

# Analysis of the Fitness Effect of Compensatory Mutations

Liqing Zhang

Department of Computer Science

Virginia Polytechnic Institute & State University

Blacksburg, VA 24061

Layne T. Watson

Departments of Computer Science and Mathematics

Virginia Polytechnic Institute & State University

Blacksburg, VA 24061

January 5, 2008

## Abstract

We extend our previous work on the fitness effect of the fixation of deleterious mutations on a population by incorporating the effect of compensatory mutations. Compensatory mutations are important in the sense that they make the deleterious mutations less deleterious, thus reducing the genetic load of the population. The essential phenomenon underlying compensatory mutations is the nonindependence of mutations in biological systems. Therefore, it is an important phenomenon that cannot be ignored when considering the fixation and fitness effect of deleterious mutations. Since having compensatory mutations essentially changes the distributional shapes of deleterious mutations, we can consider the effect of compensatory mutations by comparing two distributions where one distribution reflects the reduced fitness effects of deleterious mutations with the influence of compensatory mutations. We compare different distributions of deleterious mutations without compensatory mutations to those with compensatory mutations, and study the effect of population sizes, the shape of the distribution, and the mutation rates of the population on the total fitness reduction of the population.

**Key words:** effective population size, compensatory mutation, generalized Riemann zeta function, incomplete Gamma function.

# INTRODUCTION

The distribution of the fitness effects of spontaneous mutations is an important subject that has many implications. For example, in conservation biology, knowing the distribution of the fitness effects of naturally occurring mutations in a species will help researchers understand the speed of fitness decay of the population due to the fixation of deleterious mutations and help theoretical estimation of the critical population size that is required to prevent the so-called “mutational meltdown” and thus achieve the conservation goal (Lande, 1998; Lynch *et al.*, 1993; Whitlock, 2000). The distribution of mutation fitness effects also has an important effect on the amount of genetic load that human and other species have to bear. In fact, it is the calculation of the genetic load that led population geneticists to realize that despite the fact that the detailed distributional shape of mutation fitness effects harbored in the population is unknown, most of the mutations cannot be too deleterious, because otherwise the genetic load would have been so unbearable that we and other species would not have been able to survive (Kimura, 1968).

The commonly accepted notion is that most of the spontaneous mutations are deleterious (i.e., have negative fitness effects on the survival of the individuals who carry them) and only a small fraction of them are beneficial. Mutagenesis and mutation-accumulation experiments suggest that the majority of deleterious mutations have very small fitness effects, with an average effect of only about 1% or less (Davies *et al.*, 1999; Estes *et al.*, 2004; Keightley, 1994; Vassilieva *et al.*, 2000). To date, there has been a number of studies computationally estimating the distributional shapes of the fitness effects of deleterious mutations in different species (Eyre-Walker *et al.*, 2006; Nielsen and Yang, 2003; Piganeau and Eyre-Walker, 2003; Sanjuan *et al.*, 2004; Yampolsky *et al.*, 2005). No single continuous distribution type has been found to be descriptive for all observations in different species. So far, a variety of continuous distributions (some of which are a special case of others)

have been suggested and applied in the literature to quantify the fitness effect of deleterious mutations, including normal, exponential, gamma, and log-normal distributions.

However, compared with the research directed towards characterizing deleterious mutations, an important class of mutations, compensatory mutations, is relatively neglected. These mutations are also deleterious by themselves, however, they can compensate for the negative effect of other deleterious mutations when they appear together with the deleterious mutations. Empirical studies of compensatory mutations are largely owing to the studies of RNA structures. It is well known that the folding of RNA molecules into the stem-loop structure requires base pair matching in the stem part of the molecule, and mutations occurring to one segment of the stem part will disrupt the matching, and therefore, have a deleterious effect on the folding and stability of the molecule. It has been observed that mutations in the complementary segment can rescind the deleterious effect by mutating into a base pair that matches the already mutated base, thus recovering the fitness of the original molecule (Kelley *et al.*, 2000; Wilke *et al.*, 2003). Other direct empirical observations of compensatory mutations include the fitness recovery in RNA virus (Burch and Chao, 1999) and the compensatory effect of molecular chaperones (Fares *et al.*, 2002).

Theoretical studies on compensatory mutations are few and far between. Poon and Otto (2000) recently adopted Fisher’s geometrical model of adaptation to study the population fitness effect of the fixation of deleterious mutations in small populations with the consideration of compensatory mutations. They found that the fixation of deleterious mutations causing extinction of a small population becomes increasingly unlikely when the scope of compensatory mutations increases, contradictory to what previous studies claim about the high likelihood of extinction of small populations due to “mutational meltdown” (Lynch *et al.*, 1993). Similarly, Whitlock (2000) considered the effect of beneficial mutations on fitness recovery of small populations. However, Whitlock’s model does not incorporate the effect of compensatory mutations explicitly; instead, he considered two classes of mutations,

deleterious and beneficial. There should be an important distinction between compensatory mutations and beneficial mutations. Beneficial mutations are mutations that have positive fitness effect on the individuals that carry them, whereas compensatory mutations should be deleterious by themselves, and only become beneficial or reduce the negative fitness effect of certain deleterious mutations when they are present in the same individuals. This can be easily imagined in the population—some individuals carry deleterious mutation A, some carry deleterious mutation B, some carry beneficial mutation C, some carry A and B, which lead to a reduced negative fitness effect, most of the rest of the individuals are mutation free. Therefore, it is evident that compensatory mutations could be the key to understanding the maintenance of small populations in nature without signs of going to extinction, and much research needs to be done in understanding the scope of compensatory mutations.

In this study, we examine the effect of population size on the compensatory mutations, investigating whether compensatory mutations have different rescue effects on populations of different sizes. Given the observation that some natural populations can be very small, if compensatory mutations have a relatively bigger effect on small populations' fitness, then we should not ignore them, especially when considering small populations. We compare the effect of different effective population sizes on the net reduction of total fitness due to the fixation of deleterious mutations alone vs. that due to the combination of both deleterious and compensatory mutations. We examine different combinations of parameters that describe the distributional shapes of the fitness effects of deleterious mutations and compensatory mutations, to see how the reduction of total fitness changes as a result.

# MATHEMATICAL DERIVATIONS

Consider a randomly mating diploid population and assume that the population is at an equilibrium between the number of new alleles due to mutation and the loss of alleles due to fixation. Let  $N$  be the population size,  $N_e$  the effective population size, and  $s$  the selective coefficient of a mutation. Crow and Kimura (1970) derive the probability  $u(s)$  of fixation of a new allele as

$$u(s) = \frac{1 - e^{-2s(N_e/N)}}{1 - e^{-4N_e s}}. \quad (1)$$

The cumulative effect of deleterious mutations or the net fitness decay should depend on the number  $U_D$  of deleterious mutations per generation per individual, the total number  $N$  of individuals in the population, the fitness effect  $s$  of mutations, and the fixation probability  $u(s)$  of mutations. In reality the distribution density  $f_D(s)$  of the fitness effect of deleterious mutations is unknown a priori, and since the fixation probability  $u(s)$  of the deleterious mutation is a function of the fitness effect  $s$ , some sort of continuous distribution density  $f_D(s)$  must be assumed. Assuming that the deleterious effect is additive, one can integrate over the distribution density  $f_D(s)$  to obtain the net fitness change  $\Delta W_D$  due to deleterious mutations on the population. Precisely,

$$\Delta W_D = U_D N \int_{-\infty}^0 2su(s)f_D(s) ds, \quad (2)$$

where the factor 2 is explained below, and the integral is only over negative fitness effects. Assume that all deleterious mutations are entirely recessive, so both wild type homozygotes and heterozygotes have a fitness of 1, whereas the mutant homozygotes have a fitness of  $1 - 2|s|$ . Therefore, the (negative) fitness effect of such a deleterious mutation is  $2s$ .

Assume that the distribution density  $f_D(s)$  of  $s$ ,  $-\infty < s \leq 0$ , can be described by a Gamma distribution on  $|s|$  with mean  $\lambda_D < 0$  and coefficient of variation  $C_D$ . After a

series of mathematical derivations (see Zhang and Watson (2007) for details), we obtain the following expression:

$$\Delta W_D = \frac{2U_D N \Delta \mathcal{Z}_D}{|\lambda_D|^{1/C_D^2} (4N_e C_D^2)^{1+1/C_D^2}}, \quad (3)$$

where

$$\Delta \mathcal{Z}_D = \mathcal{Z} \left( 1 + \frac{1}{C_D^2}, 1 + \frac{1}{4N_e |\lambda_D| C_D^2} \right) - \mathcal{Z} \left( 1 + \frac{1}{C_D^2}, 1 + \frac{1}{4N_e |\lambda_D| C_D^2} - \frac{1}{2N} \right) \quad (4)$$

and

$$\mathcal{Z}(x, a) = \sum_{i=0}^{\infty} \frac{1}{(i+a)^x}$$

is the generalized Riemann zeta function.

In order to examine quantitatively the contribution of different intervals of selective coefficients to the population fitness, it is necessary to get a general expression of the above equation for any interval of selective coefficients within  $(-\infty, 0]$ . Specifically, for  $0 < a < b$ , the population fitness reduction due to deleterious mutations with a selective coefficient  $s$  such that  $a \leq |s| \leq b$  can be expressed as

$$\Delta W_D(a, b) = \frac{U_D \Delta \mathcal{Z}_D}{2N_r (|\lambda_D| C_D^2 4N_e)^{1/C_D^2}} \Gamma^*(a, b), \quad (5)$$

where  $N_r = N_e/N$ ,

$$\Gamma^*(a, b) = \frac{\gamma(1 + \frac{1}{C_D^2}, a) - \gamma(1 + \frac{1}{C_D^2}, b)}{\Gamma(\frac{1}{C_D^2})} \quad (6)$$

and

$$\gamma(\alpha, x) = \int_x^{\infty} t^{\alpha-1} e^{-t} dt \quad (7)$$

is the incomplete Gamma function, which becomes the Euler Gamma function  $\Gamma(\alpha)$  when

$x = 0$ .  $\Delta W_D(a, b)$  was derived by following exactly the derivation of  $\Delta W_D$  in Zhang and Watson (2007) and replacing the integrals  $\int_0^\infty$  by integrals  $\int_a^b$ . Since beneficial mutations are not being considered here, we will drop the subscript  $D$ , e.g., writing  $\Delta W_D$  as  $\Delta W$ .

## RESULTS AND DISCUSSION

Before directly addressing compensatory mutations, we discuss some implications of the Gamma distribution assumption for the fitness effects of beneficial, deleterious, and compensatory mutations. Gamma distributions are defined by two parameters,  $\alpha$  and  $\theta$ , related to  $\lambda_D$ ,  $C_D$  used in (3)–(6) by  $\alpha = 1/C_D^2$ ,  $\theta = |\lambda_D|C_D^2$ ,  $|\lambda_D| = \alpha\theta$ ,  $C_D = 1/\sqrt{\alpha}$ . Figure 1 shows five Gamma distributions of selective coefficients with different sets of parameters, illustrating that the distribution of selective coefficients can assume different shapes and scales, which correspond biologically to different fitness effects of deleterious (or beneficial) mutations, and the spread of these effects. For example, compared to the distribution with  $\alpha = 5$  and  $\theta = 1$ , the distribution with  $\alpha = 1$  and  $\theta = 1$  has a higher proportion of mutations with small fitness effects, and lower proportion of mutations with large fitness effects.

Figure 2 shows the effect of the squared coefficient of variation of selective coefficients ( $C_D^2 = 1/\alpha$ ) on the total fitness change of a population. Several observations can be made. First, there exists a squared coefficient of variation of selective coefficients  $\tilde{C}_D^2$  that maximizes the total population fitness reduction  $|\Delta W|$ , and its value depends on the distribution of fitness effects of the deleterious mutations. For certain distributions of fitness effects of deleterious mutations, one can find the exact value of  $\tilde{C}_D^2$  by setting the derivative of  $\Delta W$  with respect to  $C_D^2$  equal to zero. For example, when  $N = 1000$ ,  $U_D = 100$ ,  $N_e = 100$ , and  $\lambda_D = -0.02$ ,  $\tilde{C}_D^2 = 1.46$ ; when  $N = 1000$ ,  $U_D = 100$ ,  $N_e = 100$ , and  $\lambda_D = -0.2$ ,  $\tilde{C}_D^2 = 4.51$ . Second, in general, regardless of the distributional shapes of



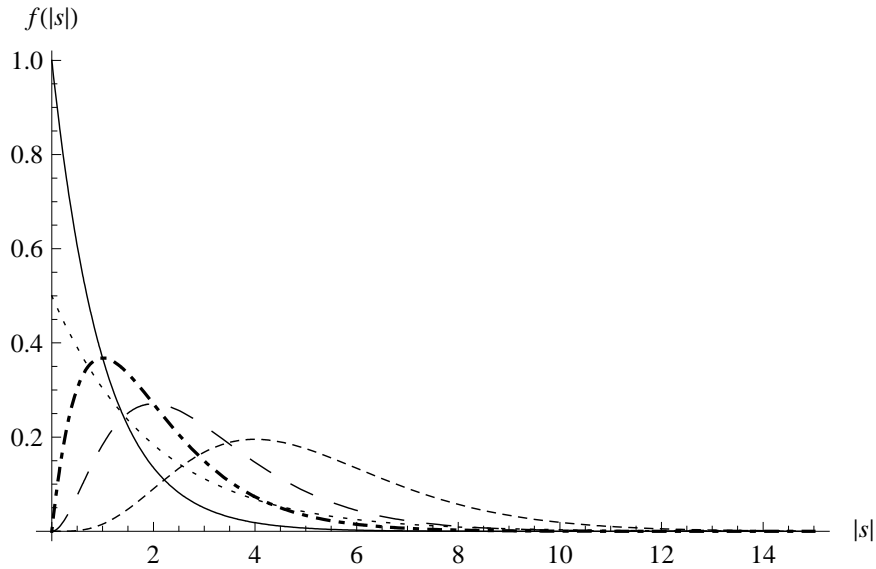


Figure 1: The Gamma distribution of selective coefficients.  $\alpha = 1, \theta = 1$  (solid);  $\alpha = 2, \theta = 1$  (dot-dashed);  $\alpha = 3, \theta = 1$  (long-dashed);  $\alpha = 5, \theta = 1$  (short-dashed);  $\alpha = 1, \theta = 2$  (dotted).

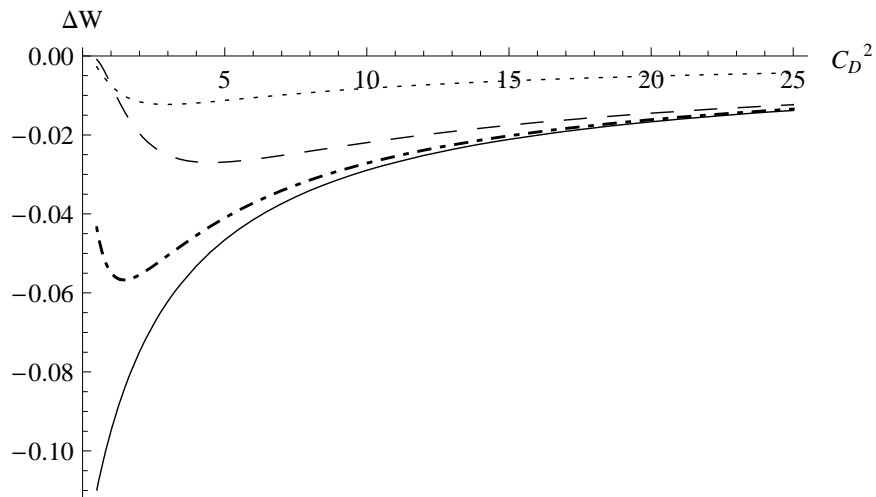


Figure 2: The effect of the squared coefficient of variation of selective coefficients ( $C_D^2$ ) on the total fitness change of a population ( $\Delta W$ ). All curves have  $N = 1000$  and  $U_D = 100$ .  $N_e = 0.1 N$  and  $\lambda_D = -0.002$  (solid);  $N_e = 0.1 N$  and  $\lambda_D = -0.02$  (dot-dashed);  $N_e = 0.1 N$  and  $\lambda_D = -0.2$  (long-dashed);  $N_e = 0.3 N$  and  $\lambda_D = -0.02$  (dotted).

the fitness effects of deleterious mutations, the higher the squared coefficient of variation of selective coefficients, the less fitness reduction  $|\Delta W|$  the overall deleterious mutations cause the population. This suggests that even for small populations, as long as there is a large variation in the fitness effects of deleterious mutations, the fixation of deleterious mutations will have a small negative effect on the population. Third, for the same distributional shape of fitness effects, the reduction  $|\Delta W|$  of overall population fitness decreases greatly as the effective population size increases (e.g., dot-dashed vs. dotted curve in Figure 2).

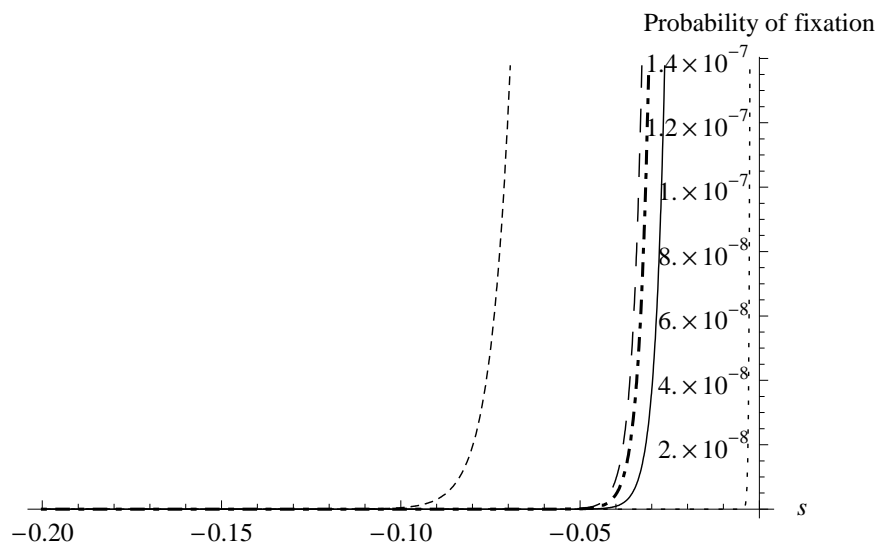


Figure 3: The fixation probability decreases rapidly with the increasing negative effect  $s$  of a deleterious mutation.  $N_e = 100$ ,  $N_e/N = 0.1$  (solid);  $N_e = 100$ ,  $N_e/N = 0.5$  (dot-dashed);  $N_e = 100$ ,  $N_e/N = 1$  (long-dashed);  $N_e = 1000$ ,  $N_e/N = 1$  (dotted);  $N_e = 50$ ,  $N_e/N = 1$  (short-dashed).

Fourth, when holding all other parameters constant, the total fitness of a population decreases more rapidly as  $C_D$  decreases when the mean fitness effect of the deleterious mutations ( $\lambda_D$ ) is small (in magnitude). The same holds true for fixed  $C_D$ . This is counter-intuitive. The observation that the higher (in magnitude) the mean negative fitness effects of the distribution, the less total fitness reduction the deleterious mutations create, suggests that it is probably due to the fact that deleterious mutations with large (in magnitude)

negative effects have lower probability of fixation in the population, therefore, they do not contribute much to the final fitness reduction. To address this, we examine the fixation probability  $u(s)$  as a function of the selective coefficient of mutation  $s$ . Figure 3 supports the above explanation for the counter intuitive observation deduced from Figure 2. Clearly, the fixation probability  $u(s)$  of a deleterious mutation is very sensitive to the effective population size ( $N_e$ ), but not sensitive to the real population size ( $N$ ). Moreover, the probability of fixation of a deleterious mutation decreases rapidly as its negative fitness effect increases in magnitude. Therefore, it seems that deleterious mutations with large (in magnitude) negative effects, because the probability of their fixation is very low, do not eventually contribute much to the reduction of the total fitness of the population, which seems to be more the result of deleterious mutations with small effects. In classical models, these small effect mutations are considered to be effectively neutral when  $N_e|s| \ll 1$ .

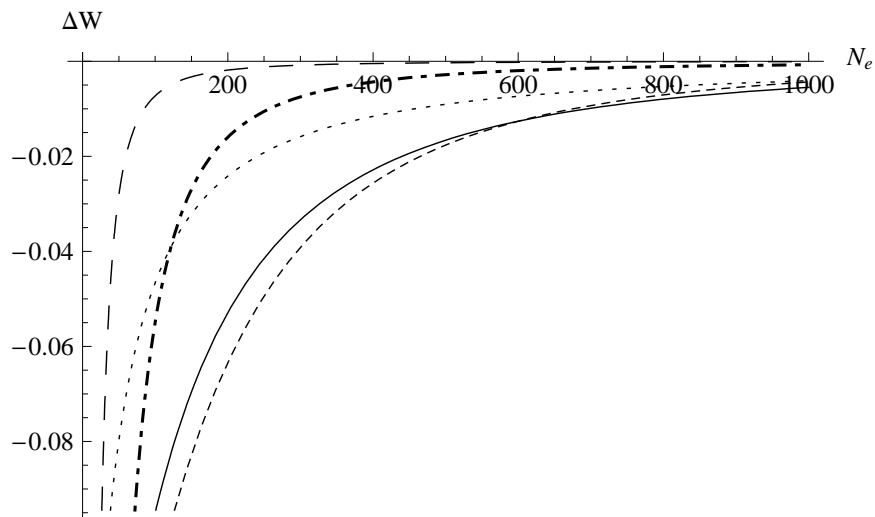


Figure 4: The total fitness reduction of a population as a function of the effective population size. All curves assume  $N = 1000$  and  $U_D = 100$ .  $C_D^2 = 1$  for  $\lambda_D = -0.002$  (solid),  $\lambda_D = -0.02$  (dot-dashed), and  $\lambda_D = -0.2$  (long-dashed). The remaining two curves correspond to  $C_D^2 = 5$  and  $\lambda_D = -0.002$  (dotted),  $C_D^2 = 0.5$  and  $\lambda_D = -0.002$  (short-dashed).

Figure 4 shows the total fitness reduction as a result of fixation of deleterious mutations for different effective population sizes ( $N_e$ ). Figure 4 indicates several things. First, the

reduction of the total fitness of a population decreases (in magnitude) when its effective population size increases and the reduction levels off at a certain  $N_e$  (hereafter called the *level-off*  $N_e$ ), which depends on the mean negative fitness effect of deleterious mutations. The smaller (in magnitude) the mean negative fitness effect, the larger the  $N_e$  required for the level off of the reduction in total fitness. Since the total fitness of a population decreases rapidly when the effective population size gets smaller than the level-off  $N_e$ , and it has been shown that the effective population size of a natural population is on average around 10% of the census population size (Frankham, 1995), in this case  $N_e \approx 100$ , we can see that for many distributions, especially the ones with small (in magnitude) mean fitness effects (i.e.,  $\lambda_D \approx 0$ ), the total population fitness reduction can be very high.

We now turn our attention to the main topic of the paper. The distribution of fitness effects of spontaneous mutations is unknown, so to study the effect of compensatory mutations (under the Gamma distribution assumption for mutations), we choose a number of different combinations of parameters for illustration. Specifically, we compare the total fitness reduction in a population taking into account compensatory mutations ( $\Delta W_c$ ) to that without compensatory mutations ( $\Delta W_{nc}$ ):  $\Delta W_{\text{diff}} = \Delta W_c - \Delta W_{nc}$ .

First, we compare different Gamma distributions that can be used to represent distributions with/without the influence of compensatory mutations. As compensatory mutations should make deleterious mutations less deleterious, for a specific Gamma distribution with certain  $\alpha$  and  $\theta$  (see Figure 1) corresponding to noncompensatory deleterious mutations, compensatory mutations will move the Gamma distribution towards the origin (i.e., the distribution with compensatory mutations should have either smaller  $\alpha$  than the original, or smaller  $\theta$  than the original, or both. In terms of  $\lambda_D$  and  $C_D^2$ , the distribution with compensatory mutations should tend to have smaller (in magnitude) mean negative fitness effects and possibly, larger squared coefficient of variation of fitness effects. Based on this idea, we choose several distributional shapes to examine the compensatory effect. The

distributions are shown in Figure 5. All distributions have the same squared coefficient of variation ( $C_D^2 = 5$ ), but different mean fitness effects ( $\lambda_D = -0.0001$  for the distribution of compensatory mutations, vs.  $-0.0002, -0.001, -0.01, -0.1, -0.2$  for noncompensatory mutations). Therefore, compared with the noncompensatory mutation's fitness distributional shapes, the compensatory mutation's distributional shape of selective coefficient  $f(|s|)$  has a larger proportion of deleterious mutations that have small (in magnitude) negative fitness effects.

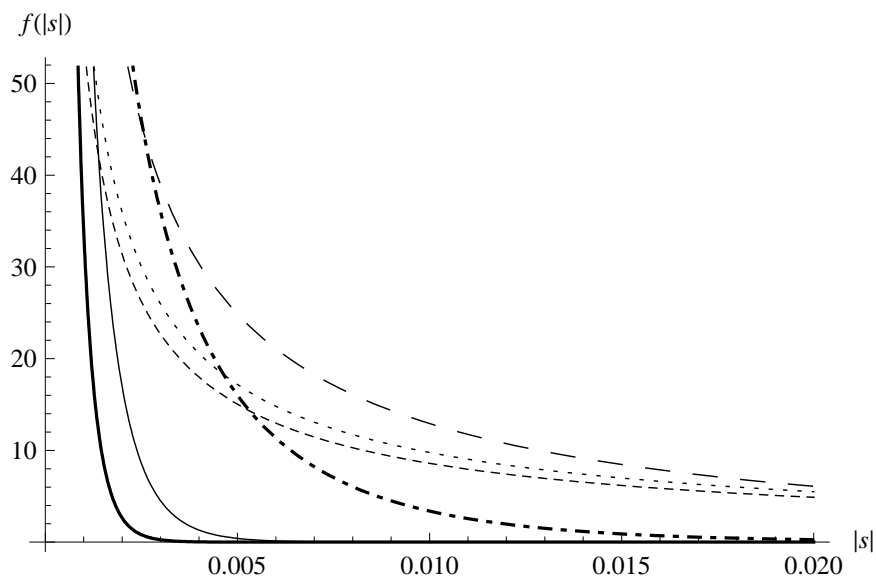


Figure 5: Thick solid curve represents the distributional shape of the fitness effects of deleterious mutations with compensatory mutations, and all other curves represent only deleterious mutations. All curves have the same  $C_D^2 = 5$ , and varying  $\lambda_D$ .  $\lambda_D = -0.0001$  (thick solid),  $\lambda_D = -0.0002$  (solid),  $\lambda_D = -0.001$  (dot-dashed),  $\lambda_D = -0.01$  (long-dashed),  $\lambda_D = -0.1$  (dotted),  $\lambda_D = -0.2$  (short-dashed).

Figure 6 shows the compensatory effects as a function of effective population size. The compensatory effect is very significant when the effective population size is small, and the compensatory effect decreases sharply over a small range of  $N_e$ . The smaller the effective population size, the larger the compensatory effect is on the total population fitness. If we assume that the effective population size is about 10% of the real population size, in all

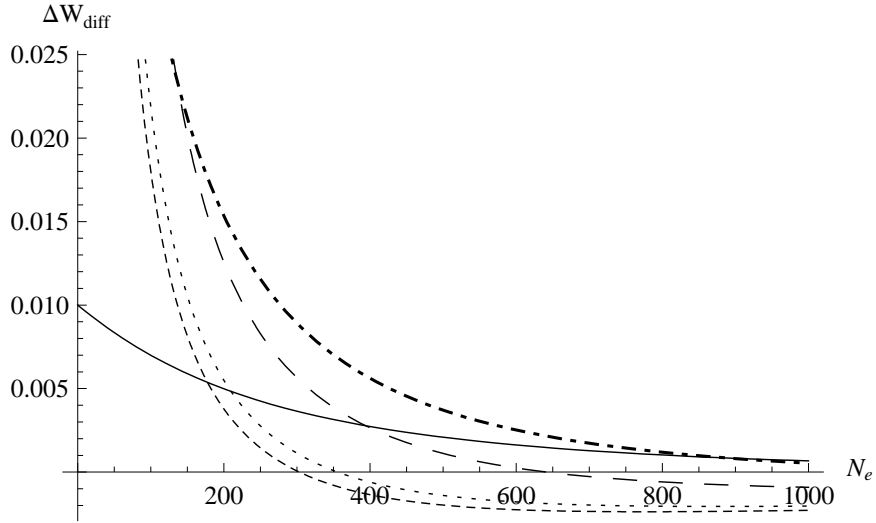


Figure 6: The compensatory effect  $\Delta W_{\text{diff}}$  for  $\Delta W_c$  with  $\lambda_D = -0.0001$ ,  $C_D^2 = 5$ ,  $N = 1000$ ,  $U_D = 100$ , and  $\Delta W_{nc}$  with the same  $C_D^2$ ,  $N$ ,  $U_D$ , and varying  $\lambda_D$ .  $\lambda_D = -0.0002$  (solid),  $\lambda_D = -0.001$  (dot-dashed),  $\lambda_D = -0.01$  (long-dashed),  $\lambda_D = -0.1$  (dotted),  $\lambda_D = -0.2$  (short-dashed).

the cases considered in Figure 6, the compensatory effect is very large, i.e., considering the effect of compensatory mutations on the total population fitness is quite different from not doing so. A second observation is that the compensatory effect essentially “loses” when the effective population size reaches a certain point. In fact, the effect reverses: the total fitness reduction with compensatory mutations is greater than that without compensatory mutations (i.e.,  $\Delta W_{\text{diff}} < 0$ ). This is a bit puzzling. One plausible explanation is that as the effective population size increases, for the distribution without compensatory mutations with large mean deleterious effects, most of the mutations do not get fixed in the population, only a small fraction of them get fixed. However, this altogether does not contribute as much negative total fitness as the distribution with compensatory mutations; for the latter, a possibly larger fraction of deleterious mutations, each with less negative fitness effect, get fixed in the population, and because their fraction is large, the fixation of these mutations actually results in more fitness reduction. Therefore, the efficiency of compensatory effects

is higher in small populations.

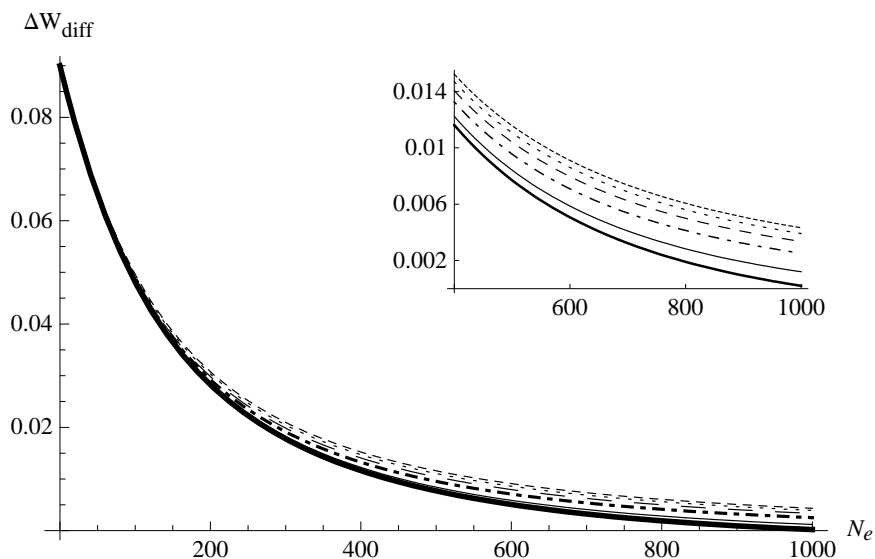


Figure 7: The compensatory effect  $\Delta W_{\text{diff}}$  with varying  $C_D^2$ . All curves are comparisons between  $\lambda_D = -0.0001$  for compensatory distribution of  $s$  vs.  $\lambda_D = -0.001$  for noncompensatory distribution of  $s$  with the same  $U_D = 100$  and  $N = 1000$ . The noncompensatory distribution of  $s$  has  $C_D^2 = 2$ , the compensatory distribution of  $s$  has  $C_D^2 = 1$  (thick solid),  $C_D^2 = 2$  (solid),  $C_D^2 = 4$  (dot-dashed),  $C_D^2 = 6$  (long-dashed),  $C_D^2 = 8$  (dotted),  $C_D^2 = 10$  (short-dashed).

Although it is clear that compensatory mutations should reduce (in magnitude) the mean negative fitness effect of deleterious mutations, it is not so clear how the coefficient of variation will change as a result. This is because the coefficient of variation is the ratio of the standard deviation ( $\sigma_D$ ) to the absolute mean fitness effect ( $|\lambda_D|$ ), so when the absolute mean fitness effect is reduced due to compensatory mutations, the standard deviation may also be reduced, but the ratio of the two can be either decreased or increased. Therefore, we take into account this uncertainty and examine the compensatory effect on the population fitness reduction with varying squared coefficients of variation due to compensatory mutations, i.e., we assume the same distributional shape for the fitness effect of deleterious mutations without compensatory mutations ( $\lambda_D = -0.001$  and  $C_D^2 = 2$ ) but several distributional shapes for the fitness effect of deleterious mutations with

compensatory mutations with the same (reduced in magnitude) mean negative fitness effect ( $\lambda_D = -0.0001$ ) but varying squared coefficient of variation, either the same ( $C_D^2 = 2$ ), lower ( $C_D^2 = 1$ ), or higher ( $C_D^2 = 4, 6, 8, 10$ ) than the original. Figure 7 shows the result. Interestingly, regardless of whether the squared coefficient of variation due to compensatory mutations is lower or higher than that for the distribution without compensatory mutations, the compensatory effect still exists. Moreover, although the larger the squared coefficient of variation, the larger the compensatory effect on the total population fitness reduction, the difference between compensatory effects for different squared coefficients of variation reduces as the effective population size decreases. For small effective population sizes, the difference between distributions with different squared coefficients of variation seems minor. Previous study shows that large coefficient of variation in the fitness effect of mutations, be it deleterious or beneficial, can help prevent or slow down the extinction of small populations (Zhang and Watson, 2007). The result shown here suggests that the compensatory effect of compensatory mutations in small populations is quite robust: as long as the mean negative fitness effect is reduced in magnitude, the compensatory effect persists regardless of the direction of change in the coefficient of variation.

To further examine the relationship between the effective population size and compensatory effect, we change the rate of deleterious mutations  $U_D$ . Figure 8 shows that when the distributions of fitness effects with and without compensatory mutations are both set, for the same effective population size, the compensatory effect increases as the rate of deleterious mutations increases, and the compensatory effect is more dramatic for smaller effective population sizes than larger ones with the change in the rate of deleterious mutations.

It is clear from the comparison of different distributional shapes with and without compensatory mutations that there is great uncertainty as to what distributional shapes of the fitness effects spontaneous mutations have, and how spontaneous compensatory mutations change the shapes. Alternatively, we ask “what classes of deleterious mutations



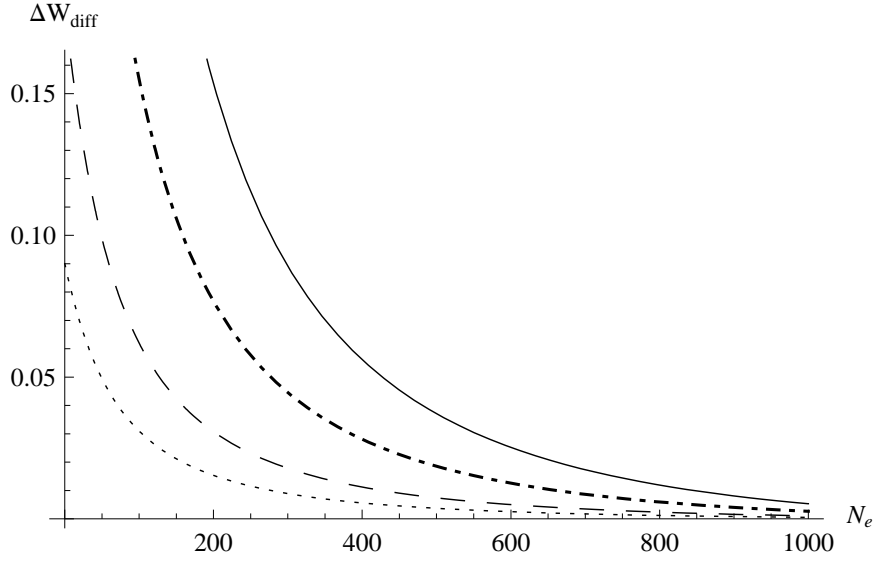


Figure 8: The compensatory effect  $\Delta W_{\text{diff}}$  for  $\Delta W_c$  with  $\lambda_D = -0.0001$ ,  $C_D^2 = 5$ , and  $N = 1000$ , and  $\Delta W_{nc}$  with the same  $C_D^2$ ,  $N$ , and  $\lambda_D = -0.001$ .  $U_D = 1000$  (solid),  $U_D = 500$  (dot-dashed),  $U_D = 200$  (long-dashed),  $U_D = 100$  (dotted).

do compensatory mutations compensate for?” If we classify deleterious mutations’ selective coefficients into different categories based on the degree of deleteriousness, then the question is “which categories of mutation do compensatory mutations have an impact on?” To examine this issue quantitatively, we use the intervals of selective coefficients for deleterious mutations that have been used in recent studies (Eyre-Walker *et al.*, 2006; Yampolsky *et al.*, 2005), which divided the spectrum of selective coefficients  $s$  into five intervals:  $(-\infty, -L]$ ,  $(-L, -10^{-2}]$ ,  $(-10^{-2}, -10^{-4}]$ ,  $(-10^{-4}, -10^{-5}]$ , and  $(-10^{-5}, 0]$ , where anything below  $-L$  is considered “lethal.” For each interval, we use Equation (5) to compute the population fitness reduction due to the fixation of the deleterious mutations with fitness effects in the interval. Here we quantitatively evaluate the compensatory effect of the compensatory mutations with different selection coefficient distributions shown in Figure 6, all of which have  $C_D^2 = 5$ . Note from Equation 5 that for each interval  $(-b, a]$  and fixed  $C_D^2$ , the compensatory effect  $\Delta W_{\text{diff}} = \Delta W_c - \Delta W_{nc}$  is the product of a factor independent of

$(-b, a]$  with  $\Gamma^*(a, b)$ , which also depends on  $C_D^2$ . Hence the compensatory effect  $\Delta W_{\text{diff}}$  is proportional to  $\Gamma^*(a, b)$ .

Taking  $L = 1$  and  $C_D^2 = 5$ ,  $\Gamma^*(a, b)$  for the five intervals is, respectively, 0.0906094, 0.108672, 0.000715842,  $2.69523 \times 10^{-6}$ , and  $1.8152 \times 10^{-7}$ , which values are proportional to the fitness loss  $\Delta W(a, b)$  caused by deleterious mutation effects in these intervals given a Gamma fitness effect distribution with  $C_D^2 = 5$ . It can be seen that the fitness loss is mainly due to the mutations with highly deleterious effects, and this suggests that compensatory mutations, if having a large effect, should “rescue” the effect of deleterious mutations with especially large effects. With the same  $C_D^2 = 5$ ,  $N = 1000$ , and  $U_D = 100$ , we compare the compensatory effect in the five intervals for three effective population sizes (100, 200, and 400). Assuming that compensatory mutations change the mean fitness effect from a noncompensatory value of  $\lambda_D$  to the compensatory value  $\lambda_D = -0.0001$ , then  $\Delta W_{\text{diff}}$  is given in the following table:

	noncompensatory $\lambda_D = -0.001$		
	$N_e = 100$	$N_e = 200$	$N_e = 400$
$(-\infty, -L]$	0.0140434	0.00695439	0.00254328
$(-L, -10^{-2}]$	0.0168428	0.00834071	0.00305027
$(-10^{-2}, -10^{-4}]$	0.000110947	0.0000549418	0.0000200927
$(-10^{-4}, -10^{-5}]$	$4.17729 \times 10^{-7}$	$2.06863 \times 10^{-7}$	$7.56515 \times 10^{-8}$
$(-10^{-5}, 0]$	$2.81334 \times 10^{-8}$	$1.39319 \times 10^{-8}$	$5.09501 \times 10^{-9}$
	noncompensatory $\lambda_D = -0.1$		
	$N_e = 100$	$N_e = 200$	$N_e = 400$
$(-\infty, -L]$	0.00990151	0.00249074	-0.00027254
$(-L, -10^{-2}]$	0.0118753	0.00298726	-0.00032687
$(-10^{-2}, -10^{-4}]$	0.000078225	0.0000196776	$-2.15315 \times 10^{-6}$

$(-10^{-4}, -10^{-5}]$	$2.94527 \times 10^{-7}$	$7.40887 \times 10^{-8}$	$-8.10688 \times 10^{-9}$
$(-10^{-5}, 0]$	$1.98359 \times 10^{-8}$	$4.98976 \times 10^{-9}$	$-5.45986 \times 10^{-10}$

This data shows that the compensatory effect of compensatory mutations decreases as the effective population size increases, and the efficiency of the compensatory effect is higher in small populations than large ones.

The general picture that emerges from the examination of different combinations of parameters is that compensatory mutations are an important class of mutations that cannot be ignored, especially in small populations. Together with previous studies showing that large variation in fitness effects of deleterious mutations can be an important factor in how some small natural populations sustain themselves, the present results clearly show that compensatory mutations can be another key mechanism for the survival of small populations. Given that compensatory mutations do occur in nature, sometimes quite frequently, our results suggest the hypothesis that compensatory mutations are more prevalent in small populations than large populations, and future empirical studies should be designed to test this prediction.

## ACKNOWLEDGMENTS

This work was supported in part by NIGMS/NIH Grant 1 R01 GM078989-01, NSF Grants DMI-0355391 and IIS-0710945, Department of Energy Grant DE-FG02-06ER25720, and a VPI & SU ASPIRES grant.

## References

Burch, C. L. and Chao, L., 1999. Evolution by small steps and rugged landscapes in the RNA virus phi6. *Genetics*. 151, 921–927.

- Crow, J. F. and Kimura, M., 1970. *An Introduction to Population Genetics Theory*. Harper & Row, New York, NY .
- Davies, E. K., Peters, A. D., and Keightley, P. D., 1999. High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science*. 285, 1748–1751.
- Estes, S., Phillips, P. C., Denver, D. R., Thomas, W. K., and Lynch, M., 2004. Mutation accumulation in populations of varying size: the distribution of mutational effects for fitness correlates in *Caenorhabditis elegans*. *Genetics*. 166, 1269–1279.
- Eyre-Walker, A., Woolfit, M., and Phelps, T., 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics*. 173, 891–900.
- Fares, M. A., Ruiz-Gonzalez, M. X., Moya, A., Elena, S. F., and Barrio, E., 2002. Endosymbiotic bacteria: groEL buffers against deleterious mutations. *Nature*. 417, 398.
- Frankham, R., 1995. Effective population size/adult population size ratios in wildlife: a review. *Genet. Res. Camb.*. 66, 95–106.
- Keightley, P. D., 1994. The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics*. 138, 1315–1322.
- Kelley, S. O., Steinberg, S. V., and Schimmel, P., 2000. Functional defects of pathogenic human mitochondrial tRNAs related to structural fragility. *Nat. Struct. Biol.*. 7, 862–865.
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature*. 217, 624–626.
- Lande, R., 1998. Risk of population extinction from fixation of deleterious and reverse mutations. *Genetica*. 102-103, 21–27.
- Lynch, M., Burger, R., Butcher, D., and Gabriel, W., 1993. The mutational meltdown in asexual populations. *J. Hered.*. 84, 339–344.

- Nielsen, R. and Yang, Z., 2003. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol. Biol. Evol.* 20, 1231–1239.
- Piganeau, G. and Eyre-Walker, A., 2003. Estimating the distribution of fitness effects from DNA sequence data: implications for the molecular clock. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10335–10340.
- Poon, A. and Otto, S. P., 2000. Compensating for our load of mutations: freezing the meltdown of small populations. *Evolution Int. J. Org. Evolution.* 54, 1467–1479.
- Sanjuan, R., Moya, A., and Elena, S. F., 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8396–8401.
- Vassilieva, L. L., Hook, A. M., and Lynch, M., 2000. The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution Int. J. Org. Evolution.* 54, 1234–1246.
- Whitlock, M. C., 2000. Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. *Evolution Int. J. Org. Evolution.* 54, 1855–1861.
- Wilke, C. O., Lenski, R. E., and Adami, C., 2003. Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. *BMC Evol. Biol.* 3, 3.
- Yampolsky, L. Y., Kondrashov, F. A., and Kondrashov, A. S., 2005. Distribution of the strength of selection against amino acid replacements in human proteins. *Hum. Mol. Genet.* 14, 3191–3201.
- Zhang, L. and Watson, L. T., 2007. Note on the computation of critical effective population sizes. *J. Comput. Biol.* 14, 950–960.