's  $< 1$  in equations (2) and (3)), the mean mutant genotype ( $M(X)$  in equation (4)) will be more deleterious with increasing  $L$  and  $K$  (Fig. 2A). However, in Model 2 (average neutral epistatic effects ( $\mu_e$ 's) = 1 in equations (2) and (3)), the mean mutant genotype will be the same in fitness, regardless of  $L$  and  $K$  (Fig. 2C). Here, the mean mutant fitness is only determined by the mean allelic effect of a mutation, which is deleterious in the Figure 2C example. Finally, in both Models 1 and 2 the mutant



**Figure 2.** Fitness distributions of all mutant genotypes (A, C) and sweeping beneficial mutant genotypes (B, D). Plots show the change in the distribution of mutants as a function of number of loci in an epistatic network for Model 1 (A, B: average deleterious epistasis) and Model 2 (C, D: average neutral epistasis). The value 1 equals selective neutrality, with all genotypes to the left of 1 deleterious and to the right of 1 beneficial. The above results assume  $K = 100\%$ . For Model 1 (A, B)  $\mu_{\text{allelic effects}} = 0.99$ ,  $\mu_{\text{epistatic effects}} = 0.99$ , and  $\sigma^2_{\text{all effects}} = 0.0004$ . For Model 2 (C, D)  $\mu_{\text{allelic effects}} = 0.99$ ,  $\mu_{\text{epistatic effects}} = 1.0$ , and  $\sigma^2_{\text{all effects}} = 0.0004$ .

variance ( $V(X)$  in eq. 5) should increase with  $L$  and  $K$ , generating more extreme deleterious and beneficial mutations (Fig. 2A, C). Including higher-order epistasis should exaggerate the above results because the fitness of a mutation would depend on more epistatic terms.

Because our interest is to describe adaptation, we focus on the fitness distribution of only mutant genotypes that undergo selective sweeps. The probability that a beneficial mutation occurs is given by equation (6) with  $X = 1 + s$ , where  $s$  is the positive selection coefficient. The probability of fixation of a beneficial mutation can be approximated by diffusion equations (Kimura 1962) or by  $\sim 2s$  in large populations when  $s$  is not too large (Haldane 1927). Considering all possible beneficial mutations, to obtain the probability that a sweep occurs ( $P$ ), one integrates from zero to infinity

$$P = \int_0^\infty \left( \frac{e(-[\ln(1+s) - m]^2/2v)}{(1+s)\sqrt{2v\pi}} \right) 2s \, ds. \quad (7)$$

Therefore, in Model 1 (average deleterious epistasis), as  $L$  and  $K$  increases, the likelihood of a sweep should decrease, whereas its mean fitness effect should increase (Fig. 2B). However, in Model 2 (average neutral epistasis), the likelihood and the mean fitness effect of a sweep should both increase (Fig. 2D). The waiting time to selective sweeps ( $T$ ) is described by

$$T = \frac{1}{NLu \cdot P}, \quad (8)$$

where  $NLu$  (the product of population size, number of loci, and per locus mutation rate) is the genome-wide per-generation rate of new mutations. Given the distribution of sweeping mutations (see Fig. 2D and eq. 7)) and weighting by  $s$ , the mean fitness effect of a sweep ( $f$ ) can be found numerically. This leads to a simple relationship for the genome-wide rate of evolution:

$$\Delta \bar{w} = NLu \cdot P \cdot f. \quad (9)$$

Because we are interested in how epistasis between different mutations affects dynamics, we must allow for the possibility of standing genetic variation to influence the evolutionary fate of mutations (see Introduction). Thus we study dynamics using individual-based simulations below. We later discuss how the above theoretical framework compares with our simulation results.

### GENERAL SIMULATION DYNAMICS

When a relatively beneficial mutation appears in the population, it may undergo a selective “sweep” or substitution (this is a “quasi-fixation” event because there is a continuous input of new mutations; Kimura 1954). Because there is a constant input of mutations, we define a selective sweep somewhat differently from a standard definition, as any mutation that overtakes the majority frequency of the previously established allele at a locus. Under the strong selection scenario, the locus-specific waiting time for a sweep (as defined above) is greater than the time it takes for a mutation to sweep.

Figure 3 illustrates an example of an adaptive walk under the dynamic mutation distribution model (for adaptive walks with fixed mutation distribution, see Fig. 8 below). The trajectory reveals distinct adaptive steps, each step representing a selective sweep at some locus (Fig. 3A). With each sweep, a new multilocus genotype succeeds and becomes the majority in the population (Fig. 3A, B). The mean fitness is first below the fitness of the current genotypic sweep, but then exceeds that fitness, indicating that more beneficial mutant genotypes begin to segregate in the population. The whole process then repeats.

We focus on the genome-wide waiting times to sweeps, which are the times between two sequential sweeps at any locus in the genome (horizontal lines in the adaptive steps of Fig. 3A). These correspond to trajectories of beneficial mutant genotypes in Figure 3B. Even though each sweep at a particular locus approaches fixation, the respective mutant genotype may only represent a fraction of the population because of other mutations at other loci (Fig. 3B). Thus, a beneficial mutant may be replaced by an even more beneficial genotype as a result of a sweep at another locus.

Also, depending on the level of genetic variation and recombination rates among loci, sweeping mutations can recombine to form double-mutant genotypes, which may also sweep. In the context of epistatic genetic networks, which allow epistatic effects between mutations, the double-mutant genotypes may deviate from their expected products of single mutant effects in either the positive (synergistic epistasis) or the negative (antagonistic epistasis) direction (see below).

Each time a selective sweep occurs, we measure its fitness effect relative to its parental genotype (see vertical lines in the adaptive steps of Fig. 3A). Further, Figure 3C shows the number

of genotypes that are able to segregate in the population and their fitness variance (see Fig. 3B for their frequencies). The presence of such genetic variation in our simulations highlights the need to study evolutionary dynamics using individual-based models. Below we present summary statistics describing the averages of simulation runs.

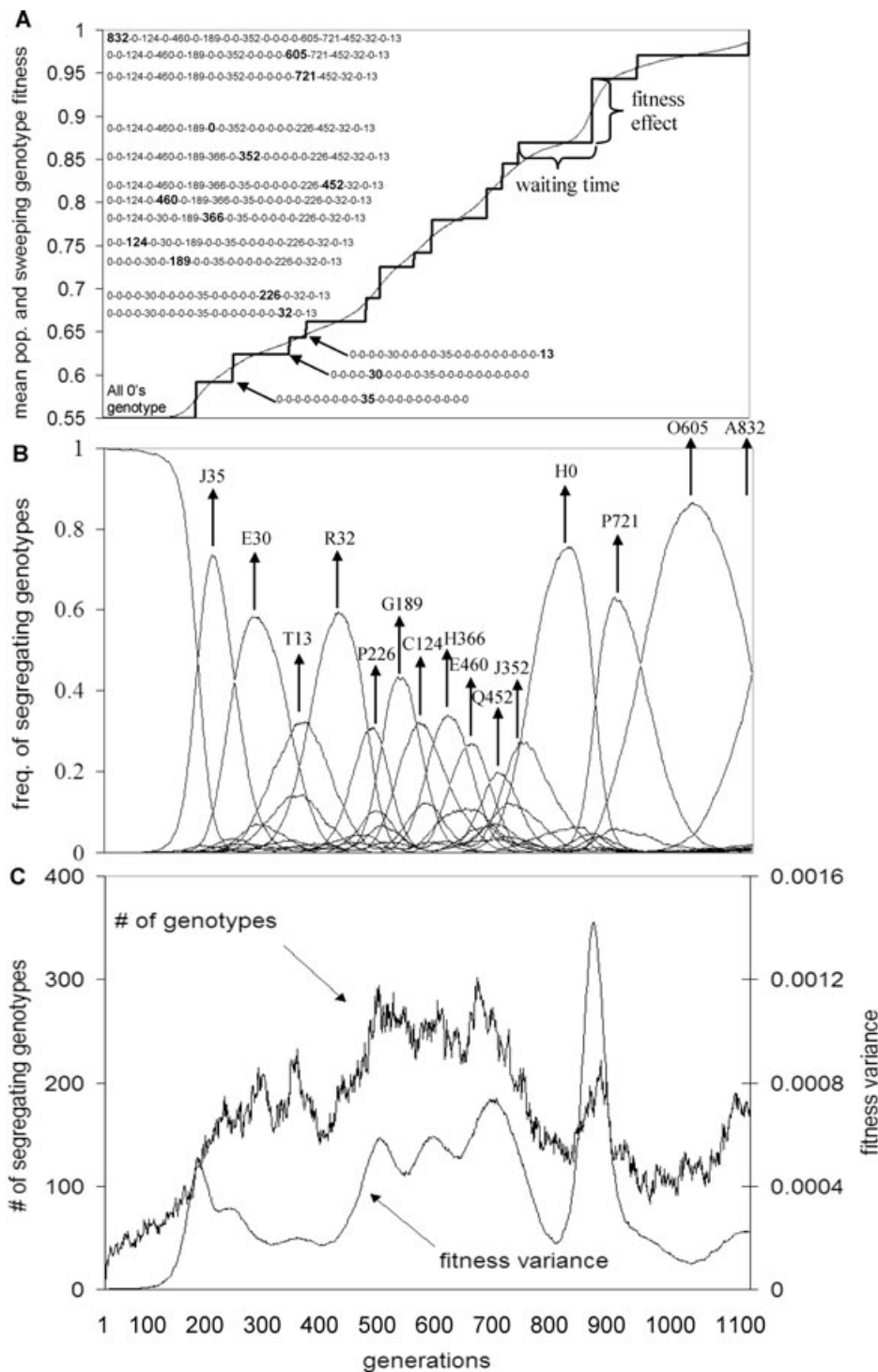
### ADAPTIVE WALKS ON SMOOTH FITNESS LANDSCAPES

We first study adaptive walks under the dynamic mutation distribution scenario, which assumes that the mean and variance of mutations shift proportionately with the evolving population over time (see Model). This generates a high correlation in fitness between parental genotypes and their single mutants (see Fig. 1; i.e., landscapes are fairly smooth). We later show that our findings on smooth fitness landscapes extend well to adaptive walks on rugged fitness landscapes.

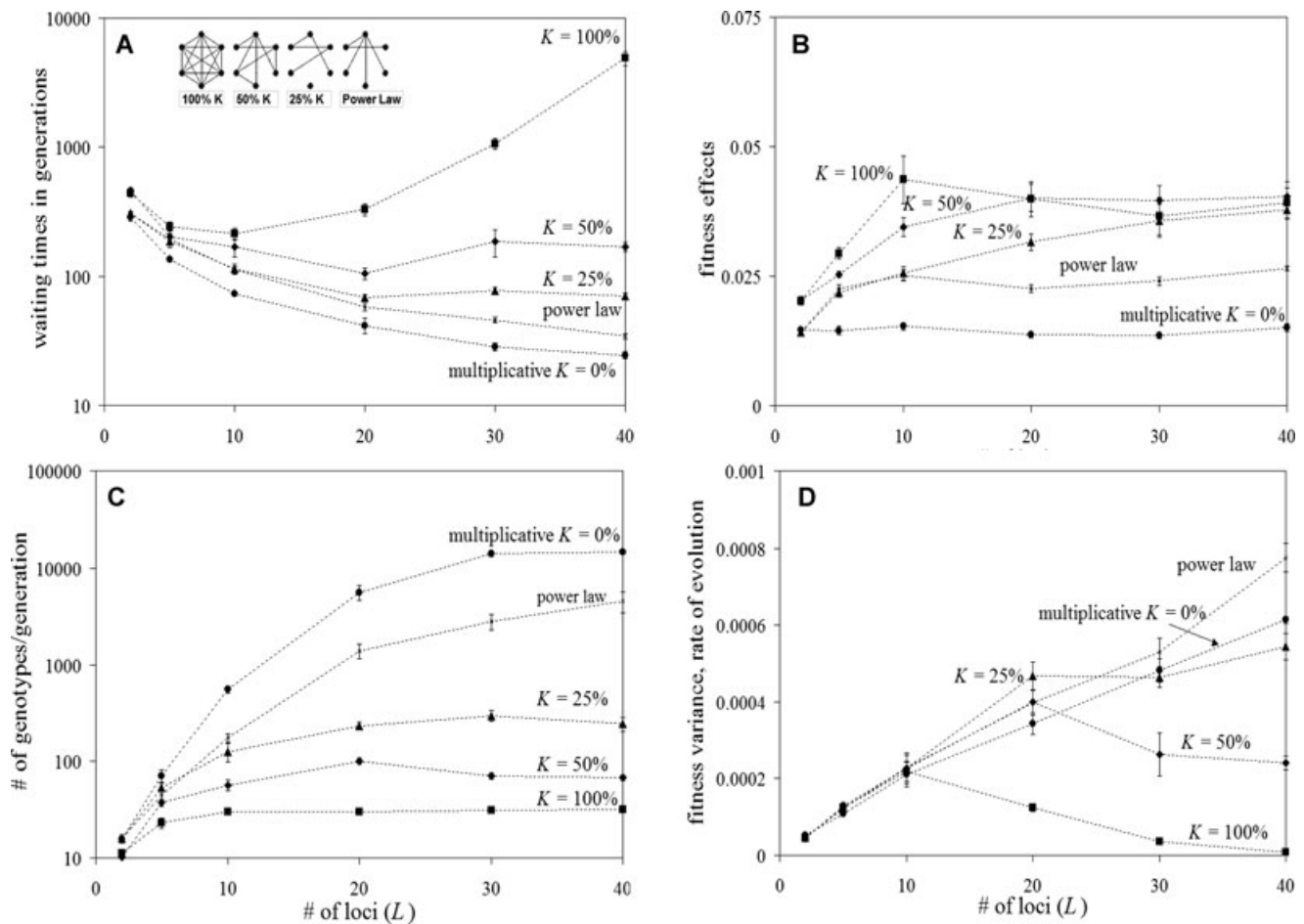
#### *Dynamics of model 1: Average deleterious epistasis*

First, consider the average epistatic effects of mutations to be relatively deleterious compared to those of their parental alleles throughout the adaptive walk. In a multiplicative model, average genome-wide waiting times to sweeps decrease linearly with  $L$  (Fig. 4A). This is expected from quantitative genetics theory because the mutational input into the population increases proportionately (Felsenstein 1965; Lewontin 1974; Ewens 1979; Falconer and Mackay 1996). In epistatic networks, however, the average waiting times to sweeps reveal a nonlinear, hyperbolic function with  $L$ . Waiting times are lowest at intermediate  $L$  but dramatically increase as  $L$  gets larger. This occurs because the mutational input into the population increases, but then as  $L$  increases further, on average mutations have more deleterious interactions with alleles in the genome, thus eventually reducing the probability of beneficial mutations (see Fig. 2B for how  $L$  changes the beneficial mutant distribution). The same effect occurs as  $K$  increases while  $L$  is kept constant. Thus our results show that in Model 1, epistasis can lead to prolonged evolutionary stasis in larger and more connected networks relative to a multiplicative model.

Further, when mutations sweep through the population, their average fitness effects are generally larger in epistatic networks compared to a multiplicative model (Fig. 4B). This effect increases with initial  $L$  and generally with  $K$ . We test whether this result is sensitive to the mutational variance of allelic and epistatic effects ( $\sigma^2$ ). By reducing the mutational variance of these effects, the difference in fitness effects of substitutions between the two models tends to diminish (Table 1). Under low variance parameters, fitness effects in epistatic networks may actually decrease with increasing  $L$  or may become deleterious altogether (Table 1). However, under most “strong selection” scenarios,



**Figure 3.** An example of a simulation run and some of its basic outputs. The parameters in this simulation are as follows:  $L = 20$ ,  $K = 25\%$ ,  $\mu_{\text{allelic effects}} = 0.99$ ,  $\mu_{\text{epistatic effects}} = 0.99$ ,  $\sigma^2_{\text{all effects}} = 0.0004$ , mutation rate/locus =  $10^{-5}$ ,  $N_{\text{adults}} = 50,000$ ,  $N_{\text{zygotes}} = 10^5$ ,  $(\theta = 1)$ ,  $\text{rec. rate/all loci pairs} = 0.25$ . In (A) the population mean fitness (smooth line) and sweeping genotypes fitness (step line) are shown over time (in generations). The genotypes with their actual allelic composition are shown adjacent to their corresponding sweep. Novel mutations of each genotype are shown as enlarged/bold values; the actual value indicates the number of attempted mutations before the present sweep at a given locus. In (B) the evolutionary trajectories of segregating genotypes are shown over time (each genotype's frequency increases and is then replaced by subsequent genotypes). Genotypes whose mutations led to selective sweeps at their respective loci (A-T) are labeled with upward arrows, which correspond to fitness jumps in panel A. In (C) the number of segregating genotypes and their overall fitness variance in the population are shown over time (fitness variance =  $\hat{\alpha}[\text{Freq}_i \times (W_i - W_{\text{mean}})^2]$ , where  $\text{Freq}_i$  and  $W_i$  are the frequency and fitness of each genotype, respectively, and  $W_{\text{mean}}$  is the mean population fitness).



**Figure 4.** Summary statistics of adaptive walks assuming mutations exhibit on average deleterious epistatic effects (Model 1). The plots show: average waiting times to genome-wide sweeping beneficial mutations on a log scale (A), average fitness effects of genome-wide sweeping beneficial mutations (B), average number of segregating genotypes per generation on a log scale (C), and average fitness variation and average rate of evolution (D) in epistatic networks with  $L$  loci and  $K$  connectivity. For each run we averaged each output; variation during a run is not shown, but see example in Figure 2. The rate of evolution of each run (panel D), was determined by dividing the total fitness change from start to end of a run by the total number of generations. Each scenario was studied using five independent replicate runs (standard errors represent  $n = 5$ ). Each run generated from 25 (2-locus system) to 10 (40-locus system) substitutions during the adaptive walk for a total of 125 to 50 substitutions for five runs. For each run,  $\mu_{\text{allelic effects}} = 0.99$ ,  $\mu_{\text{epistatic effects}} = 0.99$ ,  $\sigma^2_{\text{all effects}} = 0.0004$ , mutation rate/locus =  $10^{-5}$ ,  $N_{\text{adults}} = 50,000$ ,  $N_{\text{zygotes}} = 10^5$ , ( $\theta = 1$ ), *rec. rate/all loci pairs* = 0.25. Partially connected networks ( $K = 25\%$ ,  $50\%$ ) had randomly assigned connections, whereas Power Law curves have a scale-free power law distribution (see Siegal et al. 2007).

epistatic networks still produce larger fitness effect sweeps relative to a multiplicative model. We also find that the distribution of fitness effects of substitutions over the adaptive walk is uniform. This is because with a dynamic mutation distribution, the mean and variance remain identical relative to wild-type alleles throughout the adaptive walk (but see fixed mutation distribution results in Fig. 8 below).

Power law (or scale-free) networks (see Siegal et al. 2007) exhibit intermediate behavior between a nonepistatic system and a network with  $K = 25\%$  (Fig. 4A, B). Similar behavior is seen for other sparsely connected networks, such as ring and star topologies (Fig. A1; Appendix). These results indicate that unlike the

overall connectivity within networks, the actual connection topology has very little effect on the evolutionary dynamics of the entire genetic network (consistent with Kauffman 1993; also see Welch and Waxman 2005: pg. 19).

The average number of segregating genotypes and their fitness variation within the population are lower in epistatic networks compared to a multiplicative model (Fig. 4C, D; note that Fig. 4D describes both fitness variance as well as total rate of evolution because these patterns are identical in our model). Similar results to Figure 4C are seen for average allelic heterozygosity and genotypic diversity measures.  $K$  has a dramatic effect on the average number of segregating genotypes and their fitness

**Table 1.** Mean fitness effects of substitutions as a function of standard deviation ( $\sigma_{\text{all effects}}$ ) of allelic and epistatic effects.

$N \times SD$	1000	450	300
<i>SD</i>	0.02	0.009	0.006
<i>N</i>	50,000	50,000	50,000
<i>K=0%</i> model (all loci)	0.0144	0.0081	0.0047
<i>K=100%</i> models			
<i>L=2</i>	0.0202	0.0100	0.0055
<i>L=5</i>	0.0294	0.0124	0.0063
<i>L=10</i>	0.0436	0.0139	0.0067
<i>L=20</i>	0.0399	0.0150	0.0051
<i>L=30</i>	0.0365	0.0154	NA
<i>L=40</i>	0.0392	0.0145	NA

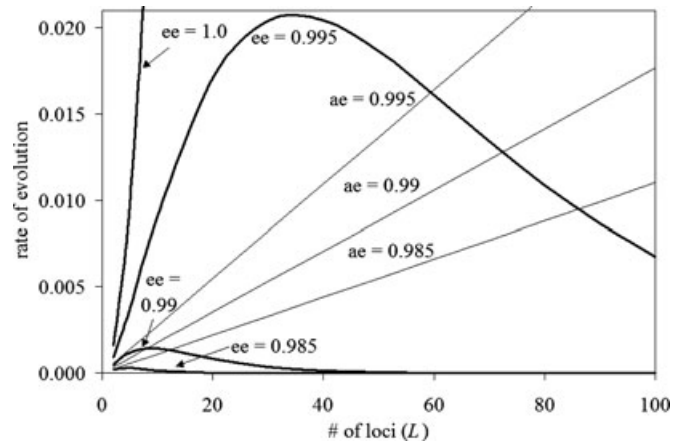
Note: Data averaged across 5 replicate runs for each set of loci as in Figure 4. "NA" for  $L = 30, 40$  indicates that beneficial mutations are not available (see text).

variation, with higher connectivity leading to much lower standing variation (Fig. 4C, D). Similarly, as  $L$  increases, genetic variation increases little in high  $K$  networks and fitness variation actually decreases (Fig. 4C, D).

These results may seem surprising at first, but can be explained by the observation that larger  $L$  and  $K$  decrease the probability of beneficial mutations because mutants on average have more deleterious interactions with alleles at other genes (see distributions in Fig. 2A). Therefore, most genotypes segregate at low frequencies and thus cannot contribute much to fitness variance in the population. In agreement, as  $L$  increases, statistical epistatic variance in the population actually decreases (Fig. A2; Appendix). This somewhat counterintuitive result indicates that increasing  $L$  and  $K$  in the network will make the selective value of a mutation less dependent on genetic context.

Given that epistatic networks have longer waiting times to sweeps and lower genetic and fitness variation, but yet have substitutions of larger fitness effect relative to a multiplicative model, we ask which model shows overall higher rates of evolution. Our results indicate that unlike a linear change in the rate of evolution with  $L$  in a multiplicative model, in epistatic networks, the rate changes hyperbolically with  $L$  (Fig. 4D). Therefore, with small or intermediate  $L$ , epistatic networks either evolve faster than a multiplicative model or show similar rates. However, as  $L$  and  $K$  increase, epistatic networks ultimately lead to evolutionary constraint (Fig. 4D). Thus larger effect mutations of epistatic systems are able to compensate for their longer waiting times only when networks are moderately sized.

In Figure 5 we ask how general are these patterns when we modify the mean epistatic effects of mutations. As these effects become less deleterious on average, network sizes that previously had lower rates of evolution than a multiplicative model are now able to evolve faster (Fig. 5). Thus, depending on specific param-



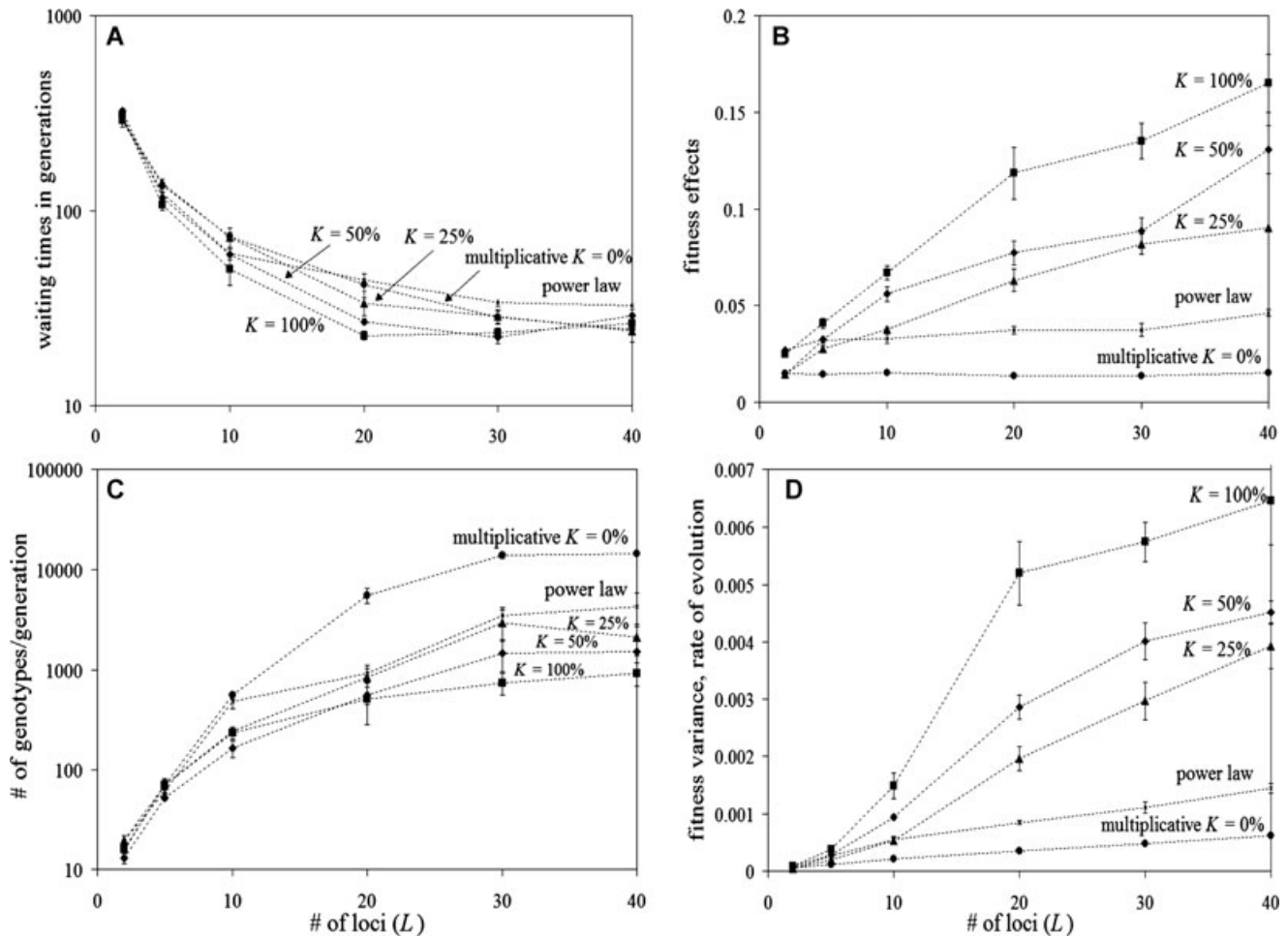
**Figure 5.** Approximations of average rate of evolution of a multiplicative model (mutations have only allelic effects; designated as "ae") and an epistatic network with  $K = 100\%$  (mutations also have epistatic effects; designated as "ee"). The plot shows the relationship between the average allelic and epistatic effects of mutations and the rate of evolution. All parameters are the same as in Figure 4 except for variable  $\mu_{\text{allelic effects}}$  and  $\mu_{\text{epistatic effects}}$ . Note that Figure 4 assumed  $\mu_{\text{allelic effects}}$  and  $\mu_{\text{epistatic effects}} = 0.99$ . Results were derived from numerical approximations of equation (9).

eters of  $L$  and  $K$  and average epistatic effects, traits governed by epistatic networks may not necessarily show evolutionary constraint relative to a multiplicative model.

#### Dynamics of model 2: Average neutral epistasis

Results in Figure 5 lead us to study a scenario in which the mean epistatic effect is selectively neutral relative to the effect of the parental allele (Model 2), while still keeping the assumption of average deleterious allelic effects (see Fig. 2C for description of mutational distributions). As mutations become less deleterious on average, the population is expected to harbor much higher levels of genetic variation. Thus we can address how epistatic systems evolve when there are multiple beneficial mutations co-segregating and recombining in the population.

In this case, the genome-wide waiting times to sweeps become very similar between epistatic networks and a multiplicative model (Fig. 6A). This differs substantially from results above (compare Fig. 4A with Fig. 6A). Indeed, the shortest average waiting times occur in epistatic networks with intermediate  $L$  (Fig. 6A). This occurs because epistatic networks no longer generate more deleterious mutant genotypes with increasing  $L$  and  $K$  (see Fig. 2C). Further, epistatic networks select for mutations of much larger fitness effect compared to a multiplicative model (Fig. 6B; also compare to Fig. 4B). Also notice that fitness effects increase more linearly with increasing  $L$  (see also Fig. 2D). Unlike in Model 1, this result is no longer sensitive to the mutational variance.



**Figure 6.** Summary statistics of adaptive walks assuming mutations exhibit on average neutral epistatic effects (Model 2). The plots show: average waiting times to genome-wide sweeping beneficial mutations on a log scale (A), average fitness effects of genome-wide sweeping beneficial mutations (B), average number of segregating genotypes per generation on a log scale (C), and average fitness variance and average rate of evolution (D) in epistatic networks with  $L$  loci and  $K$  connectivity. All parameters are identical to Figure 4 except for  $\mu_{\text{epistatic effects}} = 1.0$ .

Even though epistatic networks again segregate lower genetic variation compared to a multiplicative model (Fig. 6C), much more genetic variation can now be maintained (compare Fig. 4C with Fig. 6C). Fitness variation is also greater in epistatic networks relative to a multiplicative model and increases proportionately with  $L$  and  $K$  (Fig. 6D). This is strikingly different from results above (compare Fig. 4D with Fig. 6D). This occurs because the likelihood of beneficial mutations, instead of decreasing with  $L$  and  $K$  (as in Model 1), now increases (compare Fig. 2B with Fig. 2D). Thus in Model 2, the overall rate of evolution is always higher in epistatic networks relative to a multiplicative model and increases with  $L$  and  $K$  (Fig. 6D). This result is robust to asymmetrical distributions of mutant epistatic effects. Consistent with these results, epistatic networks generate larger statistical epistatic variance compared to Model 1 (Fig. A2: Appendix). However, epistatic variance is still extremely low dur-

ing the adaptive walk (see also Whitlock et al. 1995). As shown below, much of the accelerated evolution in epistatic networks in Model 2 occurs principally because multiple beneficial mutations sweep simultaneously and recombine to form even more beneficial allelic combinations.

#### *Sequential versus cosegregating and recombinant evolution of adaptive mutations*

In our simulations, mutations either sweep sequentially through the population, where the second mutation directly appears on the first mutant's genetic background, or both appear independently on an ancestral (parental) genetic background and then increase in frequency simultaneously (i.e., cosegregate). In Model 1, these alternative evolutionary histories of sweeping mutations occur equally frequently, whereas in Model 2 all sweeping single mutations originate on the ancestral background and thus always

**Table 2.** Percentage of sequential vs. cosegregating origin of mutations that result in sweeping double mutant genotypes and magnitude versus sign epistasis between these mutants.

	Model 1	Model 2	Diagram
Sequential	52% (n=15)	0	
Cosegregating	48% (n=14)	100% (n=46)	
Both magnitude	72% (n=21)	100% (n=46)	
Sign/magnitude	28% (n=8)	0	
Both sign	0	0	

Note: Mutations are sampled equally from 2, 5, 10, 20, 30, and 40 locus systems with  $K = 100\%$  in simulations from Figures 4 and 6. The asterisks in diagrams represent mutations at different loci. Sequential scenario means that the second mutation originated on the background of the first mutant. Cosegregating scenario means that the two mutations independently appeared on an ancestral background and then recombined to form the double mutant genotype. In the two-locus adaptive landscape diagrams, the vertical axis represents genotypic fitness. The ancestral genotype is designated as  $A_0B_0$ , with subsequent single and double mutant genotypes shown accordingly. Both loci show magnitude epistasis because  $A_0 \rightarrow A_1$  and  $B_0 \rightarrow B_1$  mutations are both beneficial on either background of the other locus. Each mutation only deviates in magnitude from its multiplicative expectation. Locus A shows sign epistasis because  $A_0 \rightarrow A_1$  mutation is beneficial on  $-B_1$  background but deleterious on the  $-B_0$  background, whereas locus B shows magnitude epistasis because  $B_0 \rightarrow B_1$  mutation is beneficial on both locus A backgrounds. Sign epistasis between sweeping mutations at both loci is not observed in simulations (Weinreich et al. 2005).

cosegregate (Table 2). This occurs because waiting times to selective sweeps in Model 1 are often much longer than in Model 2 (compare Figs. 4A and 6A).

For the case of cosegregating sweeping mutations recombination is necessary to form double mutants, whereas double mutants form without recombination when two single mutations appear sequentially at different loci. In both epistatic and nonepistatic genetic systems, recombination accelerates adaptation in terms of

decreasing waiting times to substitutions and increasing segregating genetic variation (compare  $R = 0$  with  $R = 0.5$  in Table 3). Waiting times decrease because cosegregating beneficial mutations can recombine to form double mutants much faster. However, recombination does not seem to influence the fitness effect of sweeps (Table 3). In a multiplicative model, the double mutant fitness is simply a product of two independent single mutations. However, in epistatic networks the double mutant fitness may deviate due to functional epistasis. With epistasis and recombination, mutations are exposed to multiple segregating genetic backgrounds, which makes their fitness variable across backgrounds. This reduces their overall mean fitness (see also Gillespie 1974; Barton 1995). As a result, in the case of independent assortment ( $R = 0.5$ ), genetic variation increases much more dramatically in a multiplicative compared to an epistatic model (Table 3). Below we study what types of epistatic interactions evolve within epistatic gene networks during adaptive walks.

*Synergistic versus antagonistic epistasis between adaptive mutations*

Because we are interested in adaptive evolution, we study only epistatic interactions between mutations that undergo selective sweeps, ignoring functional epistasis between segregating deleterious mutations. Our results reveal rampant functional epistasis within epistatic gene networks (Fig. 7A, B). In Model 1, mutations typically show positive synergistic epistasis (31 synergistic, 3 antagonistic;  $\chi^2$ -test = 23.06,  $P < 1.6 \times 10^{-6}$ ; Fig. 7A). Thus in Model 1, epistasis typically increases the fitness advantage of double mutants (i.e., directional; Carter et al. 2005). This gives an extra boost to the fitness effects of sweeps in epistatic networks and hence facilitates their adaptation over the multiplicative model. On the other hand, in Model 2, mutations show both synergistic and negative antagonistic epistasis (23 synergistic, 33 antagonistic;  $\chi^2$ -test = 1.78,  $P = 0.18$ ; Fig. 7B). Hence, in Model 2, epistasis reduces as often as it improves the fitness advantage of double mutants (i.e., nondirectional). This implies that on average, epistatic interactions between sweeping mutations have little effect on the dynamics of epistatic networks in Model 2. Of course this does not diminish the importance of epistatic effects of a mutation with a fixed genetic background (see above).

*Relationship between fitness effects of adaptive mutations and functional epistasis*

To explain the difference in results between Model 1 and Model 2, recall that these models differ substantially in the average fitness effect of sweeps (compare Figs. 4B and 6B). In Model 1, single mutant genotypes have a relatively small fitness advantage over the ancestral genotype. When mutations combine to form a double mutant in Model 1, the selective advantage of the double mutant over single mutants is small without epistasis. If these

**Table 3.** Effects of recombination rate on adaptive walks on smoother fitness landscapes.

Waiting times				Fitness effects			No. of segregating genotypes		
Complete linkage ( $R=0$ )									
L	Multiplicative	Model 1	Model 2	Multiplicative	Model 1	Model 2	Multiplicative	Model 1	Model 2
2	463	466	353	0.014	0.021	0.025	13	10	11
5	374	330	179	0.017	0.039	0.044	27	16	23
10	311	260	109	0.019	0.038	0.069	47	24	42
20	221	569	92	0.017	0.041	0.111	91	29	71
30	190	1015	43	0.02	0.038	0.096	135	31	96
40	167	5035	38	0.02	0.038	0.14	182	32	120
Independent assortment ( $R=0.5$ )									
L	Multiplicative	Model 1	Model 2	Multiplicative	Model 1	Model 2	Multiplicative	Model 1	Model 2
2	342	373	365	0.016	0.025	0.032	16	12	14
5	136	286	118	0.017	0.032	0.041	103	23	50
10	59	300	32	0.015	0.045	0.064	1277	33	288
20	39	326	35	0.015	0.04	0.113	6467	35	529
30	19	1263	19	0.016	0.056	0.15	6136	31	319
40	10	3846	28	0.016	0.035	0.15	6764	32	466

Note: Means are shown. Parameters were identical to runs in Figure 4 (Model 1) and Figure 6 (Model 2) except for recombination rates.

mutations interact antagonistically with one another, the double mutant genotype is unlikely to be favored by selection. Thus when single mutations have small fitness effects, the only double mutant that can sweep will have either no epistasis or synergistic epistasis (see Fig. 7A). On the other hand, in Model 2, single mutations typically have large fitness advantages over the ancestral genotype. The double mutant will therefore have an even larger fitness advantage, thus allowing both synergistic and some antagonistic epistasis (unless antagonistic epistasis is very strong; see Fig. 7B). The argument that fitness effects of sweeping mutations determine the direction of epistasis is supported by a negative relationship between these properties in both Models 1 and 2 (see Fig. 7C, D). Larger effect mutations allow more antagonistic epistasis than smaller effect mutations.

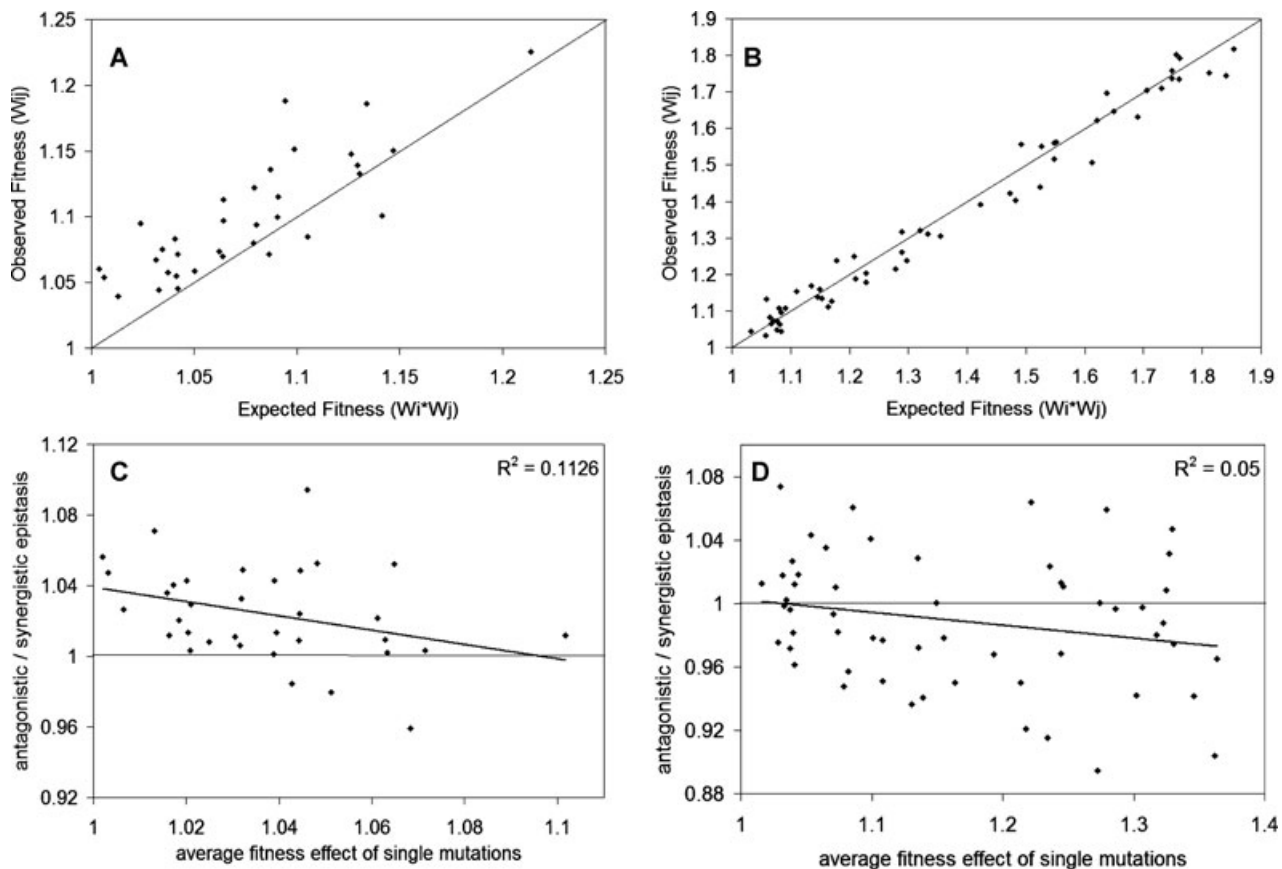
#### *Magnitude versus sign epistasis among adaptive mutations*

The epistatic interactions in the double mutant described above give rise to both magnitude and sign epistasis between sweeping mutations (Table 3). Importantly, because we only focus on sweeping double mutants, it will always be the case that the double mutant ( $A_1B_1$  in Table 3) is relatively beneficial compared to the ancestral genotype ( $A_0B_0$  in Table 3). This restricts the type of epistasis that can be observed. We find only magnitude epistasis between sweeping mutations in Model 2, whereas in Model 1 sweeping mutations show sign epistasis at one of the loci 28% of the time. This occurs because Model 2 always results in multi-

ple cosegregating beneficial mutations that are beneficial on their own as well in the presence of each other (Table 3). However, in Model 1, about half the time mutations sweep sequentially and this allows the second mutation to only be beneficial on the mutant background, not on the ancestral background of the other locus (Table 3). As expected, we do not find sign of epistasis at both loci because for this to occur both single mutations would have to be deleterious on their own, but beneficial together. This is equivalent to adaptation via peak shift, which is unlikely to occur in the model because single beneficial mutations dominate the dynamics of adaptive walks. Note however, that sign epistasis at both loci is possible among deleterious nonsweeping mutations.

#### **ADAPTIVE WALKS ON RUGGED FITNESS LANDSCAPES**

Finally, we study adaptive walks with a fixed mutation distribution, which results in uncorrelated adaptive landscapes in which the fitnesses of the parental alleles provide no information about the fitnesses of their single mutants (see Fig. 1). The overall adaptive walk of both epistatic genetic networks and a multiplicative model shows a similar pattern of diminishing fitness returns, where most rapid evolutionary change occurs early on and then slows down as the population approaches its local optimum (e.g., see Fig. 8A). This is consistent with the idea that epistatic and nonepistatic systems exhibit qualitatively similar evolutionary trajectories (Welch and Waxman 2005). However, despite this broad similarity, there are important differences.

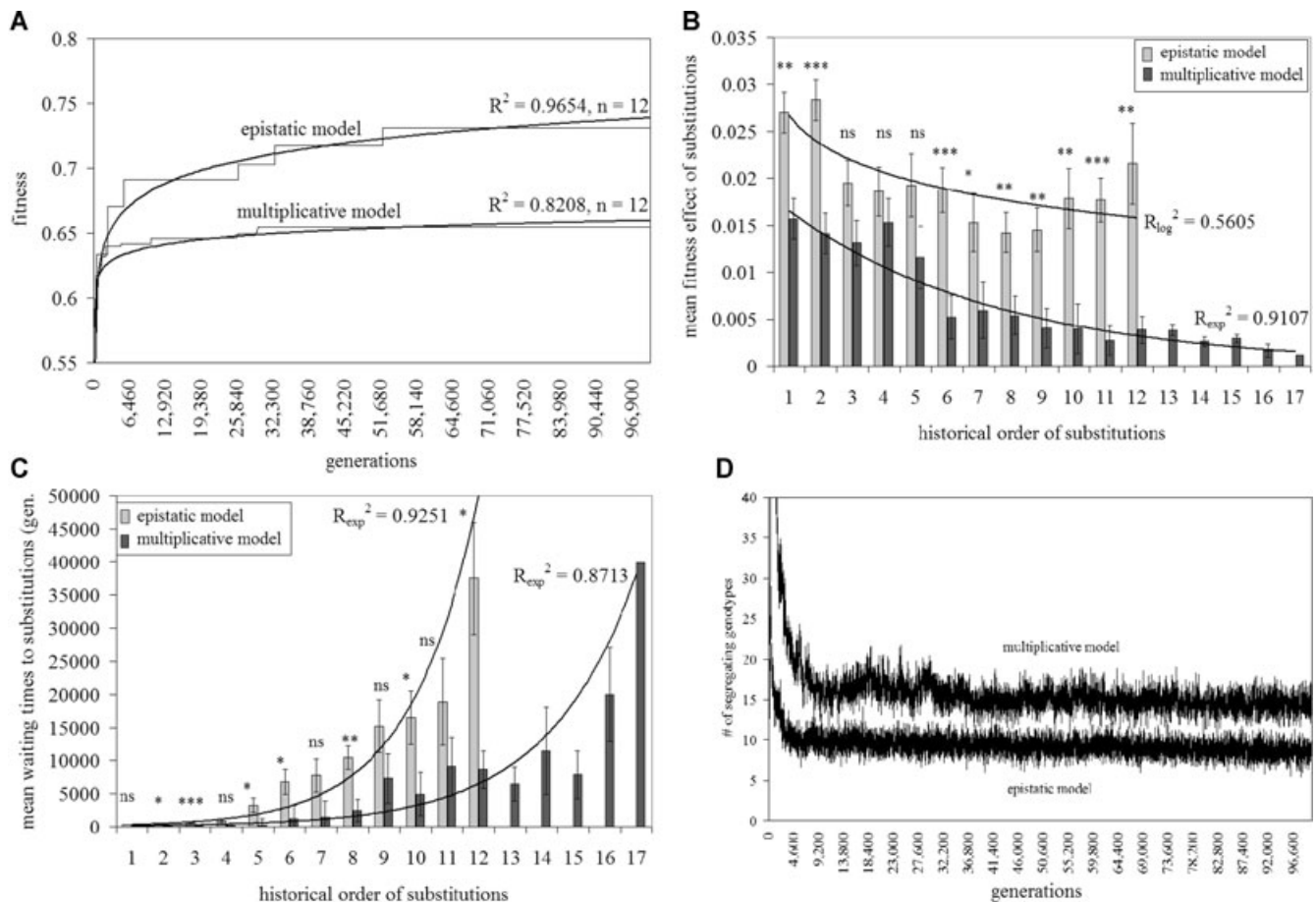


**Figure 7.** Fitness epistasis between sweeping mutations at different loci. Panels A and B describe the relationship between observed and expected (multiplicative) fitness for sweeping double-mutant beneficial genotypes in Model 1 (A) and in Model 2 (B). Fitness is relative to the ancestral genotype. The solid line represents the null expectation of pure multiplicative effects of two mutations. Deviations from this line indicate the presence of functional epistasis between mutations ( $W_{ij} \neq W_i W_j$ ; see text). Mutations are sampled equally from 2, 5, 10, 20, 30, and 40 locus systems with  $K = 100\%$  in simulation runs from Figures 4 and 6. Note that the scale of fitness effects of mutations is smaller for Model 1 than for Model 2 (see Figs. 4B and 6B). Panels C and D describe the relationship between average fitness effect of mutations that make up the sweeping double mutant genotypes and their epistatic interactions in both Model 1 (C) and Model 2 (D). The value 1 on the y-axis represents no epistasis (multiplicative fitness), and values above 1 designate synergistic epistasis, and below 1 antagonistic epistasis. Note scale differences.

First, consistent with Kauffman (1993) and adaptive walks on smoother fitness landscapes above, we once again observe that large and highly connected epistatic gene networks exhibit evolutionary constraint relative to a multiplicative model (i.e., they have fewer substitutions to local peaks and reach lower/less fit local peaks; data not shown). However, smaller and less connected networks may not necessarily exhibit these patterns as illustrated in a five-locus genetic system below (Fig. 8). On average a five-locus 100%  $K$  epistatic network reached higher local adaptive peaks compared to a five-locus multiplicative system ( $t$ -test:  $\hat{w}_{epistatic\ local\ peak} = 0.742$  (SE: 0.0057);  $\hat{w}_{multiplicative\ local\ peak} = 0.656$  (0.0057);  $df = 23$ ,  $P < 0.0001$ ; see Fig. 8A). This occurs because epistatic networks still favor substitutions of larger fitness effect compared to a multiplicative model ( $t$ -test: mean fitness effect<sub>epistatic</sub> = 0.02 (0.0009); mean fitness effect<sub>multiplicative</sub> = 0.007 (0.0009);  $df = 23$ ,  $P < 0.0001$ ), but do not have dramati-

cally longer waiting times to substitutions ( $t$ -test: mean waiting times<sub>epistatic</sub> = 6418 gen. (519); mean waiting times<sub>multiplicative</sub> = 4223 gen. (519);  $df = 23$ ,  $P < 0.0067$ ). Thus, similar to our results on smoother landscapes, larger fitness effects can compensate for longer waiting times in small epistatic networks, but not in larger networks.

Figure 8B shows the distribution of fitness effects of substitutions throughout the adaptive walk in the above example. In both models, early substitutions typically have a larger fitness effect than later ones, but this pattern is less pronounced in an epistatic model than in a multiplicative model. Indeed, only in the multiplicative model the fitness effects of sweeps are better described by an exponential distribution (consistent with Orr 2002, 2006; compare to  $R^2_{\log} = 0.864$ ). Fitness effects of sweeps in epistatic networks are better described by a log distribution (compare to  $R^2_{\text{exponential}} = 0.319$ ). This occurs because unlike the small fitness



**Figure 8.** Simulation dynamics of a five-locus genetic system on uncorrelated (rugged) fitness landscapes. (A) A log fit of average fitness trajectories of 12 simulation runs of an epistatic network model and a multiplicative model (thick lines;  $R^2$  fit is shown) and an example run of each model (thin lines). (B) Mean fitness effects of substitutions and (C) mean waiting times to substitutions in generations in the order with which substitutions appear during an adaptive walk for each model. The best-fit distributions are shown. Asterisks indicate significant difference between values for the two models for each substitution number (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ). Error bars represent standard errors. (D) Mean number of segregating genotypes accumulated in each model over the adaptive walk. All means were based on 12 replicate simulation runs of each model. For each run,  $L = 5$ ,  $K = 100\%$  (for epistatic model),  $\mu_{allelic\ effects} = 0.99$ ,  $\mu_{epistatic\ effects} = 0.99$ ,  $\sigma^2_{all\ effects} = 0.0004$ , mutation rate/locus =  $10^{-5}$ ,  $N_{adults} = 50,000$ ,  $N_{zygotes} = 10^5$ , ( $\theta = 1$ ), rec. rate/all loci pairs = 0.25. Note parameters are identical to Figure 4.

steps in a multiplicative model, epistatic networks improve fitness using large fitness jumps all the way to the local peak (see Fig. 8B). This pattern holds if fitnesses are ranked from highest to lowest.

Interestingly, even though a five-locus epistatic network reached higher local peaks and exhibited substitutions of larger fitness effects, it did so still using fewer substitutions compared to a multiplicative model ( $t$ -test: no. of substitutions<sub>epistatic</sub> = 9.8 (0.582); no. of substitutions<sub>multiplicative</sub> = 14 (0.582);  $df = 23$ ,  $P < 0.0001$ ). Thus it seems that regardless of the size of the epistatic gene network, these models exhibit shorter adaptive walks than a multiplicative model.

Figure 8C shows the distribution of waiting times to substitutions throughout the adaptive walk. The multiplicative model generates many more short waiting time events during early stages

of the walk compared to an epistatic model. However, both models follow an exponential distribution and show similar waiting times toward the end of the walk (Fig. 8C). Thus the differences we observe in average waiting times between epistatic and nonepistatic systems are driven by substitutions that occur early in the adaptive process. Finally, Figure 8D shows that epistatic networks segregate less genetic variation compared to a multiplicative model, which is again consistent with results from smooth fitness landscapes.

## Discussion

Recently, Orr (2006) noted that the theoretical study of the genetics of adaptation is in its infancy. This is especially the case for models of adaptation that incorporate epistatic genetic

architecture (for recent models see Johnson and Porter 2000; Porter and Johnson 2002; Carter et al. 2005; Hansen et al. 2006). In the present article we explored long-term adaptation using one type of epistatic genetic architecture, namely the *NK* framework of Kauffman (1993) that emphasizes a multilocus epistatic gene network approach, where epistatic connectivity among loci plays a key role in dynamics.

### WAITING TIMES TO SUBSTITUTIONS IN EPISTATIC VERSUS NONEPISTATIC MODELS

Our results revealed that regardless of whether mutations are drawn from a dynamic or fixed mutational distribution throughout the adaptive walk (i.e., smooth or rugged adaptive landscapes, respectively), epistatic genetic networks can exhibit longer genome-wide waiting times to adaptive substitutions compared to a nonepistatic multiplicative model. This result is consistent with similar findings of Kauffman (1993: p. 58) and suggests that highly epistatically connected networks will often show prolonged evolutionary stasis, characterized by purifying selection of deleterious mutants (also see Hansen and Houle 2004; Welch and Waxman 2005).

We also uncovered several important novel insights. First, this result depends on average epistatic effects of mutations being selectively deleterious compared to those of the parental allele throughout the adaptive walk (referred to as “Model 1”). This is why adaptive walks with a fixed mutational distribution always generate longer waiting times in epistatic systems (Kauffman 1993). Under a dynamic mutation distribution, however, average epistatic effects of mutations can remain close to selective neutrality throughout most of the adaptive walk. In this case, waiting times to substitutions in epistatic and nonepistatic genetic systems become similar and can even be shorter in the former (see Fig. 6). Second, we reveal that genome-wide waiting times in epistatic gene networks are characterized by a hyperbolic function as the number of loci in the network increases (see Figs. 4 and 6). Finally, waiting times during adaptive walks are well characterized by an exponential distribution toward local fitness peaks, regardless of genetic architecture (see Fig. 8C).

### FITNESS EFFECTS OF SUBSTITUTIONS IN EPISTATIC VERSUS NONEPISTATIC MODELS

Recently, Welch and Waxman (2005) noted that epistatic connectivity in *NK* gene networks is intimately linked to fitness effects of novel mutations. This occurs because the fitnesses of mutations are modified by alleles at *K* other loci within a given genotype (i.e., epistatic effects). This assumption increases the variance of mutational effects in epistatic relative to nonepistatic systems. We studied the effect of such mutational distributions on the evolution of fitness effects of sweeps during adaptive walks. Our results indicate that under most “strong selection” scenarios in which the

variance of mutational effects is assumed to be reasonably high, fitness effects of substitutions are larger in epistatic relative to nonepistatic models and increase with larger and more connected networks (see Figs. 4 and 6). However, under low mutational variance (approaching “weak selection” scenarios), fitness effects of substitutions may be similar or even lower in epistatic relative to nonepistatic systems (see Table 1).

We also find that adaptive walks on rugged fitness landscapes are characterized by diminishing fitness returns in which early substitutions have larger fitness effects than later ones (consistent with Gillespie 1984; Kauffman 1993; Orr 2002, 2006; Welch and Waxman 2005). However, this pattern is much less pronounced in epistatic systems. Indeed, fitness effects in epistatic systems are better described by a log distribution instead of an exponential distribution toward local peaks (see Fig. 8B and text results). Interestingly, this tends to disagree with several empirical studies of distribution of fitness effects among beneficial mutations before selection (e.g., Kassen and Bataillon 2006). However, it is still unclear how general such patterns are and whether they will hold after selection has taken place. Future work is required to test our predictions.

### THE RATE OF EVOLUTION IN EPISTATIC VERSUS NONEPISTATIC MODELS

We explored how waiting times and fitness effects of substitutions influence the overall relative rate of adaptation in epistatic versus nonepistatic genetic models. Interestingly, we discovered that larger effect mutations of epistatic systems are able to compensate for their longer waiting times, but only when networks are not too large. As the size and connectivity of networks increases, epistatic models typically lead to evolutionary constraint (see Fig. 4D; consistent with Kauffman 1993). Nevertheless, when epistatic networks are either small or fairly sparsely connected (such as power-law networks, Siegal et al. 2007), or both, evolution may be more rapid than in a nonepistatic model even when the average epistatic effects are substantially deleterious.

### THE ROLE OF GENETIC VARIATION AND RECOMBINATION IN ADAPTIVE WALKS

Perhaps the greatest weakness of the traditional *NK* model is that it assumes a SSWM scenario, thereby forcing substitutions to occur sequentially on a fixed genetic background (Gillespie 1984). It has been previously emphasized that this model still exhibits functional epistasis between a mutation and alleles that were fixed beforehand (Brodie 2000; Phillips 1996; Phillips et al. 2000; Weinreich et al. 2005). However, the *NK* model had clearly missed several important aspects of epistasis in nature. Epistasis may cause mutations to have variable fitness effects when exposed to multiple genetic backgrounds (i.e., fitness is context-dependent). Similarly, in the presence of epistasis, different mutations may

recombine to form double-mutant genotypes that can deviate in fitness from their independent contributions (Wright 1932; Brodie 2000; Phillips et al. 2000; Wade 2000; Weinreich et al. 2005).

By allowing mutations to segregate in our model, we let epistasis determine dynamics in the context of multiple genetic backgrounds (e.g., see Fig. 2). However, we found that epistatic gene networks segregate less genetic variation relative to a nonepistatic model during adaptive walks, regardless of whether epistatic effects are on average deleterious or neutral (although it is more pervasive in the former) and whether landscapes are smooth or rugged. This occurs because in the presence of epistasis, most mutations are incompatible with established alleles in the population, which does not occur in a nonepistatic system (see also Mayr 1963: pp. 293–294; Gavrillets and de Jong 1993; Hermisson et al. 2003). Thus our findings point to an interesting paradox, where on one hand, the role of epistatic interactions has particular relevance when there are multiple genetic backgrounds, but on the other hand, epistatic gene networks tend to reduce that role by preventing many mutations from segregating in the population. Note that this result differs from that of Whitlock et al. (1995) that emphasized that functional epistasis may not necessarily generate much statistical epistasis. Our results show that epistatic gene networks tend to reduce number of segregating genotypes in the population and thus curtail genetic variance in general. Curiously, this makes the analytical framework lacking variation a reasonable approximation of exact simulation results, especially when average epistatic effects of mutations are deleterious.

Nevertheless, a substantial number of mutations can still cosegregate during their selective sweeps (especially in Model 2) and can recombine to form double-mutant genotypes. Higher rates of recombination allow mutations to explore the genetic background space faster by forming double-mutant genotypes during adaptive walks (Table 2; see also Malmberg 1977; Misevic et al. 2006). This causes mutations to have variable fitness effects among genetic backgrounds. Compared to a multiplicative model, this forces epistatic genetic networks to segregate relatively even less genetic variation (Table 3; also see Gillespie 1974; Barton 1995).

#### THE EVOLUTION OF EPISTASIS DURING ADAPTIVE WALKS IN EPISTATIC GENETIC NETWORKS

Functional epistasis between sweeping mutations is rampant within our simulated adapting populations, revealing the evolution of positive and negative epistatic deviations in fitness (see Fig. 7). Fitness effects of sweeping mutations have a direct impact on whether mutations interact positively or negatively. Adaptive mutations with small fitness effects typically interact positively and thus directionally, whereas mutations with large fitness effects can interact positively or negatively and thus nondirectionally (see Fig. 7). These predictions assume that all possible epistatic inter-

actions are initially possible between any pair of mutations. Carter et al. (2005) argued that only when epistasis is directional does it have any major effect on response to directional selection. Here we show that this epistatic directionality is a characteristic of sweeping mutations with small fitness effects.

The relationship between the fitness effects of sweeping mutations and the direction of epistatic interactions has yet to be tested in any study that we are aware of. This can be done in long-term selection experiments where the fitness effects of mutations can be well characterized (Lenski and Travisano 1994; Elena and Lenski 2001, 2003). These results also predict that any biological condition that determines the fitness effect of adaptive mutations should also affect the direction of epistasis in nature. Even though functional epistasis between single mutations has recently received much empirical attention (reviewed by Sanjuan and Elena 2006), most studies focus on deleterious single mutations. We are aware of only a single study that looked at epistasis between adaptive mutations, which showed predominantly negative antagonistic epistasis (Sanjuan et al. 2004; see also Bahannan et al. 1999).

Predictions from our model differ in important respects from other recent attempts to predict the direction of epistasis in nature. Perhaps the most important difference is that, similar to empirical work, most theoretical studies focus on epistasis between deleterious mutations because this is of special interest for understanding the evolution of sex, genetic load, and robustness to genetic perturbations (e.g., Wilke and Adami 2001; Azevedo et al. 2006; Beerwinkel et al. 2007; Desai et al. 2007). As for adaptive mutations, Martin et al. (2007) have recently used Fisher's geometric model of adaptation to study how the mapping from phenotype to fitness can generate epistasis (see also Brodie 2000). However, in our model, epistasis does not occur as a result of a mismatch between phenotype and fitness maps, but is due to explicit epistatic effects for fitness between single mutations.

#### ASSUMPTIONS AND CAVEATS

The task of understanding the role of gene interactions in evolutionary biology is extraordinarily complex. Even though we have relaxed several important assumptions of the *NK* model, our approach most importantly relies on the concept of epistatic connectivity. This is to say that we assume that mutations interact with alleles at other loci within the genome, regardless of other segregating genetic backgrounds. The biological relevance of this concept is clear; mutations whose fitness is modified by alleles at many other loci in larger and more epistatically connected networks should experience different selection pressures and ultimately different evolutionary fates compared to mutations whose fitness is modified by alleles at only a few other loci in smaller and poorly connected gene networks. Whether this assumption is biologically realistic is open to empirical test (Kauffman and

Weinberger 1989; Fontana et al. 1993; Jeong et al. 2001; Fraser et al. 2002). Approaches that have been taken to study the effects of molecular connectivity on fitness (e.g., Featherstone and Broadie 2002; Hahn et al. 2004) can be used in the future to understand how the fitness of mutations depends on epistatic connectivity of gene networks (e.g., Fedorowicz et al. 1998; Sambandan et al. 2006).

Further, for the present analysis we have concentrated on adaptive walks assuming a “strong selection” scenario, where population sizes are large and mutations have a high chance of selective sweeps at all loci in the genetic network. Recently, Welch and Waxman (2005) argued that the epistatic *NK* model and a nonepistatic House of Cards model generate essentially identical evolutionary behavior. However, these similarities mostly apply to weak selection or nearly neutral scenarios with no genetic variation (see also Ohta 1997, 1998). Future work should focus on adaptive walks with weak selection scenarios in the presence of segregating genetic variation.

When adaptive walks are driven by strong selection, the only similarity between epistatic and nonepistatic models is that the adapting population eventually reaches fitness plateaus over time with fixed mutational distributions (Fisher 1930; Gillespie 1984; Welch and Waxman 2005; Orr 2006). We have described many important differences in how epistatic and nonepistatic genetic systems adapt under fixed and dynamic mutational distributions (see also Welch and Waxman 2005).

More generally, our model assumed that adaptive walks occur within a single panmictic population with no structure. The importance of population structure in determining the role of epistatic genetic architectures for adaptation is a critical area for future studies. Population structure should make the fitness of mutations highly variable across genetically differentiated demes (e.g., Wright 1949, 1969; Goodnight 1995, 2000; Whitlock et al. 1995; Brodie 2000; Johnson 2000; Wade 2000, 2002). Future work may especially benefit by combining weak selection with population structure to see how epistasis will affect adaptive walks.

#### ACKNOWLEDGMENTS

We are grateful to D. Futuyama, C. Goodnight, E. Haig, N. Johnson, M. Travisano, and G. Wagner and two anonymous reviewers for helpful comments on earlier draft of this manuscript. We also thank W. Gharaibeh and J. Rohlf for statistical help on partitioning of fitness variance. This study was supported by Stony Brook University. This article is 1172 contributed by the Stony Brook Ecology and Evolution Graduate Program.

#### LITERATURE CITED

Azevedo, B. R., R. Lohaus, S. Srinivasan, K. K. Dang, and C. L. Burch. 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* 440:87–90.

Bohannon, B. J. M., M. Travisano, and R. E. Lenski. 1999. Epistatic interactions can lower the cost of resistance to multiple consumers. *Evolution* 53:292–295.

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

Beerenwinkel, N., L. Pachter, B. Sturmfels, S. F. Elena, and R. E. Lenski. 2007. Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC Evol. Biol.* 7:60.

Brodie, E. D. I. 2000. Why evolutionary genetics does not always add up. Pp. 3–19 in J. B. Wolf, E. D. B., and Wade, M. J., eds. *Epistasis and the evolutionary process*. Oxford Univ. Press, Oxford.

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

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

Chambers, D. M., J. Peters, and C. M. Abbott. 1998. The lethal mutation of the mouse wasted (*wst*) is a deletion that abolishes expression of a tissue-specific isoform of translation elongation factor 1 $\alpha$ , encoded by the *Eef1a2* gene. *Proc. Natl. Acad. Sci. USA* 95:4463–4468.

Cheverud, J. M., and E. J. Routman. 1995. Epistasis and its contribution to genetic variance components. *Genetics* 139:1455–1461.

Desai, M. M., Weissman, D., and Feldman, M. W. 2007. Evolution can favor antagonistic epistasis. *Genetics* 177:1001–1010.

Elena, S. F., and R. E. Lenski. 2001. Epistasis between new mutations and genetic background and a test of genetic canalization. *Evolution* 55:1746–1752.

———. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* 4:457–469.

Ewens, W. J. 1979. *Mathematical population genetics*. Springer, New York.

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8:610–618.

Eyre-Walker, A., M. Woolfit, and T. Phelps. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173:891–900.

Falconer, D. S. and T. Mackay. 1996. *Introduction to quantitative genetics*. Longman Group Ltd, Harlow, UK.

Featherstone, D. E., and K. Broadie. 2002. Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *Bioessays* 24:267–274.

Fedorowicz, G. M., J. D. Fry, R. R. Anholt, and T. F. Mackay. 1998. Epistatic interactions between smell-impaired loci in *Drosophila melanogaster*. *Genetics* 148:1885–1891.

Felsenstein, J. 1965. The effect of linkage on directional selection. *Genetics* 52:349–363.

Fisher, R. A. 1930. *The genetical theory of natural selection*. Oxford Univ. Press, Oxford.

Fontana, W., P. F. Stadler, E. G. Bornberg-Bauer, T. Griesmacher, I. L. Hofacker, M. Tacker, P. Tarazona, E. D. Weinberger, and P. Schuster. 1993. RdA folding and combinatorial landscapes. *Phys. Rev. E* 47:2083–2099.

Franklin, I., and R. C. Lewontin. 1970. Is the gene the unit of selection? *Genetics* 65:707–734.

Fraser, H. B., A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman. 2002. Evolutionary rate in the protein interaction network. *Science* 296:750–752.

Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton Univ. Press, Princeton NJ.

Gavrilets, S., and G. de Jong. 1993. Pleiotropic models of polygenic variation, stabilizing selection, and epistasis. *Genetics* 134:609–625.

Gillespie, J. H. 1974. Natural selection for within-generation variance in offspring number. *Genetics* 76:601–606.

———. 1983. A simple stochastic gene substitution model. *Theor. Popul. Biol.* 23:202–215.

- . 1984. Molecular evolution over the mutational landscape. *Evolution* 38:1116–1139.
- . 1991. The causes of molecular evolution. Oxford Univ. Press, Oxford.
- . 1994. Substitution processes in molecular evolution. III. Deleterious alleles. *Genetics* 138:943–952.
- Gimelfarb, A. 1989. Genotypic variation for a quantitative character maintained understabilizing selection without mutations: epistasis. *Genetics* 123:217–227.
- Goodnight, C. J. 1995. Epistasis and the increase in additive genetic variance: implications for phase 1 of Wright's shifting balance process. *Evolution* 49:502–511.
- . 2000. Modeling gene interactions in structured populations. Pp. 129–145 in J. B. Wolf, E. D., B. and Wade, M. J., eds. *Epistasis and the evolutionary process*. Oxford Univ. Press, Oxford.
- Hahn, M. W., G. C. Conant, and A. Wagner. 2004. Molecular evolution in large genetic networks: does connectivity equal constraint? *J. Mol. Evol.* 58:203–211.
- Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection. *Proc. Camb. Philos. Soc.* 23:838–844.
- Hansen, T. F., and G. P. Wagner. 2001. Modeling genetic architecture: a multilinear theory of gene interaction. *Theor. Popul. Biol.* 59:61–86.
- Hansen, T. F. and D. Houle. 2004. Evolvability, stabilizing selection, and the problem of stasis. Pp. 130–150 in M. Pigliucci, Pigliucci, K., ed. *Phenotypic integration: studying the ecology and evolution of complex phenotypes*. Oxford Univ. Press, Oxford, UK.
- Hansen, T. F., J. M. Alvarez-Castro, A. J. Carter, J. Hermisson, and G. P. Wagner. 2006. Evolution of genetic architecture under directional selection. *Evolution* 60:1523–1536.
- Hartl, D. L., D. D. Dykhuizen, and K. M. Dean. 1985. Limits of adaptation. The evolution of selective neutrality. *Genetics* 111:655–674.
- Hermisson, J., T. F. Hansen, and G. P. Wagner. 2003. Epistasis in polygenic traits and the evolution of genetic architecture under stabilizing selection. *Am. Nat.* 161:708–734.
- Jeong, H., S. P. Mason, A. L. Barabasi, and Z. N. Oltvai. 2001. Lethality and centrality in protein networks. *Nature* 411:41–42.
- Johnson, N. A. 2000. Gene interactions and the origin of species. Pp. 197–212 in J. B. Wolf, E. D., B. and Wade, M. J., eds. *Epistasis and the evolutionary process*. Oxford Press, Oxford UK.
- Johnson, N. A., and Porter, A. H. 2000. Rapid speciation via parallel, directional selection on regulatory genetic pathways. *J. Theor. Biol.* 205:527–542
- Kassen, R., and T. Bataillon. 2006. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat. Genet.* 38:484–488.
- Kauffman, S. A. 1993. The origins of order: self-organization and selection in evolution. Oxford Univ. Press, Oxford, UK.
- Kauffman, S. A., and S. Levin. 1987. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* 128:11–45.
- Kauffman, S. A., and Weinberger. 1989. The NK model of rugged fitness landscapes and its application to the maturation of the immune response. *J. Theor. Biol.* 141:211–245.
- Kimura, M. 1954. Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. *Genetics* 39:280–295.
- . 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713–719.
- Kingman, J. F. C., 1978. A simple model for the balance between selection and mutation. *J. Appl. Prob.* 15:1–12.
- Lenski, R. E., and Travisano M. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. USA* 91:6808–6814.
- Lewontin, R. C. 1974. The genetic basis of evolutionary change. Columbia Univ. Press, NY.
- . 1978. Adaptation. *Sci. Am.* 239:212–218, 220, 222 passim.
- Lynch, M., and B. Walsh. 1998. *Genetics and analysis of quantitative traits*. Sinauer Assoc., Inc. Sunderland, MA.
- Mackay, T. F. 2001. Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.* 2:11–20.
- Malmberg, R. L. 1977. Evolution of epistasis and advantage of recombination in populations of bacteriophage T4. *Genetics* 86:607–621.
- Malmberg, R. L., and R. Mauricio. 2005. QTL-based evidence for the role of epistasis in evolution. *Genet. Res.* 86:89–95.
- Martin, G., S. F. Elena, and T. Lenormand. 2007. Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat. Genet.* 39:555–560
- Maynard Smith, J. 1970. Natural selection and the concept of a protein space. *Nature* 225:563–564.
- Mayr, E. 1959. Where are we? Cold Spring Harbor Symp. Quant. Biol. 24:1–14.
- . 1963. *Animal species and evolution*. Belknap Press, Cambridge, MA.
- Misevic, D., Ofria, C., and Lenski, R. E. 2006. Sexual reproduction reshapes the genetic architecture of digital organisms. *Proc. R. Soc. Lond. B.* 273:457–464.
- Nielsen, R., and Z. Yang. 2003. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol. Biol. Evol.* 20:1231–1239.
- Ohta, T. 1977. Extension of the neutral mutation drift hypothesis. Pp. 148–167 in M. Kimura ed., *Molecular evolution and polymorphism*. National Institute of Genetics, Mishima.
- . 1997. Role of random genetic drift in the evolution of interactive systems. *J. Mol. Evol.* 44:S9–S14.
- . 1998. Evolution by nearly-neutral mutations. *Genetica* 102/103: 83–90.
- Orr, A. H. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52:935–949.
- . 2002. The population genetics of adaptation: the adaptation of DNA sequences. *Evolution* 56:1317–1330.
- . 2006. The population genetics of adaptation on correlated fitness landscapes: the block model. *Evolution* 60:1113–1124.
- Phillips, P. C. 1996. Waiting for a compensatory mutation: phase zero of the shifting balance process. *Genetical Research* 67:271–283.
- . 1998. The language of gene interaction. *Genetics* 149:1167–1171.
- Phillips, P. C., Otto, S. P. and Whitlock, M. C. 2000. Beyond the average, the evolutionary importance of gene interactions and variability of epistatic effects. Pp. 20–38 in J. B. Wolf, E. D. Brodie III and Wade, M. J., ed. *Epistasis and the evolutionary process*. Oxford Univ, Oxford UK.
- Porter, A. H., and Johnson, N. A. 2002. Speciation despite gene flow when developmental pathways evolve. *Evolution* 56:2103–2111
- Sambandan, D., A. Yamamoto, J. J. Fanara, T. F. Mackay, and R. R. Anholt. 2006. Dynamic genetic interactions determine odor-guided behavior in *Drosophila melanogaster*. *Genetics* 174:1349–1363.
- Sanjuan, R., and S. F. Elena. 2006. Epistasis correlates to genomic complexity. *Proc. Natl. Acad. Sci. USA* 103:14402–14405.
- Sanjuan, R., A. Moya, and S. F. Elena. 2004. The contribution of epistasis to the architecture of fitness in an RdA virus. *PdAS* 101:15,376–15379.
- Siegal, M. L., D. E. Promislow, and A. Bergman. 2007. Functional and evolutionary inference in gene networks: does topology matter? *Genetica* 129:83–103.
- Sokal, R. R., and J. F. Rohlf. 1995. *Biometry*. W. H. Freeman and Company, New York.
- Threadgill, D. W., A. A. Dlugosz, L. A. Hansen, T. Tennenbaum, U. Lichti, D. Yee, C. LaMantia, T. Mourton, K. Herrup, R. C. Harris, *et al.* 1995.

Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science* 269:230–234.

van Swinderen, B., and R. J. Greenspan. 2005. Flexibility in a gene network affecting a simple behavior in *Drosophila melanogaster*. *Genetics* 169:2151–2163.

Wade, M. J. 1992. Sewall Wright: gene interaction and the shifting balance theory. Pp. 35–62 in D. J. Futuyma and J. Antonovics, ed. *Oxford surveys in evolutionary biology*. Oxford Univ. Press, Oxford.

———. 2000. Epistasis as a genetic constraint within populations and an accelerant of adaptive divergence among them. Pp. 20–38 in J. B. Wolf, E. D. Brodie III and Wade, M. J., ed. *Epistasis and the evolutionary process*. Oxford Univ. Press, Oxford, UK.

———. 2002. A gene's eye view of epistasis, selection and speciation. *J. Evol. Biol.* 15:337–346.

Wagner, G. 2002. To epistasis—and beyond. *Evolution* 56:852–855.

Weinberger, E. D. 1990. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biol. Cybern.* 63:325–336.

Weinreich, D. M. 2005. The rank ordering of genotypic fitness values predicts genetic constraint on natural selection on landscapes lacking sign epistasis. *Genetics* 171:1397–1405.

Weinreich, D. M., and L. Chao. 2005. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution* 59:1175–1182.

Weinreich, D. M., R. A. Watson, and L. Chao. 2005. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59:1165–1174.

Welch, J. J. and Waxman, D. 2005. The NK model and population genetics. *J. Theor. Biol.* 234:329–340.

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

Whitlock, M. C., P. C. Phillips, F. B.-G. Moore, and S. J. Tonsor. 1995. Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* 26:601–629.

Wright, S. 1931. *Evolution in Mendelian populations*. *Genetics* 16:97–159.

———. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth Inter. Congr. Genet.* 1:356–366.

———. 1949. *Adaptation and selection*. Pp. 365–389 in G. G. S. G. L. Jepsen, and E. Mayr, ed. *Genetics, paleontology and evolution*. Princeton Univ. Press, Princeton, NJ.

———. 1969. *Evolution and the genetics of populations*. Univ. Chicago Press, Chicago.

Zhivotovskiy, L. A. and Gavrilits, S. 1992. Quantitative variability and multilocus polymorphism under stabilizing selection. *Theor. Popul. Biol.* 42:254–283.

Associate Editor: M. Travisano

## Appendix of Additional Monte-Carlo Simulations

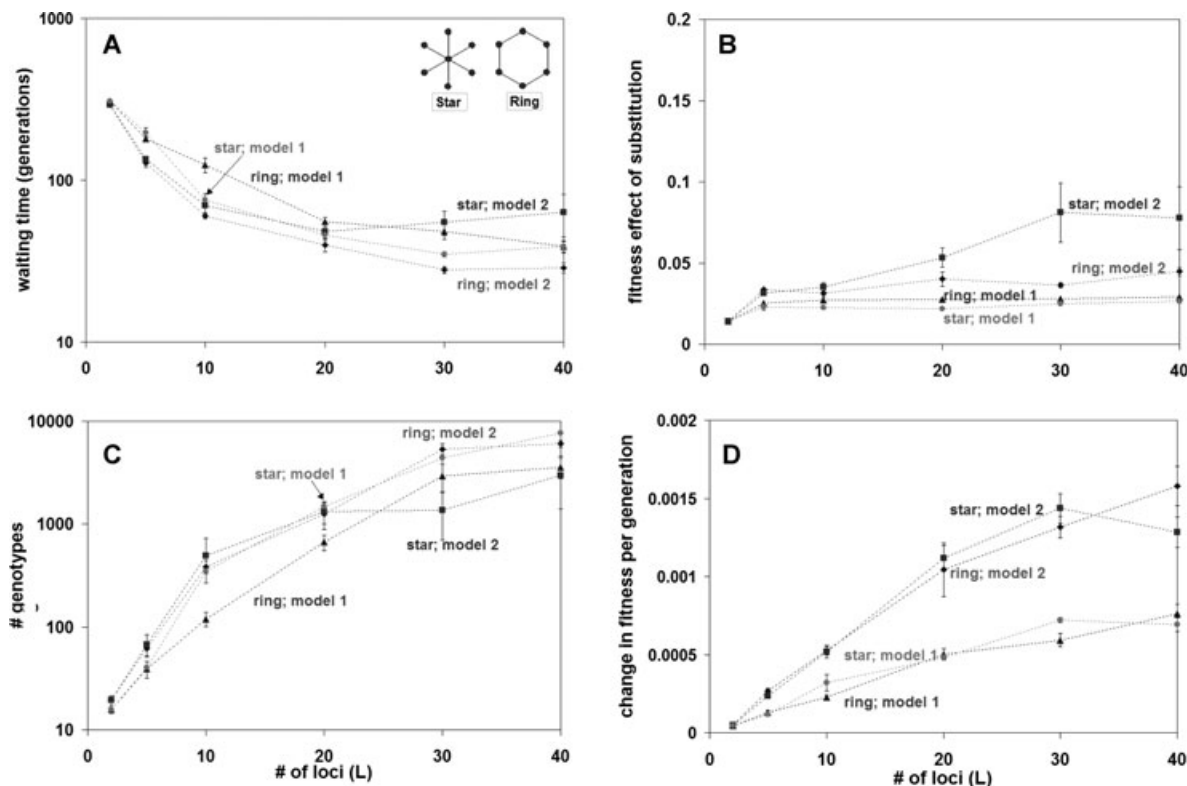
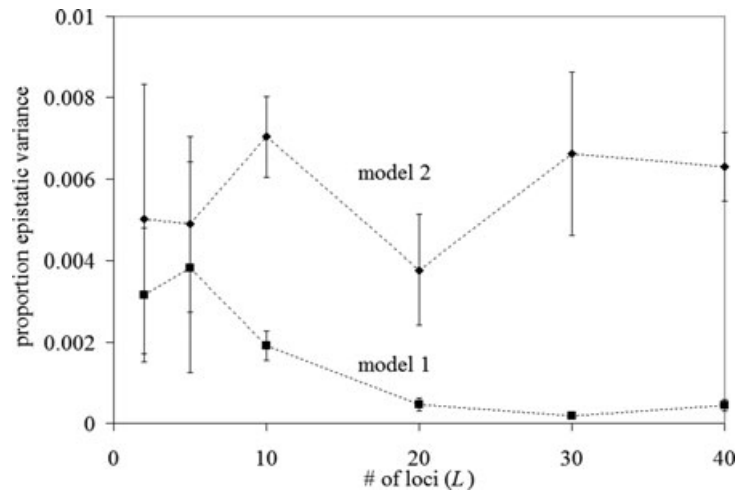


Figure A1. The effect of gene network topology on the statistics of adaptive walks when mutations exhibit on average either deleterious (Model 1) or neutral epistatic effects (Model 2). In particular, two additional topologies are explored explicitly, a ring topology, where only mutations of neighboring genes interact, and a star topology, where only mutations between one “hub” gene and peripheral genes interact (see inset in (A)). The plots show: average waiting times to genome-wide sweeping beneficial mutations on a log scale (A), average fitness effects of genome-wide sweeping beneficial mutations (B), average number of segregating genotypes per generation in log scale (C), and average rate of evolution (D) as a function of the size (x-axis) of a gene network. All other parameters are identical to Figure 4.



**Figure A2.** Approximation of average proportion fitness variance among segregating genotypes explained by epistatic effects between alleles (Cheverud and Routman 1995; Falconer and Mackay 1996) for Model 1 (average deleterious epistasis) and Model 2 (average neutral epistasis). Five replicate runs determine the standard errors for each scenario above. The epistatic variance is calculated using the Generalized Least Squares method of General Linear Model (e.g., Lynch and Walsh 1998). This approach assumes that the genotypic fitness is a linear function of allelic and epistatic effects. All genotypic fitnesses are converted into a log-linear scale (Sokal and Rohlf 1995). The vector for genotypic observations in matrix form is  $y = X\beta$ , where  $y$  are genotypic fitness observations,  $X$  is design matrix, and  $\beta$  is a vector of parameters to be estimated using  $\beta = (X^T X)^{-1} X^T y$ . If environmental variance in fitnesses among genotypes = 0, the full design matrix explains all variance in fitness. The vector of parameters ( $\beta$ ) is calculated only using the design matrix with allelic effects. Sum of squares due to epistatic interactions ( $SS_{\text{epistasis}}$ ) are then deviations of observed fitness values from those predicted by the allelic effects model. Variance or Means Squares ( $MS_{\text{epistasis}} = SS_{\text{epistasis}}/\text{sample size}$ ). The proportion of variance due to epistasis =  $MS_{\text{epistasis}}/\text{total fitness variance among genotypes}$  ( $\log(MS_{\text{total}})$ ). Note the exact proportion cannot be estimated using this method given our unbalanced design (e.g., Lynch and Walsh 1998). However, the magnitude of epistatic variance and the relative difference between Models 1 and 2 is informative (J. Rohlfs, pers. comm.).