Getting There First: An Evolutionary Rate Advantage for Adaptive Loss-of-Function Mutations

Michael J. Behe[1]

[1]*Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015, USA.*
mjb1@lehigh.edu

Abstract

Over the course of evolution organisms have adapted to their environments by mutating to gain new functions or to lose pre-existing ones. Because adaptation can occur by either of these modes, it is of basic interest to assess under what, if any, evolutionary circumstances one of them may predominate. Since mutation occurs at the molecular level, one must look there to discern if an adaptation involves gain- or loss-of-function. Here I present a simple, deterministic model for the occurrence and spread of adaptive gain-of-function versus loss-of-function mutations, and compare the results to laboratory evolution experiments and studies of evolution in nature. The results demonstrate that loss-of-function mutations generally have an intrinsic evolutionary rate advantage over gain-of-function mutations, but that the advantage depends radically on population size, ratio of selection coefficients of competing adaptive mutations, and ratio of the mutation rates to the adaptive states.

Key words: gain-of-function mutation, loss-of-function mutation, rate of fixation

1. Introduction

In *On the Origin of Species* Charles Darwin emphasized that natural selection is relentless, continuously monitoring each organism for its fitness, selecting those with an advantage and weeding out the disadvantaged [1]. However, as Darwin also knew, an organism's advantage in a particular set of circumstances did not have to involve the gain of a new ability, such as the power to fly or swim. Indeed, it could involve the loss of those abilities. Flightless birds had adapted to their habitats partially by abandoning such a faculty. Some organisms went even further. Darwin described some barnacles in which the male was reduced to a transparent sac, with little but a reproductive system remaining [2]. By specializing in this way, the barnacles and their descendants presumably gained an adaptive advantage over competitors.

In the nineteenth century Darwin and his contemporaries could identify mutations only through their phenotypic effects. However, with the progress of biology especially in the last half-century, contemporary science can now characterize

mutations also by their molecular effects to the genetic material of a species. In order to understand the roles of loss-of-function (LOF) versus gain-of-function (GOF) mutations, one must keep phenotypic versus molecular changes separate. An altered, visually observable phenotype may be due to any of a number of disparate underlying molecular mutations. For example, a mutant mouse that is 50% larger than its litter mates might have had the gene for a repressor protein that switches off production of growth hormone deleted. At the molecular level, that would be an LOF mutation, since a functional molecular feature was deleted, even though the increased size of the mouse may strike the casual observer as a gain-of-function. On the other hand, a large mutant mouse might be due to the formation of a new promoter site for a transcription factor near a gene involved in growth, which would be a GOF mutation, since a new functional molecular feature (the promoter site) was produced. In this paper I will consider LOF and GOF mutations as affecting functional molecular features such as genes and regulatory elements, no matter what their possible phenotypic effects may be.

2. The model

Consider a population of organisms that comes under a new selective pressure. To respond to the pressure two different adaptive mutations are postulated to be potentially available: one which results in the gain of a molecular function, and another which results in the loss of one. What factors might affect the probabilities of either kind of mutation becoming fixed in the population in competition with the other? One factor of immediate importance is the rate of appearance of the adaptive mutations. It is very often possible to eliminate a molecular function by a variety of mutations. GOF mutations, on the other hand, are generally much more specific, sometimes being produced in only one way.

As an illustration, consider several mutations to human genes that give a measure of resistance to malaria. The best known such mutation is the sickle cell gene in which, by means of a single A→T transversion, the codon for a glutamic acid residue in the sixth position of the β-chain globin gene is converted to a codon for valine [3]. This can be considered a GOF mutation, because the hemoglobin gains a self-association site on its surface, allowing the individual proteins upon deoxygenation to aggregate into microtubular-like structures [4]. By an as-yet-unknown mechanism, the polymerization negatively affects the growth of the malarial parasite (which spends part of its life cycle in the red blood cell) [5, 6]. Another mutation which confers a measure of resistance to malaria is deficiency of glucose-6-phosphate dehydrogenase (G6PD), in which a mutant gene produces little or no functional enzyme [7]. For reasons that are unclear, this interferes with

parasite viability. Population genetic studies have shown that hundreds of separate mutations have led to deficiency of wild-type G6PD in populations at risk for malaria. On the other hand, the mutation producing the sickle gene is thought to have arisen de novo only a few times in the last 10,000 years, or perhaps only once [8].

The reason for the disparity in the number of *de novo* mutations is straightforward. To secure a sickle mutation a particular nucleotide of the β-globin gene must be substituted. Since the nucleotide mutation rate of humans is on the order of 10^{-8} substitutions/ generation, that is also the de novo rate of appearance of the sickle gene [9]. On the other hand, there are many ways to produce a nonfunctional protein such as malaria-resistant G6PD. For example, during replication the insertion of a nucleotide anywhere within the coding sequence results in a frame-shift and likely an inactive polypeptide. Deletion of a nucleotide in the coding region will have the same affect, as will alteration of a codon from sense to nonsense. Longer insertions and deletions will frequently have the same effect. Missense mutations, although likely not completely inactivating the protein, will often make the protein less stable or less functional. Thus, considered as a class, the mutation rate from a functional to a nonfunctional gene may be several orders of magnitude greater than the basic nucleotide mutation rate. (Indeed, the adaptation rate of *E. coli*, whose generational nucleotide mutation rate is 50-fold lower than that of humans, has recently been measured as 10^{-5})[10]. For the two classes of mutations, in this paper I explore the effect of this factor on the evolutionary rate of spread of adaptive mutations as a function of population size, mutation rate, selection coefficient, ratio of selection coefficients of the competing adaptive mutations, and ratio of mutation rates to the adaptive state.

Calculations were performed using *Mathematica* [34].

3. Results

3.1 *Relatively small population sizes*

In this section I consider small population sizes ($N_e << 1/v$), where N_e is the effective population size and v is the mutation rate per generation. Unless otherwise stated, organisms are assumed to be haploid (because most laboratory evolution experiments have been done with haploids), and the model is developed accordingly. The resulting equations can be applied to diploid organisms by replacing N_e by $2N_e$.

In order for an adaptive mutation to become fixed in a population of relatively small size two separate processes must occur, each with its own time scale: (1) if

the mutation does not yet exist in the population when the selective pressure begins, then the expected waiting time to the appearance of the selected mutation is $t_{w1} = 1/(2N_e vs)$, where s is the selection coefficient; (2) once the selected mutation appears, the time for it to fix in the population is $t_{fx1} = (2 \ln N_e)/s$ [11].

If one is comparing two distinct mutations in the same population that are responsive to the same selective pressure, however, both the rates of mutation to the adaptive state and the selection coefficients may differ. For the second mutation, the expected waiting time to the appearance of the selected mutation may be written as $t_{w2} = 1/(2N_e vsr_v r_s)$, where r_v is the ratio of the mutation rates to the adaptive state and r_s is the ratio of the selection coefficients for the two cases. The expected time for the second mutation to spread to fixation in the population can be written $t_{fx2} = (2 \ln N_e)/r_s s$. Considering the case of a GOF versus LOF mutation, if we take v to be the nucleotide mutation rate, then in general r_v will range from 1 to ~1000 for an LOF mutation. r_s can take any positive value (both selection coefficients are positive because both the GOF and the LOF mutations are postulated to be adaptive).

A useful metric for comparing the prospects of fixation for the GOF versus LOF mutations is $r_{D/fx}$, which is defined as the expected time to appearance of an adaptive GOF mutation minus that for an adaptive LOF mutation, divided by the time for the LOF mutant to spread to fixation in the population. If the difference in the expected waiting times between the selected GOF versus LOF mutations is greater than the time required for the LOF mutation to spread, then the LOF mutation will have already fixed in the population before the expected appearance of the selected GOF mutation. The expected difference in waiting time to appearance of the selected mutations is

$$t_D = t_{wG} - t_{wL} = \frac{1}{2N_e vs} - \frac{1}{2N_e vsr_v r_s} = \frac{1}{2N_e vs}\left(1 - \frac{1}{r_v r_s}\right) \tag{1}$$

The ratio of the time difference t_D to the time for the LOF mutation to spread to fixation in the population, t_{fxL}, is

$$r_{D/fx} = \frac{t_D}{t_{fxL}} = \frac{\frac{1}{2N_e vs}\left(1 - \frac{1}{r_v r_s}\right)}{\frac{2}{r_s s} \ln N_e} = \frac{1}{4N_e v \ln N_e}\left(r_s - \frac{1}{r_v}\right) \tag{2}$$

Thus whenever $r_{D/fx} > 1$, the LOF mutation is expected to fix in the population before the selected GOF mutation appears. Figure 1 illustrates this situation. Two curves are plotted for the appearance and subsequent spread of an LOF and a GOF mutation in a population of 10^6 organisms. The selection coefficient for the GOF

is 0.1 and for the LOF is 0.01; thus r_s is 0.1. The basic nucleotide mutation rate is taken to be 10^{-9}, and r_v, the ratio of the mutation rate to the adaptive state for the LOF vs GOF mutation, is set at 100. The expected waiting time to the appearance of the selected LOF mutation under these circumstances is 500 generations, while for the GOF mutation the time is 5,000 generations. On average the GOF mutation would take 276 generations to fix in the population; the LOF mutation would require 2763 generations. Figure 1 shows that under such circumstances the selected LOF mutation would be expected to fix in the population before the selected GOF mutation appeared. Equation 2 determines the ratio $r_{D/fx}$ for this situation to be 1.62.

If $r_s = 1/r_v$, then equation 2 evaluates to zero, which means there is no expected difference t_D in the waiting time to the appearance of the selected LOF versus GOF mutations — the rate advantage of the LOF mutation is exactly offset by the relative weakness of its selection coefficient. If $r_s < 1/r_v$, then $r_{D/fx}$ will be negative, which means that there is less time to the appearance of the selected GOF mutation than to the LOF mutation — the rate disadvantage of the GOF mutation is more than offset by the relative strength of its selection coefficient.

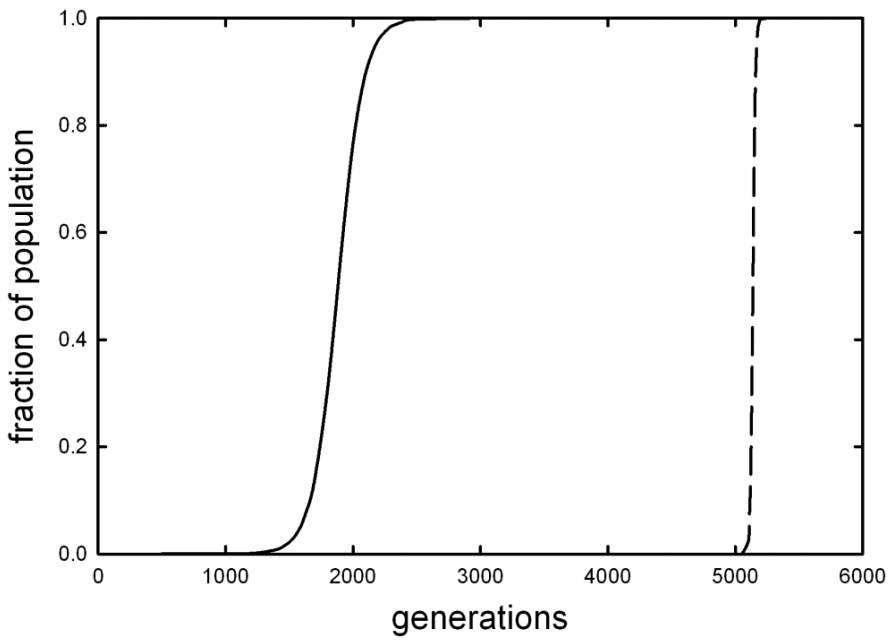

Fig. 1. Time in generations to occurrence and spread of an adaptive LOF mutation versus GOF mutation. The LOF mutant (———) has a selection coefficient 0.1-times that of the GOF mutant (— — —), but a mutation rate to the adaptive state 100-times that of the GOF mutant. The effective population size N_e is set at 10^6. The GOF mutation rate v is 10^{-9} per generation and the GOF selection coefficient $s = 0.1$.

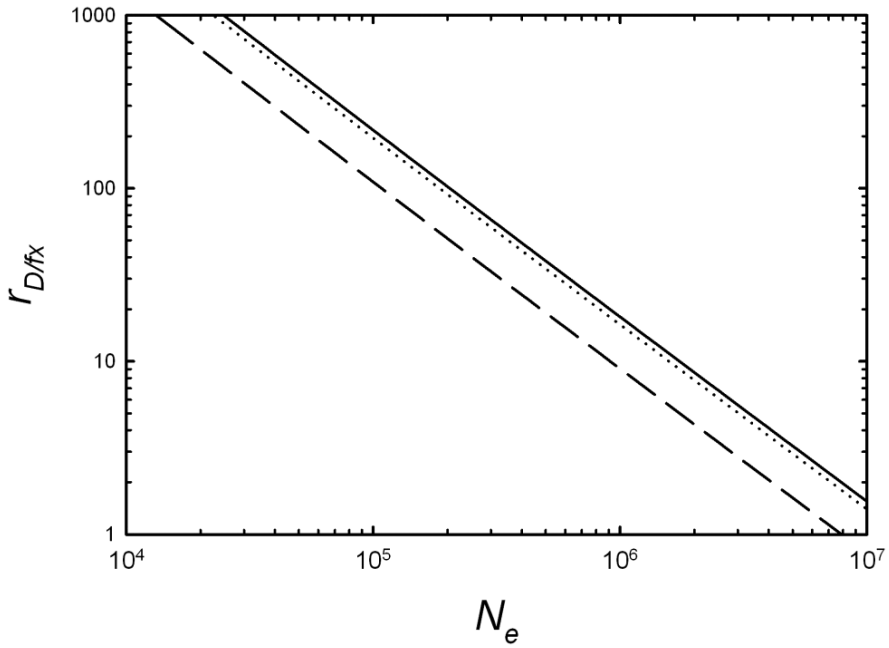

Fig. 2. The ratio $r_{D/fx}$ versus effective population size N_e. $r_{D/fx}$ is the time to appearance of an adaptive GOF mutation minus that for an adaptive LOF mutation, divided by the time for the LOF mutant to spread to fixation in the population. In this figure the LOF and GOF selection coefficients are equal. The mutation rate v is 10^{-9} per generation. (———) $r_v = 1000$; (·····) $r_v = 10$; (— — —) $r_v = 2$.

Figure 2 plots the value of $r_{D/fx}$ versus the effective population size N_e for several values of r_v, with r_s held constant at one. As can be seen, the value of $r_{D/fx}$ is largely insensitive to changes in r_v, the ratio of the mutation rates to the adaptive state. Decreasing r_v 100-fold from 1000 to 10 leaves the value of $r_{D/fx}$ little changed. In all of these circumstances (except where $r_v = 2$ at effective population sizes very near 10^7) the ratio of the time for the LOF mutation to spread to the difference in the expected waiting time to the selected GOF versus LOF mutations, $r_{D/fx}$, is well above one.

Figure 3 examines the relationship between the value of $r_{D/fx}$ versus the effective population size N_e for several values of r_s, with r_v held constant at 1000, its likely maximum for a typical gene. In this case $r_{D/fx}$ depends linearly on the ratio of the selection coefficients: at any population size in the range, a decrease of a factor of 10 in r_s decreases $r_{D/fx}$ by approximately the same factor. (The magnitude of s, the selection coefficient itself, which is absent from equation 2, does not affect the results.) Thus, when r_s is 0.01 (that is, when the selection coefficient for the LOF mutation is only 1% of that of the GOF mutation), $r_{D/fx}$ decreases to a value of one at a population size of about 1.5×10^5, versus a population size of 1.5×10^7 when r_s is one.

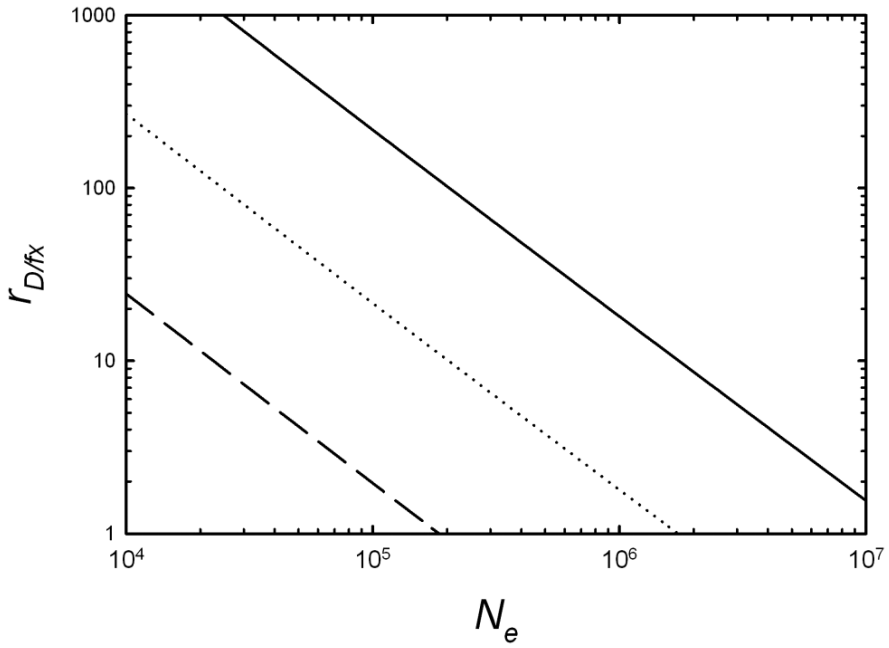

Fig. 3. The ratio $r_{D/fx}$ versus effective population size N_e. $r_{D/fx}$ is the time to appearance of an adaptive GOF mutation minus that for an adaptive LOF mutation, divided by the time for the LOF mutant to spread to fixation in the population. In this figure the rate of mutation to the adaptive state of the LOF mutant is 1000-times that of the GOF mutant. The mutation rate v is 10^{-9} per generation. (———) $r_s = 1$; (·····) $r_s = 0.1$; (— — —) $r_s = 0.01$.

Figure 4 shows the dependence of $r_{D/fx}$ on r_v and r_s at a fixed value of $N_e = 10^6$. As can be seen $r_{D/fx}$ is essentially independent of r_v over a wide range, but is linearly dependent on r_s. The pronounced curvature for both plots at lower values on the x-axis reflects the approach of the factor (r_s - $1/r_v$) to zero.

3.2 *Relatively large population sizes*

In this section I consider relatively large population sizes ($N_e \geq 1/v$). As population size increases, the expected waiting time to the appearance of either or both selected mutations can shrink to much less than the expected time for the mutations to spread in the population. In fact, one or both mutations may be present continuously in the population at a low percentage as a neutral or detrimental allele before the new selective pressure makes it adaptive. Thus in this population size range a different metric is required to follow the relative advantage of LOF versus GOF mutations.

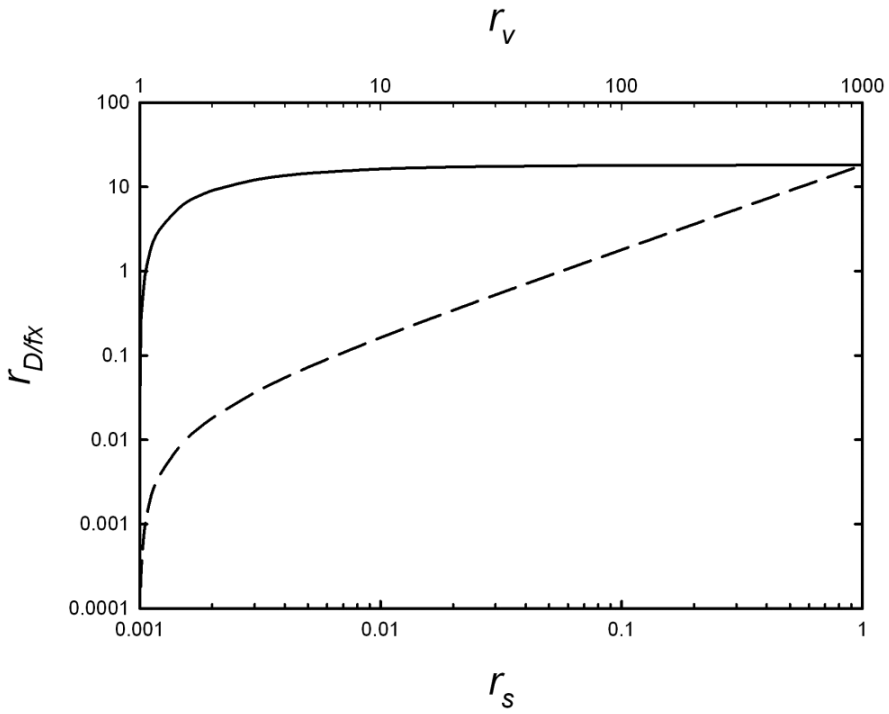

Fig. 4. The ratio $r_{D/fx}$ versus r_s and r_v. $r_{D/fx}$ is the time to appearance of an adaptive GOF mutation minus that for an adaptive LOF mutation, divided by the time for the LOF mutant to spread to fixation in the population. r_s is the ratio of the LOF to GOF selection coefficients. r_v is the ratio of the rate of LOF to GOF mutation to the adaptive state. The effective population size N_e is set at 10^6 and the GOF mutation rate v is 10^{-9} per generation. (———) r_s is set at 1 and r_v ranges from 1 to 1000; (— — —) r_v is set at 1000 and r_s ranges from 0.001 to 1.

A useful measure is the ratio of the fractions of LOF to GOF mutations in the population when the sum of those fractions first increases to 1.0. The time t in generations required to increase the frequency of a selected mutation from a value of q_0 to q_t can be calculated from [11]:

$$q_t = \frac{1}{1+\left(\frac{1-q_0}{q_0}\right)e^{-st}}$$

Thus (ignoring double mutants) the number of generations required for the fractions of an LOF and GOF mutation to sum to one can be calculated from:

$$\frac{1}{1+\left(\frac{1-q_{0G}}{q_{0G}}\right)e^{-st}}+\frac{1}{1+\left(\frac{1-q_{0L}}{q_{0L}}\right)e^{-r_s st}}=1 \qquad (3)$$

The initial fraction q_0 when a selected mutation begins to increase in a haploid population is at a minimum $1/N_e$. However, for population sizes greater than the inverse of the mutation rate, numerous mutants are expected to be present in the initial population. For example, if the mutation rate v is 10^{-9} and the population size is 10^{12}, then there will be 10^3 mutants produced in the first generation. So the initial fraction q_{0G} is at least $\frac{1}{N_e}+\frac{N_e v}{N_e}=\frac{1}{N_e}+v$, and q_{0L} is at least $\frac{1}{N_e}+r_v v$.

The time t in equation (3) is the time required for the selected mutation to spread. Thus if we are counting generations from the first application of the selective pressure, then the expected waiting time for the selected mutation must be accounted for. As mentioned previously, for a haploid GOF mutation this is $t_{wG} = 1/(2N_e vs)$ and for an LOF mutation $t_{wL} = 1/(2N_e vsr_v r_s)$. Equation (3) can then be re-written as:

$$\frac{1}{1+\left(\frac{1-q_{0G}}{q_{0G}}\right)e^{-s(t_{fx}-t_{wG})}}+\frac{1}{1+\left(\frac{1-q_{0L}}{q_{0L}}\right)e^{-r_s s(t_{fx}-t_{wL})}}=1 \tag{4}$$

where $(t_{fx} - t_w)$ is the time for the mutations to spread to a sum fraction of 1.0 after the waiting time for at least one kind of selected mutation to first appear in the population. Given N_e, v, s, r_v, and r_s, equation 4 can be solved for t_{fx} and the value used to determine r_{fx}, which is the fraction of adaptive LOF mutations divided by the fraction of adaptive GOF mutations in the population when the two fractions first sum to one:

$$r_{fx}=\frac{1+\left(\frac{1-q_{0G}}{q_{0G}}\right)e^{-s(t_{fx}-t_{wG})}}{1+\left(\frac{1-q_{0L}}{q_{0L}}\right)e^{-r_s s(t_{fx}-t_{wL})}} \tag{5}$$

Figure 5 plots r_{fx} from equation 5 at $r_s = 1$ and $r_v = 1000$ for population sizes N_e ranging from 10^7 to 10^{10}. It is seen that at lower values of N_e, r_{fx} increases very rapidly. Indeed, at population sizes of 10^7 or less, r_{fx} is greater than N_e, reflecting the fact that less than one GOF mutant is expected to be present in the population when the LOF mutant has fixed. As N_e increases, r_{fx} approaches a constant value of approximately 31.6. Thus when the population initially consists entirely of LOF and GOF mutants and $N_e \geq 10^9$, under the circumstances described in Figure 5 LOF mutants will represent about 97% of the population.

Figure 6 plots r_{fx} as a function of r_v for population sizes from $10^{6.5}$ to 10^{12}, with $r_s = 1$. At the smallest population sizes the fixation ratio is extremely sensitive to the ratio of mutation rates. As N_e increases, however, and it becomes more likely

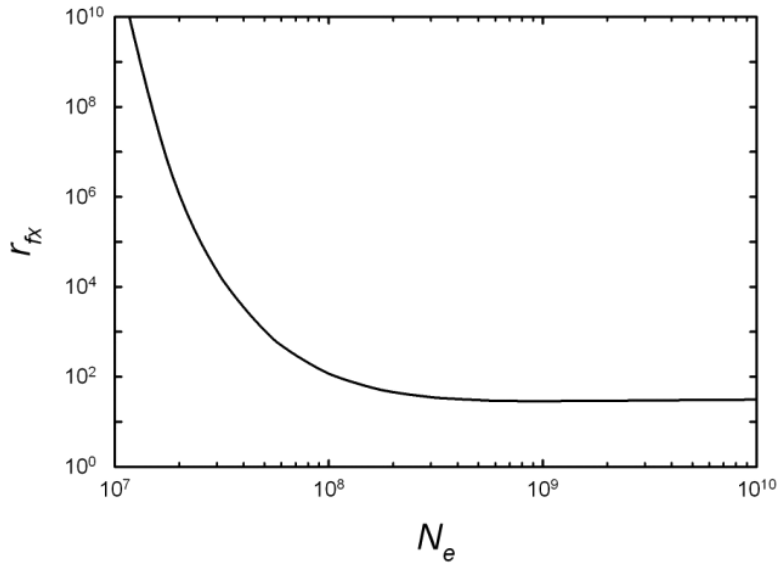

Fig. 5. The ratio r_{fx} versus the effective population size N_e. r_{fx} is the fraction of adaptive LOF mutations divided by the fraction of adaptive GOF mutations in the population when the two fractions first sum to one. $r_s = 1$; $r_v = 1000$; $s = 0.1$; $v = 10^{-9}$ per generation.

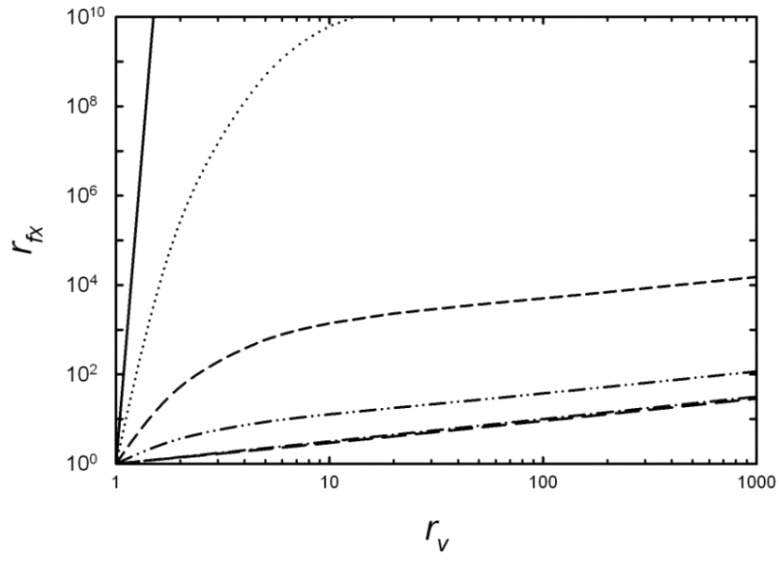

Fig. 6. The ratio r_{fx} versus the ratio r_v. r_{fx} is the fraction of adaptive LOF mutations divided by the fraction of adaptive GOF mutations in the population when the two fractions first sum to one. r_v is the ratio of the rate of LOF to GOF mutation to the adaptive state. r_s is set at 1; v is 10^{-9}. (———) $N_e = 10^{6.5}$; (·····) $N_e = 10^{7}$; (- - - - -) $N_e = 10^{7.5}$; (– ·· – ·· – ·· –) $N_e = 10^{8}$; (— — —) $N_e = 10^{9}$; (– · – · – · –) $N_e = 10^{12}$.

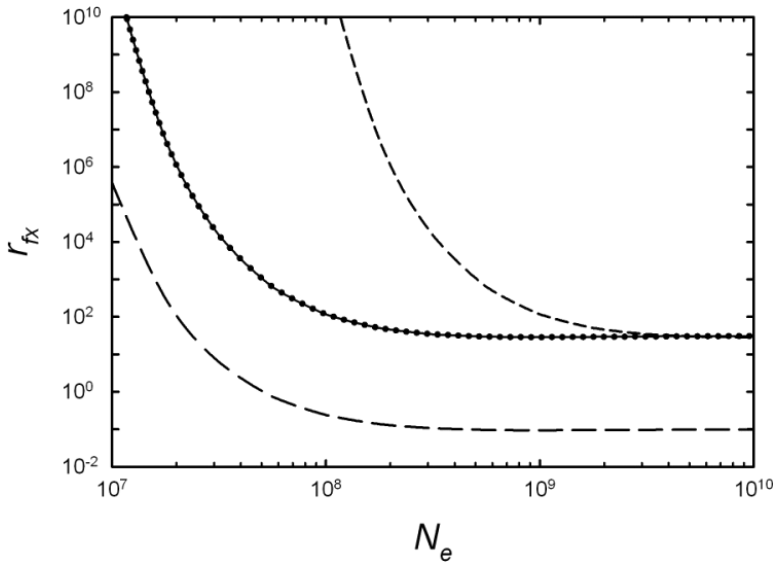

Fig. 7. The ratio r_{fx} versus the effective population size N_e. r_{fx} is the fraction of adaptive LOF mutations divided by the fraction of adaptive GOF mutations in the population when the two fractions first sum to one. For all curves r_v is set to 1000. (———) $s = 0.1$, $r_s = 1$, $v = 10^{-9}$; (·····) $s = 0.0001$, $r_s = 1$, $v = 10^{-9}$; (— — —) $s = 0.1$, $r_s = 0.5$, $v = 10^{-9}$; (- - - - -) $s = 0.1$, $r_s = 1$, $v = 10^{-10}$.

that the mutants are present in the population from the first generation, the sensitivity decreases. As seen in the figure, the plots of r_{fx} versus r_v for values of $N_e \geq 1/v$ are essentially superimposable, and closely approximate the relationship $r_{fx} = \sqrt{r_v}$.

Figure 7 plots r_{fx} versus N_e for several variables. The solid curve reproduces the values from Figure 5 of $s = 0.1$ and $r_s = 1$. Coinciding with the solid curve is a dotted curve for which $s = 0.0001$, demonstrating the insensitivity of the curve to changes in the selection coefficient itself. The long-dashed curve uses the same parameters as the solid curve except that the value of r_s has been decreased to 0.5. As can be seen, this decreases the value of r_{fx} by several orders of magnitude, so that at large population sizes the value is below one, and the GOF mutation predominates at fixation, despite the initial 1,000-fold advantage of the LOF mutation rate. The short-dashed curve uses the same parameters as the solid curve except that the value of v has been decreased from 10^{-9} to 10^{-10}. As can be seen, this has the effect of simply moving the curve an order of magnitude to the right on the population axis.

Figure 8 plots r_{fx} versus r_v at three values of r_s with $N_e \gg 1/v$. As seen, modestly varying the ratio of the selection coefficients displaces the curve considerably along the r_{fx} axis and slightly alters its slope. Figure 9 compares r_v and r_s versus

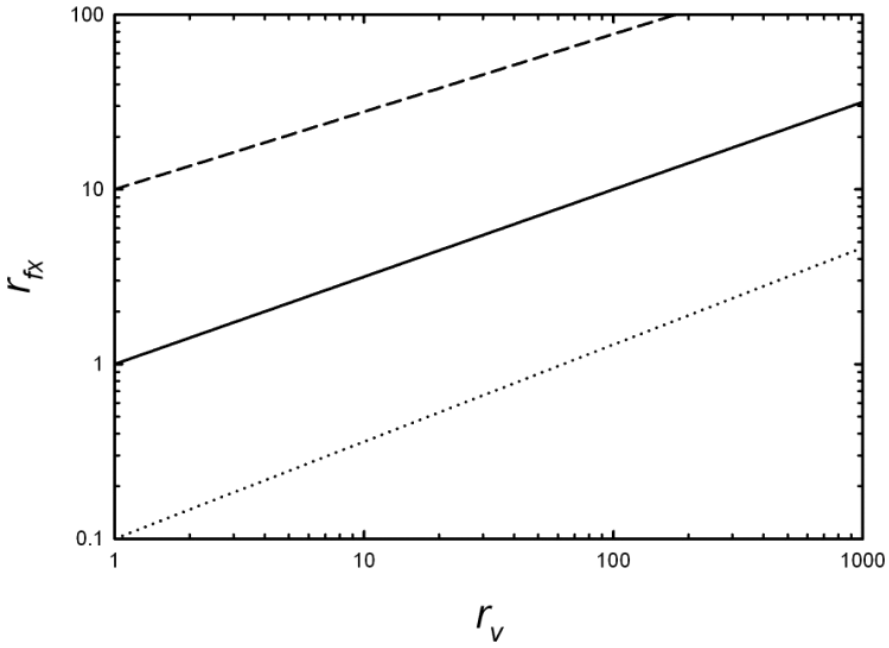

Fig. 8. The ratio r_{fx} versus the ratio r_v. r_{fx} is the fraction of adaptive LOF mutations divided by the fraction of adaptive GOF mutations in the population when the two fractions first sum to one. r_v is the ratio of the rate of LOF to GOF mutation to the adaptive state. For all curves, N_e is set at 10^{12} and v is 10^{-9}. (———) $r_s = 1$; (·····) $r_s = 0.8$; (- - - - -) $r_s = 1.25$.

r_{fx}, showing the relative sensitivity of the fixation ratio to those parameters at large N_e. Figure 9 plots values for r_s including from one to 100; that is, for situations in which the selection coefficient of the LOF mutation is greater than or equal to that of the GOF mutation. r_{fx} is greater than one and increases rapidly in this region. In general, whenever $r_s \geq 1$ and $r_v > 1$, r_{fx} will be greater than one at any population size. That is, the LOF mutation will always be the majority of the population when the entire population is initially comprised of LOF and GOF mutations.

4. Discussion

4.1 *LOF versus GOF adaptive mutations*

Organisms can adapt to their environment either by acquiring new abilities or by abandoning old ones. This can be observed in such examples as legless snakes and sightless cave fish. Science has learned especially in the last fifty years that altered, observable phenotypes are the manifestation of changes to the genetic endowment

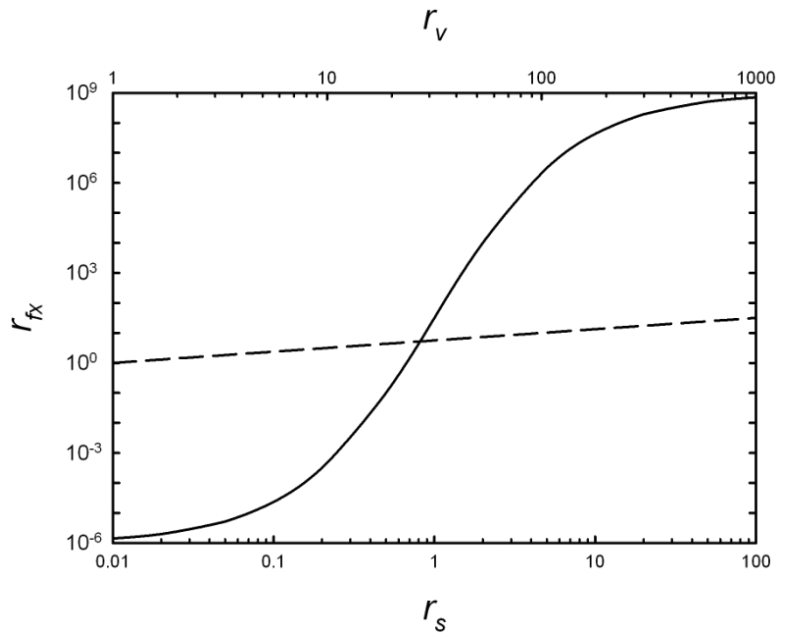

Fig. 9. The ratio r_{fx} versus r_s and versus r_v. r_{fx} is the fraction of adaptive LOF mutations divided by the fraction of adaptive GOF mutations in the population when the two fractions first sum to one. r_s is the ratio of the LOF to GOF selection coefficients. r_v is the ratio of the rate of LOF to GOF mutation to the adaptive state. For both curves, N_e is set to 10^{12} and v is 10^{-9}. (———) r_{fx} versus r_s, r_v is set at 1000; (— — —) r_{fx} versus r_v, r_s is set at 1.

of an organism. It has also learned that there is not a necessary correlation between loss or gain of an ability at the phenotypic level and loss or gain of a functional genetic element at the molecular level. In other words, what strikes an observer as a phenotypic gain of function may be caused by either a molecular loss or gain of function. The same holds for a phenotypic loss of function: it may be the result of a genetic gain or loss of function. Because organisms can adapt by either molecular GOF mutations or LOF mutations it is of basic interest to determine which, if either, kind of mutation will dominate under various circumstances.

Research over the past fifty years has shown that many genetic elements consist of multiple nucleotides. Protein coding regions can be thousands of nucleotides in length; RNA genes can be hundreds of nucleotides; regulatory elements and processing signals can be several nucleotides to dozens of nucleotides long. A substantial portion of possible mutations in these elements will result in the diminution or loss of their function. Thus, as a class, LOF mutations for a particular genetic element will occur at a rate from several times to several-orders-of-magnitude times the basic nucleotide substitution rate.

That is not the case for GOF mutations. Consider two examples: First, a transcription factor binding site that is 10 nucleotides in length, and a second DNA sequence which has 9 of 10 nucleotides that are necessary to form a second regulatory site. Suppose that in response to a new selective pressure an adaptive effect could be secured either by mutating the first site so that it lost its function or by mutating the single mismatching residue of the second site so that it gained function. The LOF mutation would on average appear at 10-times the nucleotide substitution rate, simply because there are multiple ways to break the functioning element. The GOF mutation, however, would appear at even less than the basic rate of nucleotide substitution (because for a currently-nonfunctional, potential genetic element there it is possible that one of the "correct" nucleotides in the sequence will mutate before the "incorrect" one [12]). Second, consider a recently duplicated gene which could provide an adaptive effect in response to a new selective pressure if a certain nucleotide in the gene were altered (allowing the duplicate gene product to, say, diverge productively in activity from the parent gene product). Suppose, however, that an adaptive effect could also be had by reducing or eliminating the activity of another, separate gene. Because of the many ways in which a gene can be altered to lose function, the LOF mutation would have a rate several orders of magnitude greater than that of the GOF mutation for the duplicated gene.

There can be cases in which a GOF mutation may appear at several times the nucleotide substitution rate. I discussed earlier the sickle mutation, in which a single particular nucleotide in the β-globin gene must be changed. Yet in other cases of GOF, there can be several possible nucleotides to change, each of which will suffice. For example, Couñago *et al*. [13] replaced the essential gene for adenylate kinase in *Geobacillus stearothermophilus* — a moderate thermophile — with that of *Bacillus subtilis*, a mesophile, which they then grew in a turbidostat at increasing temperatures. Over the course of 1500 generations they isolated six thermostable mutants of the enzyme — one single point mutant and five double point mutants derived from the single mutant. Thus in this circumstance the enzyme could gain the function of being active in a hostile environment by altering any of six positions. Nonetheless, the number of ways to break a functional element will almost always be much greater than the number of ways to construct one, so that in almost all cases r_v would be expected to be greater than one.

4.2 *Effect of disparity in adaptive rate*

In this chapter I investigate the effect of the disparity in rate of mutation to an adaptive state for LOF and GOF mutations as a function of several parameters.

The model presented here is a simple, deterministic one, which does not consider the probabilistic nature of changes in allele frequencies [11]. Because of its simplicity, the general behavior of the investigated model is visible with considerable clarity and the issue of the evolutionary rate advantage of adaptive LOF mutations is highlighted.

The behavior at relatively small population sizes is governed by equation 2, which accounts for the two separate phases of fixation of a new mutation: the expected waiting time to the appearance of the selected mutation, and the time taken for the mutation to spread within the population. An interesting aspect of the equation is that it does not contain the selection coefficient *s*; that is, the ratio of the selection coefficients r_s influences the competition between the two mutations rather than the absolute value of either or both selection coefficients. (This also is the case at relatively large population sizes, as shown by Figure 7.) Whenever equation 2 evaluates to $r_{D/fx} > 1$, then the LOF mutation is expected to fix in the population before a selected GOF mutation appears. Thus, as illustrated in Figure 1, an LOF mutation whose selection coefficient is ten-fold weaker than an adaptive GOF mutation can outrace it to fixation, due to its greater rate of mutation to an adaptive state.

Figures 2 and 3 show that this effect exerts substantial influence at relatively low population sizes. For a population size of $< 10^7$, if $r_s \geq 1$ and $r_v > 1$, then an LOF mutant is always expected to fix in the population before a selected GOF mutant appears. Because of an increasing disparity in waiting times, at population sizes $<<10^7$ an LOF mutant may be fixed in the population first even if its selective advantage is considerably less than that of a GOF mutant. For example, for a population size of 10^5, an LOF adaptive mutation will become fixed first at $r_v \geq 1$ even if its selection coefficient is only one-hundredth that of a GOF adaptive mutation; that is, if $r_s \geq 0.01$. At smaller population sizes, the advantage for the LOF mutation increases linearly with $1/N_e$.

If an LOF mutation with a smaller selection coefficient is first fixed in a population, what scenario is most likely to occur after the GOF mutation eventually appears? The answer to that question is likely to depend sharply on the specific genetic elements involved. One possible scenario is that the GOF mutation also spreads to fixation, and the LOF mutation remains fixed. A second possibility is that, depending on the physical nature of the mutation, the LOF mutation may be repaired by subsequent mutation after the GOF mutation spreads in the population. If it cannot be repaired, it may be replaced by horizontal gene transfer or by having its function taken over by another genetic element, or the organism may adapt in other ways to its loss. Penman *et al*. [14] recently demonstrated that the outcome in competition between the sickle mutation (which is highly protective against malaria) and various thalassemic

disorders (which are less protective) is quite difficult to predict because of epistatic effects unrelated to their anti-malarial activities. Thus the future course of the evolution of a system after initial fixation of an LOF mutation might be considerably more complex than a linear succession of mutations with increasing selective value.

For $v = 10^{-9}$, at population sizes $N_e > 10^8$ an LOF mutation is no longer expected to fix in the population before a selectable GOF mutation appears, even if $r_s = 1$, because the larger population sizes produce both types of mutations within the time it would take for the LOF mutation to spread in the population. Nonetheless, even though the metric $r_{D/fx}$ decreases below one in this range, in many cases the LOF mutation will become the dominant mutation in the population. In order to assess the advantages of LOF versus GOF mutations in this population range, a new metric, r_{fx}, was introduced in equation 5. r_{fx} is the ratio of LOF to GOF mutants when their fraction of the population first sums to one.

Figure 6 shows that LOF mutations always possess a rate advantage over GOF mutations if the respective selection coefficients are equal; that is, if $r_s = 1$. Under these circumstances at large population sizes ($N_e \geq 1/v$), $r_{fx} \approx \sqrt{r_v}$, and the ratio of LOF to GOF mutations when their fraction first sums to one will range from 1.41 to 31.6 for values of r_v ranging from 2 to 1000. Thus the LOF mutant will comprise from 59% of the population to 97% of the population. If at this point the mutants then drift neutrally in the population (because it is postulated that neither has a selective advantage over the other), the LOF mutant is expected to become fixed with a probability equal to its population fraction [15].

Under what circumstances would two selection coefficients be expected to be equal? If two mutations both met the new selective pressure without causing deleterious pleiotropic effects, then their selection coefficients would be expected to be the same. Thus whenever such a situation presents itself, the LOF mutation would have an advantage.

If the selection coefficients are not equal, how likely is it that a GOF mutation will have a value of s greater than that of an LOF mutation, or vice-versa? The answer to that question is not known, but both LOF and GOF mutations can have significant selection coefficients. The selection coefficient for LOF mutations of the *rpoS* gene of *E. coli* has been measured at 0.217, a substantial value [16]. The selection coefficient for the GOF sickle mutation has been estimated as 0.05 to 0.18, again a large value [17]. If in general there is no overall correlation between adaptive GOF versus LOF mutations and the magnitude of the selection coefficient, then the intrinsic rate advantage enjoyed by LOF mutations will bias long-term evolution in that direction.

4.3 *Comparison to laboratory evolution experiments*

Over the past forty years many laboratories have conducted evolution experiments, observing adaptation of micro-organisms to varying environmental conditions, and in many cases identifying the molecular changes that comprised the adaptive mutation [4]. How do the results obtained in this chapter bear on the interpretation of those experiments?

Comparison to experiments where $N_e < 1/v$: The most extensive laboratory evolution experiment to date has been performed under the direction of Richard Lenski at Michigan State University [18]. Starting in the early 1990s, Lenski and colleagues began growing 10 ml cultures of *E. coli*, which undergo six to seven doublings per day. Each day they transferred 1% of the culture to fresh medium. Over the years the cultures have undergone more than 50,000 generations. All adaptive mutations identified to date appear to be LOF ones [4]. The single most beneficial mutation was the destruction of the *rbs* operon by insertion sequences. The value of the selective coefficient for this was approximately 0.02 [19]. Other identified LOF mutations include ones in the *pykF, nadR, pbpA-rodA, hokB/sokB*, *malT*, and *topA* genes. A number of other adaptive genes have been identified to date, but the natures of the mutations, whether LOF or GOF, have not yet been reported [20].

The rate of nucleotide mutations per generation of *E. coli* is ~5×10^{-10} [21]. The effective population size N_e of Lenski's [18] cultures of *E. coli* is ~2×10^7, which is the harmonic mean between the initial population of the day's culture (5×10^6) and the final population of the day (5×10^8) if the population is assumed to double in discrete generations [11]. Substituting these numbers into equation 2 shows that $r_{D/fx}$ would be 1.47 — greater than one — if r_s were one and r_v were 100. An LOF mutation would thus be expected to be fixed in the population before a GOF mutation appeared if their selection coefficients were equal. How great of a selective advantage must a GOF mutation have to outcompete an LOF mutation under these circumstances? Using equation 2 it is seen that if r_s is 0.68, then $r_{D/fx}$ falls slightly below one. In other words, a GOF mutation would have to have a selection coefficient about 50% greater than an LOF mutation in these circumstances in order to at least appear in the population before the LOF mutation were fixed.

To find out how much greater the selection coefficient must be to actually outcompete the LOF mutation, we must use equations 4 and 5 to calculate r_{fx}. Assuming r_s were 0.68, there would be approximately one GOF allele in the population per ~2×10^7 LOF alleles. In order to overcome the LOF rate advantage, however, r_s would have to fall to ~0.25. In other words, if the selection coefficient of the GOF mutation were approximately four times that of the LOF mutations,

then the GOF mutation would be slightly more than half the population. In order to dominate the population by ~90% r_s would have to be ~0.2; that is, the selection coefficient of the GOF mutation would have to be about five-fold that of the LOF mutation. Since no GOF mutations have yet been seen, we can tentatively conclude that there are no GOF mutations available whose selective value is five-fold greater than the least-adaptive LOF mutations seen in this series of experiments. (Lenski's group recently reported a very adaptive Cit+ phenotype, which apparently required both LOF and gene duplication mutations [22]). If an LOF mutation appeared within the first 25,000 generations, it would require a minimum selection coefficient of 0.00076 to spread to fixation in the next 25,000 generations. To outcompete it, a GOF mutation would require a minimum selection coefficient of five-times this value, i.e. ~0.0038. Thus it can be concluded that there are no GOF mutations available under the circumstances of the experiment whose selection coefficients exceed that number.

The question might be asked, what if a potential GOF mutation with a sufficiently strong selection coefficient existed, but simply failed to arise during the term of the experiment? That is always a possibility, but an unlikely one. Given the scale of the Lenski experiment [20], with an effective population size of 2×10^7 over 50,000 generations and a nucleotide mutation rate of $\sim5\times10^{-10}$, each nucleotide is expected to be substituted 500-fold over the course of the experiment. Deletions, additions, and other kinds of mutations would similarly be expected to occur multiple times. There were many redundant opportunities for all simple mutations to arise (the Cit^+ phenotype apparently needed several mutations to arise). Thus we can be confident that if a particular mutation, or kind of mutation, was not observed, then it is very unlikely to have the necessary selection coefficient.

Comparison to experiments where $N_e > 1/v$: As seen in Figures 7–9, at $N_e \geq 1/v$, r_{fx} is much more sensitive to r_s than at smaller population sizes. Just a slight advantage in the selection coefficient for a GOF mutation is sufficient to offset a 1,000-fold advantage in the rate of LOF mutation. This great sensitivity can be used to infer whether such a GOF mutation is available under particular environmental circumstances. That is, if a certain selective pressure is applied, one or more LOF mutations are observed, and $N_e \geq 1/v$, then the failure to observe a GOF mutation would imply that no GOF mutation is available within a single mutational step that had a somewhat greater selection coefficient than the LOF mutations(s). Conversely, if a GOF mutation were observed but no LOF mutation, we could deduce that no LOF mutation was available that had a selection coefficient greater than or equal to the GOF mutation.

As mentioned earlier, Couñago *et al*. [13] replaced the essential gene for adenylate kinase in *Geobacillus stearothermophilus* — a moderate thermophile — with that of

Bacillus subtilis, a mesophile, which they then grew in a turbidostat at increasing temperatures. Over the course of 1500 generations they isolated six thermostable mutants of the enzyme — one single point mutant and five double point mutants derived from the single mutant, which can all be classified as GOF mutations. The mutation rate of *G. stearothermophilus* can be estimated by using a value for the mutation rate of approximately 0.003 per genome per generation for DNA-based microbes, which yields a value of about 5×10^{-10} mutations per generation [21]. Since the authors maintained a continuous culture, the population was not subject to the large changes in size seen in Lenski's experiments, so the effective population of microbes per generation in the turbidostat was ~5×10^{10}. In other words, in the Couñago *et al*. [13] experiment, $N_e > 1/v$. Inserting these values into equations 4 and 5 shows that if a potentially adaptive LOF mutation were available with the same selection coefficient as a GOF mutation, then it would dominate the population with an r_{fx} of 9.9; in other words, it would comprise ~91% of the population. Thus it can be concluded that, despite the frequency of adaptive LOF mutations in Lenski's work, no LOF mutation with an $r_s \geq 1$ compared to the observed GOF mutations was available in the experiment conducted by Couñago *et al*. [13]. The likely reason for the disparate results is the differing experimental regimens. Lenski did not put strong constraints on the direction for *E. coli* to evolve, but Couñago *et al*. [13] replaced an essential gene with a substitute optimized for a different growth temperature before applying selective pressure, which they termed a "weak link" method. Furthermore, Couñago *et al*. [13] used an N_e that was more than three orders of magnitude greater than Lenski's group. The activity of the thermophilic adenylate kinase activity had to be replaced or compensated for. Apparently, the fastest way available to do so at high N_e was by GOF point mutations to the mesophilic substitute gene.

4.3.1 *Comparison to experiments where two selective routes were potentially available*

An interesting conceptual blend of the Lenski [18] and Couñago *et al*. [13] approaches was recently published by Gauger *et al*. [23]. This group mutated two amino acid residues of a plasmid-borne trpA gene of *E. coli*, transfected a Trp^- bacterial strain with the plasmid, and grew it in a tryptophan-limiting medium. One of the mutations (E49V) alone completely inactivates the gene product; the other mutation (D60N), when present alone, allows weak Trp^+ activity and supports growth in Trp^- media when the plasmid-borne gene is overexpressed. The authors expected cells containing the double mutant plasmid to take a short, selected route to full Trp^+ activity when grown in tryptophan-limiting medium by first reverting the inactivating mutation at position 49 (allowing the resumption of

weak Trp^+ activity) and then reverting the second mutation at position 60 to regain full activity. However, almost all mutants recovered after sustained growth had not taken even the first step on that expected pathway. Rather, the expression of the *trpA* gene was decreased either by deletion, insertion of an IS element, or by various point mutations, apparently saving the cell the energy of overproducing the protein.

The *E. coli* point mutation rate is 5×10^{-10}. Gauger *et al*. [23] grew liquid cultures to an effective population size N_e of ~0.6×10^7 cells per generation. Substituting these numbers into equation 2 shows that $r_{D/fx}$ would be 5.3 — greater than one — if r_s were one and r_v were 100. That is, if the selective advantage the cell received from shutting down overexpression of the plasmid-borne gene were equal to the selective advantage it would receive from taking the first GOF mutational step to partial Trp^+ activity, the LOF mutation would be expected to easily be fixed in the population well before a GOF mutation appeared. For one partial-revertant to be expected to appear before the LOF mutant fixed under the conditions of the experiment r_s would have to be about 0.2. That is, the selection coefficient for the GOF mutation would have to be approximately five-fold greater than that of the LOF mutation. Equations 4 and 5 can be used to show that for the GOF mutant to be expected to dominate the population to >90%, the GOF selection coefficient would have to be about 12.5-times that for the LOF mutation. Apparently, regaining merely limited Trp^+ activity did not have 12.5-times the selective value of the decrease in expression of the plasmid gene caused by the LOF mutations. Thus, under the conditions of the experiment, the selective pathway back to full Trp^+ activity is blocked at the first step. Interestingly, if cells transfected with either singly-mutated plasmid (E49V or D60N) were grown in liquid culture, Trp^+ revertants quickly took over the culture, indicating the selection coefficient for full-reversion was greater than 12.5-times the selection coefficient for saving the cell the energy of overproducing the protein [23].

4.4 *Comparison to short-term evolution in the wild*

A possible objection to results from laboratory evolution experiments is that they are artificial. The organisms are housed in special environments and not exposed to the rigor and variety of challenges they would encounter in nature. Thus the many advantageous LOF mutations observed in experimental work may not reflect what happens in nature, since presumably the great majority of an organism's genes are required in the wild, and therefore few if any adaptive LOF mutations are available in nature.

While that may turn out to be the case, and more data will be required to come to a definitive conclusion, an increasing number of results from nature appear to ratify the importance of adaptive LOF mutations in the wild. One class of such LOF mutations which I have mentioned previously includes genes that help adapt humans to the presence of malaria [4]. Other important human adaptive mutations are also LOF mutations: immunity to HIV due to a deletion variant of CCR-5 [24]; and resistance to tuberculosis by a deletion variant of SLC11A1 [25]. Development of lactose tolerance in adult humans [26] also seems a good candidate for an adaptive LOF mutation, perhaps by loss of a repressor binding site, although that has not yet been confirmed. In a recent survey of multiple human genomes it has been determined that for humans, "On average, each person is found to carry approximately 250 to 300 loss-of-function variants in annotated genes...," over 1% of the total number of human genes [27].

A second example of LOF mutation in nature is seen in the evolution of the plague bacterium *Yersinia pestis*. A plausible evolutionary scenario to explain its great virulence is that it serially acquired several plasmids which conferred on it the ability to be transferred between mammalian hosts by flea bite [28, 29]. After the acquisition of these plasmids (which are GOF events), the *Y. pestis* genome lost several hundred genes, apparently because they were no longer necessary for its new life cycle [29, 30]. Thus, after several GOF events, the plague bacterium adjusted to its new environment by much more numerous and rapid LOF adaptive mutations.

Nadeau and Jiggins [31] have recently reviewed genomic studies of adaptation in natural populations and note that "Many of the well-studied examples of adaptive evolution have involved trait loss, such as the loss of bony structures in freshwater stickleback populations and the reduction of pigmentation and eyes in cavefish." Although, as mentioned earlier in this chapter, there is not a necessary correlation between phenotypic trait loss and adaptive LOF mutations, in the cases mentioned by Nadeau and Jiggins [31] they coincide. Loss of pelvic spines in freshwater sticklebacks has been traced to deletion of a Pitx1 enhancer [32]. Eye reduction in cavefish apparently involves multiple genes [33]. Of those that have been identified three involve decreased expression of the gene (*γ-M crystallin*, *rhodopsin*, and *αA crystallin*). One gene, *hsp90α*, has increased expression, and it appears to be involved in promoting apoptosis.

5. Conclusion

Organisms have adapted over evolutionary history both by gaining and losing functions. Therefore it is of basic interest to determine if one or the other

dominates during particular circumstances. Until the past few decades, however, the molecular events underlying these processes were obscure. In recent decades science has in some cases gained the ability to determine whether the events behind a phenotypic adaptation involve an adaptive GOF mutation or an adaptive LOF mutation [4].

Both experimental laboratory work over the past few decades and recent genomic studies of adaptation in natural populations attest to the importance, even dominance, of LOF mutations in short term evolutionary episodes. The work presented in this paper helps show why this should be the case. Functional genetic elements such as genes and regulatory regions are built of multiple nucleotides, and a substantial fraction of mutations to these elements will cause them to lose their function. Thus the LOF mutation rate can be orders of magnitude greater than the nucleotide substitution rate. On the other hand, GOF mutations tend to be quite specific. So the rate for adaptive GOF mutations tends to be equal or very similar to the nucleotide mutation rate. As shown here, for some population size regions and for some values for the ratio of selection coefficients, the greater rate of mutation to the adaptive state for LOF versus GOF gives adaptive LOF mutations an intrinsic edge over adaptive GOF mutations.

In retrospect, the result is straightforward. Yet it also seems somewhat surprising because, as Nadeau and Jiggins [31] write, "there clearly are complex structures that are gained during evolution ... and we currently know little about how this process takes place." It may be hoped that understanding how organisms survive in the short term by adaptive LOF mutations will be a step toward understanding how complex structures are built over the long term.

References

1. Darwin C (1859) The Origin of Species. Bantam Books, New York.
2. Stott R (2003) Darwin and the Barnacle. W. W. Norton, New York.
3. Chang JC, Kan YW (1981) Antenatal diagnosis of sickle cell anaemia by direct analysis of the sickle mutation. Lancet 2:1127–1129.
4. Behe MJ (2010) Experimental evolution, loss-of-function mutations, and "The First Rule of Adaptive Evolution". Quart Rev Biol 85:419–445.
5. Friedman MJ (1978) Erythrocytic mechanism of sickle cell resistance to malaria. Proc Natl Acad Sci USA 75:1994–1997.
6. Verra F, Simpore J, Warimwe GM, Tetteh KK, Howard T, Osier FH, Bancone G, Avellino P, Blot I, Fegan G, Bull PC, Williams TN, Conway DJ, Marsh K, Modiano D (2007) Haemoglobin C and S role in acquired immunity against *Plasmodium falciparum* malaria. PloS ONE 2:e978.

7. Ruwende C, Khoo SC, Snow RW, Yates SN, Kwiatkowski D, Gupta S, Warn P, Allsopp CE, Gilbert SC, Peschu N, Newbold CI, Greenwood BM, Marsh K, Hill AVS (1995) Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. Nature 376:246–249.
8. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The History and Geography of Human Genes. Princeton University Press, Princeton, NJ.
9. Lynch M (2010) Rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci USA 107:961–968.
10. Perfeito L, Fernandes L, Mota C, Gordo I (2007) Adaptive mutations in bacteria: high rate and small effects. Science 317:813–815.
11. Li WH (1997) Molecular Evolution. Sinauer Associates, Sunderland, Mass.
12. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155.
13. Couñago R, Chen S, Shamoo Y (2006) In vivo molecular evolution reveals biophysical origins of organismal fitness. Mol Cell 22:441–449.
14. Penman BS, Pybus OG, Weatherall DJ, Gupta S (2009) Epistatic interactions between genetic disorders of hemoglobin can explain why the sickle-cell gene is uncommon in the Mediterranean. Proc Natl Acad Sci USA 106:21242–21246.
15. Kimura M (1983) The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.
16. Ferenci T (2008) The spread of a beneficial mutation in experimental bacterial populations: the influence of the environment and genotype on the fixation of *rpoS* mutations. Heredity 100: 446–452.
17. Li WH (1975) The first arrival time and mean age of a deleterious mutant gene in a finite population. Am J Hum Gen 27:274–286.
18. Lenski RE (2004) Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. Plant Breeding Rev 24:225–265.
19. Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J Bacteriology 183: 2834–2841.
20. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461: 1219–1221.
21. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148:1667–1686.
22. Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489: 7899-7906.
23. Gauger AK, Ebnet S, Fahey PF, Seelke R (2010) Reductive evolution can prevent populations from taking simple adaptive paths to high fitness. BIO-Complexity 2:1–9.

24. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, Saragosti S, Lapoumeroulie C, Cognaux J, Forceille C, Muyldermans G, Verhofstede C, Burtonboy G, Georges M, Imai T, Rana S, Yi Y, Smyth RJ, Collman RG, Doms RW, Vassart G, Parmentier M (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382:722–725.
25. Barnes I, Duda A, Pybus OG, Thomas MG (2010) Ancient urbanization predicts genetic resistance to tuberculosis. Evolution doi:10.1111/j.1558–5646.2010. 01132.x
26. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P (2007) Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genetics 39:31–40.
27. The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.
28. Wren BW (2003) The yersiniae–a model genus to study the rapid evolution of bacterial pathogens. Nat Rev Microbiol 1:55–64.
29. Carniel E (2003) Evolution of pathogenic Yersinia: Some lights in the dark. In: Skurnik M, Bengoechea JA, Granfors K (eds) The Genus Yersinia: Entering the Functional Genomic Era, Kluwer Academic/Plenum, New York.
30. Chain PS, Hu P, Malfatti SA, Radnedge L, Larimer F, Vergez LM, Worsham P, Chu MC, Andersen GL (2006) Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. J Bacteriol 188:4453–4463.
31. Nadeau NJ, Jiggins CD (2010) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends Gen 26:484–492.
32. Chan YF, Marks ME, Jones FC, Villarreal G Jr., Shapiro MD, Brady SD, Southwick AM, Absher DM , Grimwood J, Schmutz J, Myers RM, Petrov D, Jónsson B, Schluter D, Bell MA, Kingsley DM (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327:302–305.
33. Jeffery WR (2009) Regressive evolution in *Astyanax* cavefish. Ann Rev Genetics 43:25–47.
34. Wolfram S (2003) *The Mathematica Book*. Wolfram Media, Champaign, IL.