

# Fitness Decline Over Long-Term Evolution of a Small Population of Asexual Computational Organisms

Douglas D. Axe\*, Jireh Gerry, Alisa D. Daniels, Sabrina Wilkerson, William Mitchell, and Sarah Randall

Biola University, La Mirada, CA

## Abstract

The vital information carried in the DNA of every organism must be protected from the mutagenic processes that tend to degrade it. Molecular systems that detect and correct chemical alterations and base-pair mismatches form the first line of defense in all kingdoms of life. Natural selection provides a second line of defense. Specifically, purifying selection (or negative selection) is the natural tendency in wild populations for the genetic lines of individuals that suffer impairment from a new mutation to terminate within a few generations of the mutation event. Purifying selection is known to be much more efficient than positive selection, though it becomes less efficient in very small populations. Here, we describe a long-term evolution experiment that tracks fitness in a population of 1,000 computational organisms. We use the previously described *Stylus* artificial-world model, in which genes encode drawings that have scorable functionality based on the degree to which they resemble one of the Chinese written characters. In our weakest-link minimal-genome model, the fitness of the organism is proportional to the lowest score of its 223 essential genes. Following a population of 1,000 model organisms through 2,000,000 fixation events, we find that fitness declines approximately as a two-phase exponential decay. The long-term result is substantial loss of function for all 223 genes, seen both in a collapse of numerical scores and in loss of legibility. The cause of the collapse is an imbalance in the initial genome: the number of ways for mutations to produce a new worst gene is so much higher than the number of ways to improve the current worst that selection is unable to prevent decline. This result raises the question of genome decay in real organisms. It seems likely that the same imbalance exists there, though life may have ways of averting the decline we describe here.

**Cite as:** Axe DD, Gerry J, Daniels AD, Wilkerson S, Mitchell W, et al. (2023) Fitness decline over long-term evolution of a small population of asexual computational organisms. *BIO-Complexity* 2023 (2):1–9. doi:10.5048/BIO-C.2023.2.

**Editor:** Robert J. Marks II

**Received:** September 12, 2022; **Accepted:** July 27, 2023; **Published:** September 30, 2023

**Copyright:** © 2023 Axe. This open-access article is published under the terms of the [Creative Commons Attribution License](#), which permits free distribution and reuse in derivative works provided the original author(s) and source are credited.

**Notes:** A Critique of this paper, when available, will be assigned doi:10.5048/BIO-C.2023.2.c

\*douglas.axe@biola.edu

## INTRODUCTION

All species depend upon genetic information stored in their genomic DNA. Because natural mutagenic processes can degrade this information, life must continually use its own restorative processes to counteract natural degradation. These processes are carried out by highly complex systems that detect and correct errors in DNA. A recent study on human cell lines reveals that the DNA damage response involves “a hierarchical organization of 605 proteins into 109 assemblies” [1]. All domains of life share some elements of the error-correction apparatus [2–4], with species that encounter exceptionally high levels of UV radiation or ionizing radiation having their own unique tools for the job [5].

Although these sophisticated countermeasures greatly reduce the frequency of chromosomal mutation, some mutations

inevitably escape them. For these mutations, natural selection provides the next line of defense, not at the level of the individual but at the level of the population. In particular, *purifying* selection (also referred to as *negative* selection) tends to reduce the frequency of genetically disadvantaged individuals in natural populations [6]. This is simply the flip side of positive selection: the comparative reproductive disadvantage of the impaired new variant is nothing other than the comparative reproductive advantage enjoyed by the wild type.

Purifying selection is quite efficient because the advantaged genotype is prevalent in the population when the mutation that produces the disadvantaged genotype first appears in a mutant individual. This contrasts with positive selection of a new variant, in which the advantaged genotype is enjoyed only by the

lone mutant individual. In effect, then, purifying selection is indistinguishable from the case in which positive selection has *nearly* completed fixation<sup>1</sup> of a recently produced advantaged genotype (fixation being complete when the prior genotype has been fully replaced by the new one).

However, despite its efficiency, purifying selection is an imperfect filter. A large majority of injuries to the information carried in genomic DNA are eliminated either at the level of the individual (by error-correction mechanisms) or at the level of the population (by purifying selection) before they become fixed, but the ongoing occurrence of these injuries inevitably means that some do become fixed.

The likelihood of this happening depends on the size of the population and the degree to which the mutation affects fitness. The latter is characterized by the selection coefficient  $s$ , which is the fractional change in fitness of the new mutant relative to the wild type ( $s$  is positive for beneficial mutations, negative for harmful ones). Although mutations that reduce fitness (and therefore have negative  $s$  values) are generally less apt to be passed on, their fixation is reliably disfavored only if the effective population size<sup>2</sup>  $N_e$  is greater than  $-1/s$  [7].<sup>3</sup> For example, a population of at least 1,000 individuals is needed if individuals with mutations that reduce fitness by  $1/1000^{\text{th}}$  are to be weeded out reliably. And if they are not reliably weeded out, then these mutations would be expected to accumulate indefinitely, which (in the absence of offsetting beneficial mutations) would result in long-term fitness decline.

This is different from what is commonly called *error catastrophe*, as originally described by Eigen and Schuster [8]. That well-studied phenomenon occurs when the average mutation rate substantially exceeds one new mutation in the coding portion of the genome per individual. In that case, the wild-type genotype is lost before selection has an opportunity to preserve it because nearly all individuals suffer new mutational damage. The accumulation we consider here is more akin to what is known as *Muller's ratchet*. As originally described [9], this ratchet effect results from an asymmetry between selection (which cannot improve fitness unless superior variants happen to exist) and genetic drift (which can always reduce fitness because inferior variants always exist). Put simply, selection is powerless to undo the small slips in fitness that drift inevitably produces.

Since Muller, many others have returned to the alarming possibility that selection may be unable to prevent fitness decline. In framing the neutral theory of molecular evolution, Kimura argued persuasively that natural selection plays a much smaller role in genomic evolution that was previously thought [10], and Ohto's refinement of Kimura's work (rebranding it the *nearly* neutral theory [11]) led to the expectation that slightly

harmful mutations would accumulate. The subtitle of a 1995 paper by Kondrashov—"Why have we not died 100 times over?"—raised the point as an important paradox [12], one that has been taken up by Lynch and coworkers (e.g., [13–15]) and Sanford and coworkers (e.g., [16–18]), among others.

On the one hand, then, so much has been said about this paradox that we might wonder what more can be said. On the other, the paradox remains as unresolved and as pressing as ever: How is genomic information preserved in the face of slightly harmful mutations, and if it is not preserved, then how are we still here?

Resolution has been slow at least in part because of the difficulty of making conclusive observations. Fitness, being statistical in nature, is notoriously hard to measure for living populations even in controlled laboratory contexts, to say nothing of natural contexts. Computer models offer an alternative that can be studied with statistical precision, but usually only because the distribution of mutation effects is a model input. The statistics that come out of these models are dictated by the statistics that went in.

Nelson and Sanford have commented on this limitation. The statistical parameters they used in one model run [18] allowed highly beneficial mutations to occur occasionally, which means they occur reliably if the run is continued long enough. When they do, they substantially increase the fitness of the population by offsetting the prior history of mildly harmful mutations. The authors concluded that this cost-offsetting is simplistic in that "it treats evolution merely as an accounting problem...as an exercise in fitness addition and subtraction, without any reference to the underlying genomic mechanisms or architecture" [18].

Our aim here is to examine genomic information loss in a model system that does have an underlying genetic mechanism. The previously described *Stylus* model is built on a real-world relationship between structures and functions—namely, the relationship between the Chinese written characters and their meanings [19]. In the Stylus world, gene-like sequences encode chains of connected vectors that, like polypeptide chains, are functional only if they have the right kind of sequence. Stylus genes work to the degree that their encoded vector chains—called *vector proteins*, by analogy—resemble one of the traditional Chinese characters. For example, a Stylus gene has the function of 段 (Unicode designation U+6BB5, meaning *section*, *piece*, or *division*) to the degree that the encoded vector protein resembles this character when rendered according to the model algorithm. The freely available<sup>4</sup> Stylus software [20] performs a geometric comparison<sup>5</sup> between the encoded vector drawing and the ideal form of a Chinese character to calculate a score ranging from zero to one. Although efforts were taken to make the Stylus scoring reflect actual legibility to a reasonable degree [19], the intent was for the Stylus world, once specified, to be its own thing with its own properties. That is, Stylus is not meant to emulate either molecular biology or human

<sup>1</sup> Fixation is the term for replacement of the wild-type genotype in a population with a new wild type. This can happen either by natural selection or by drift (i.e., by chance).

<sup>2</sup> The non-uniform distributions of resources and threats in real populations make them less able to carry genetic diversity than ideal (highly uniform) populations would be. In essence, the effective size of a real population is the size of an ideal population that carries the same genetic diversity.

<sup>3</sup> It is commonly stated that fixation of deleterious mutations is nearly impossible if  $N_e$  is two-fold larger ( $-2/s$ ; see, for example, Charlesworth [7]).

<sup>4</sup> <http://biologic.org/stylus>

<sup>5</sup> Full details of the Stylus scoring algorithm are published as a Supplement to the original Stylus paper [19].

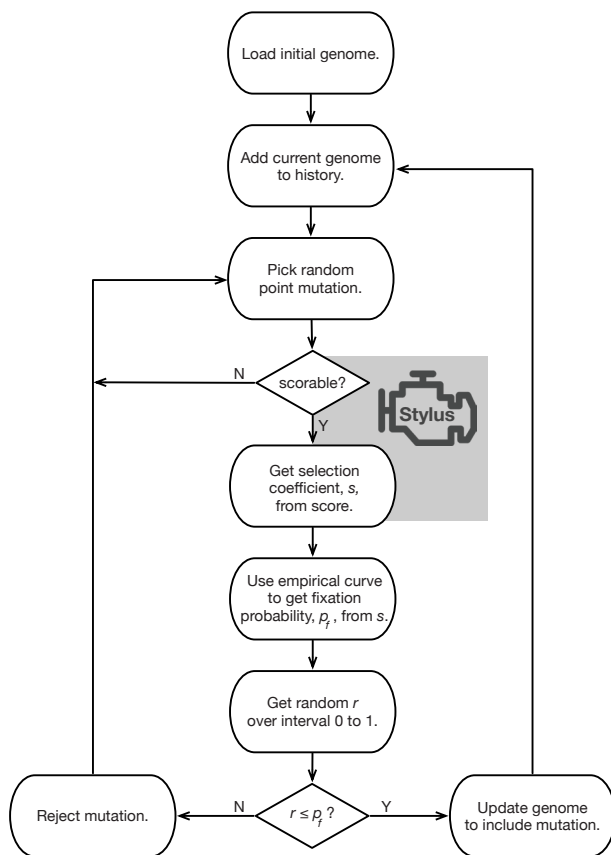
perception. Rather, it's meant to be a separate world with parallels to real-world biology that are rich enough to make for interesting and possibly informative comparisons.

A minimalistic Stylus genome has been described previously [21]. This 223-gene genome encodes a succinct text that explains in Chinese how Stylus genes are decoded. This is meant to be analogous to a minimal bacterial genome, in which the encoded proteins (components of the transcription, translation, and DNA replication complexes, etc.) “interpret” the genome by acting upon it appropriately.

In the present work, we track fitness in an ideal population of 1,000 ( $N = N_e = 1,000$ ) minimal Stylus organisms through as many generations as needed to get a clear picture of the long-term fate of the population.

## APPROACH

Prior implementations of Stylus have treated single genes in isolation, simulating evolutionary lineages of any duration from an initial given gene [22]. Any inferences about genomes had to be based on computational experiments performed separately on each gene in the genome. Because the current work examines genomes that evolve as wholes, our first objective was to extend the software to handle genomic evolution. The Python script we wrote for this handles long-term genomic evolution as shown schematically in Figure 1.



**Figure 1: Flowchart showing the process used to carry out long-term evolution.** doi:10.5048/BIO-C.2023.2.f1

The method is as follows. With the current genome<sup>6</sup> loaded and entered into the genome history, a single base substitution is chosen randomly, the likelihood of it falling in any particular gene being proportional to that gene's length. The previously described Stylus scoring engine<sup>7</sup> is then used to calculate the resulting score for the affected gene. In some cases, the chosen mutation is so disruptive that the gene cannot be scored (for example, if a new stop codon interrupts translation). Mutations of this kind are rejected (as they would be by purifying selection), and a new mutation is chosen. When the Stylus scoring engine can return a numerical score for the mutant gene, this is used to calculate the fitness of the mutant genome under consideration.

The original description of Stylus states that the fitness of a genome is calculated by “dividing the lowest proficiency of its necessary [gene] functions by the total cost of its genes” [19]. In other words, fitness is limited by the poorest performing gene in the genome. This is referred to as a *weakest-link* model of fitness, one of many models that could be used. Considering the extraordinarily complex and underdetermined connection between reproductive success and genotype in actual populations, all models require simplification. An advantage of this weakest-link approach (apart from simplicity) is that it suits an ideal conception of minimal genomes. Specifically, in a cell where the only processes occurring are those necessary for cell growth and division, the rates of these processes should be balanced according to need. Because all of them are necessary, any one of them would limit the rate of reproduction if it were to become disproportionately slow. Conversely, there would be no benefit in any one of the processes becoming disproportionately fast. The weakest-link model represents this situation in that any gene will become rate limiting if it underperforms the others.

Another aspect of this study that suits minimal genomes is that the current implementation of Stylus only scores a particular gene for its original function, specified by the Chinese character it encodes. If mutations cause this score to fall by distorting the encoded vector protein, it is conceivable that the modified protein could come to resemble a different Chinese character. But given that the original function is necessary (which follows from the definition of a minimal genome), any such fortuitous appearance of a different function would be unhelpful in that it would be accompanied by loss of a necessary function. For the purposes of the present study, then, the fitness of a genome is correctly calculated from the scores of its genes with respect to their original (necessary) functions.

From the fitness value for a new mutant genome, we calculate the selection coefficient  $s$  for the mutant genotype. As stated above, the selection coefficient is the fractional change in fitness caused by the mutation:

$$s = (f_m - f_{wt}) / f_{wt}, \quad (1)$$

where  $f_m$  and  $f_{wt}$  are the fitness values for the mutant and wild-type genomes, respectively. The selection coefficient will be

<sup>6</sup> We use the term *wild-type* to refer to whatever genome is considered to be the current one. The wild-type genome therefore changes throughout long-term evolution.

<sup>7</sup> See: doi:10.1371/journal.pone.0002246.s001

negative for mutations that reduce fitness, positive for mutations that increase fitness, and zero for mutations that leave the fitness unchanged.

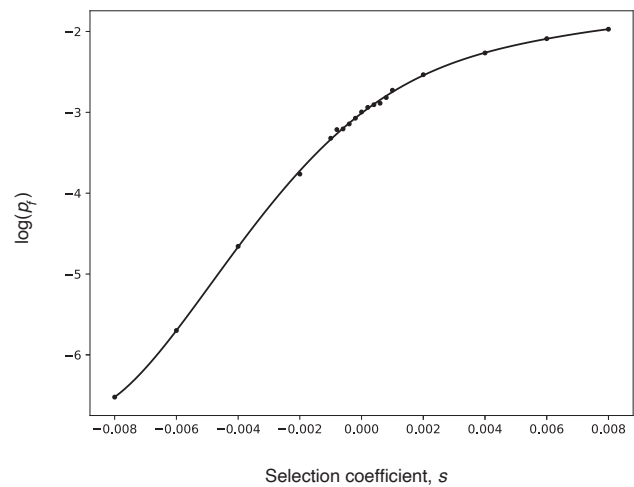
Given our chosen population size of 1,000, it follows that selection will not reliably eliminate mutations with  $s$  values between  $-0.001$  and  $0$  (as discussed above). But this is only a crude rule of thumb. Because fixation probabilities are continuous, there is no sharp cutoff at  $s = -0.001$ . Realistic modeling of long-term evolution therefore requires us to characterize the dependence of the probability of fixation  $p_f$  on  $s$  in the vicinity of  $s = 0$ .

This we accomplished by performing Monte Carlo simulations on a population of 1,000. Each simulation started with 999 individuals having identical genotypes (and therefore identical fitness values) and one individual with a slight advantage or disadvantage. In each subsequent generation, 500 individuals were chosen probabilistically for duplication (the equivalent of asexual reproduction), the simulation ending when all of these chosen individuals had the same genotype. Most often this happened by losing the genotype that was represented by the single individual at the outset. Occasionally, however, that introduced genotype ended up displacing the one that started with the huge numerical advantage. Moreover, this wholesale change of genotype happened with measurable frequency even when the introduced genotype was disadvantaged (i.e., had a negative  $s$  value, as given by Equation 1).

Figure 2 shows the result of millions of simulations of this kind performed with nineteen values of  $s$  for the introduced genotype. We chose the range of  $s$  to cover what we anticipated to be the significant values in the long-term evolution experiment. At the low end, when  $s = -0.008$ , we found the probability of fixation to be well below one in a million, which can safely be ignored. We anticipated that the high end ( $s > 0.008$ ) can be ignored for a different reason, namely that mutations that improve fitness by that amount are vanishingly rare.

For the sake of efficiency, values of  $p_f$  were multiplied by a factor of 85 during the long-term evolution experiment. This minimized the wasted cycles caused by failing the  $r \leq p_f$  test at the bottom of the flowchart in Figure 1 without altering the relative frequencies of fixation for mutations with different  $s$  values.

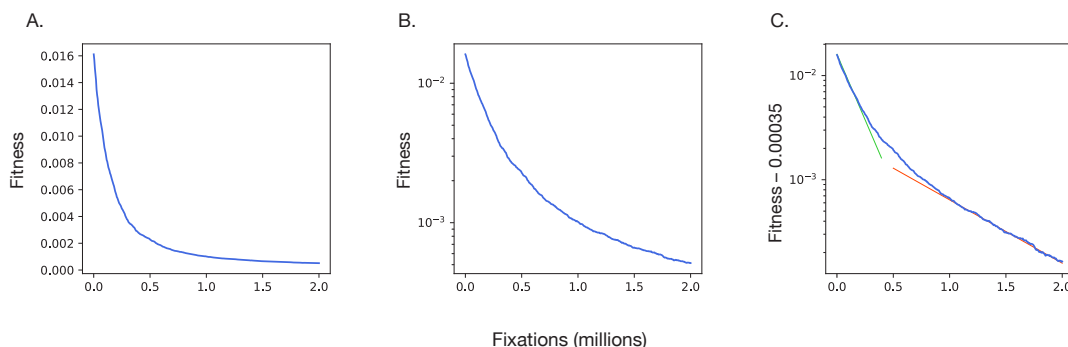
Because this is the first implementation of accurate genome-wide mutation and selection in the Stylus world, we took



**Figure 2: Monte Carlo assessment of the probability that a genotype with selection coefficient  $s$  will become fixed in a population of 1,000 individuals.** Simulations started with 999 wild-type individuals and one advantaged or disadvantaged individual (or wild-type equivalent in the case  $s = 0$ ). Each simulation was carried out until a single genotype remained. The number of simulations was adjusted in order to capture an adequate number of fixation events (e.g.,  $10^7$  simulations for  $s = -0.008$ ;  $5 \times 10^5$  simulations for  $s = 0.008$ ). The resulting data were fit to a fifth-degree polynomial using the polyfit method from the NumPy polynomial package (in Python terms: `numpy.polynomial.polynomial.polyfit`). [doi:10.5048/BIO-C.2023.2.f2](https://doi.org/10.5048/BIO-C.2023.2.f2)

advantage of the opportunity to evolve a more realistic genome to use as a starting point for long-term evolution than the genome described previously [21]. The method for making the first Stylus gene encoding any particular Chinese character starts with something very unlike natural genes—a large and highly repetitive gene produced by a gene-generation algorithm. More realistic genes are obtained from this artificial starting point by using Stylus to apply codon-deletion mutations (along with base substitutions) to reduce it to a more optimal size [21]. However, since this process was previously performed one gene at a time, there is no reason to think that the size reduction it causes matches what would occur if codon deletions were applied to the whole genome.

To remedy this, we applied a mixture of 30% codon-deletions and 70% base substitutions to the published genome [21] until one million fixation events had occurred. This was done in a



**Figure 3: Fitness trajectory over two million fixation events.** A) Linear plot. B) Semi-log plot. C) Semi-log plot of fitness minus  $3.5 \times 10^{-4}$ . The green line ( $y = 0.016e^{-5.8 \times 10^{-6}x}$ ) approximates the early decay; the red line ( $y = 0.0026e^{-1.4 \times 10^{-6}x}$ ) approximates the late decay. [doi:10.5048/BIO-C.2023.2.f3](https://doi.org/10.5048/BIO-C.2023.2.f3)

way that roughly approximates a population of  $10^8$  individuals. Specifically, instead of using the relationship between  $p_f$  and  $s$  shown in Figure 2, we took  $p_f$  to be 1 for  $s > -10^{-8}$  and 0 otherwise. The resulting genome, which we use here as the starting point for long-term evolution of a population of 1,000, turns out to be not very different in size from the published genome. The published genome has 67,410 coding bases, whereas the newly evolved one has 64,875 bases (a 3.8% reduction).

## RESULTS

As shown in Figure 3, a consistent downward trend in fitness is observed throughout the two-million-fixation-event history that was modeled. Although some fixation events increased fitness slightly, these were well outnumbered by those that slightly decreased fitness. The curve in Figure 3A is suggestive of exponential decay, but the semi-log plot shown in Figure 3B shows the curve is not a simple exponential decay (as this would produce a downward-sloping line). Rather, there is a transition in the vicinity of  $5 \times 10^5$  fixations from steep decay to less steep decay. It is evident from Figure 3A that the asymptote of the decay (i.e., the final fitness) is not zero. The linear decline after  $10^6$  fixations in the semi-log plot of Figure 3C was achieved by subtracting  $3.5 \times 10^{-4}$  from the fitness values. From this, we estimate that the final fitness would be about  $3.5 \times 10^{-4}$ , which is about 50-fold lower than the initial fitness of  $1.6 \times 10^{-2}$ .

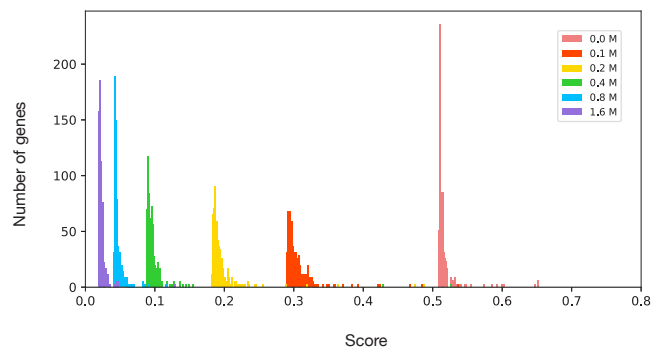
As stated above, the fitness of an individual in the Stylus world is calculated as the score of its least proficient gene divided by the total cost of its genome. This cost has two components: a cost of maintaining genetic instructions, set at  $2.38 \times 10^{-5}$  per DNA base, and a cost of making vector proteins from those instructions, set at  $9.52 \times 10^{-4}$  per unit length of vector (the three vector lengths being  $e^0$ ,  $e^{1/2}$ , and  $e^1$ ) [19]. Because mutations in this study are restricted to base substitutions, they have at most a slight effect on cost: the number of bases in the genome is unchanged throughout the long-term evolution, and the total number of vector units changes only by a small proportion with each vector substitution ( $\pm 0.005\%$ ).

The steady decline in fitness we observed is therefore caused not by cost escalation but by a steady decline in the lowest gene

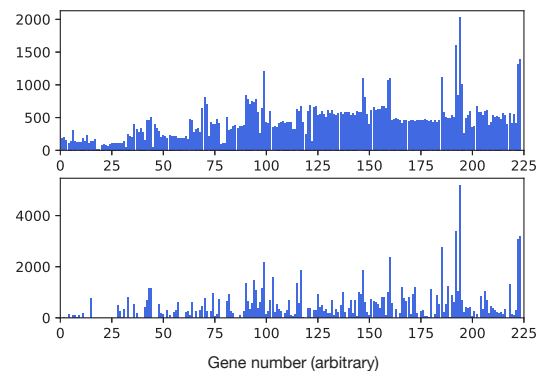
score. We should expect this to be accompanied by a steady decline in all gene scores because, while it is theoretically possible for low fitness to be caused by a single impaired gene, a population size of 1,000 is amply large to prevent fixation of mutations that single-handedly cause severe impairment. The severe impairment observed in this study must instead be caused by accumulation of a very large number of tiny injuries. This implies that the entire genome undergoes gradual degradation. The gene-score histograms shown in Figure 4 confirm this. All genes undergo functional degradation in unison.

This does not mean that all genes have equivalent roles in the process. Certain genes may carry mutations to fixation more often than others. We refer to these genes as *mutation carriers*. Similarly, certain genes (not necessarily the same ones) may more often be the poorest performers and therefore the determinants of fitness. We refer to these genes as *fitness limiters*. Figure 5 shows that there is indeed nonuniformity with respect to both of these roles. Over the first 100,000 fixations, all 223 genes served as the mutation carrier, though some served this role much more frequently than others (Figure 5A). On the other hand, 51 genes never served as the fitness limiter (Figure 5B).

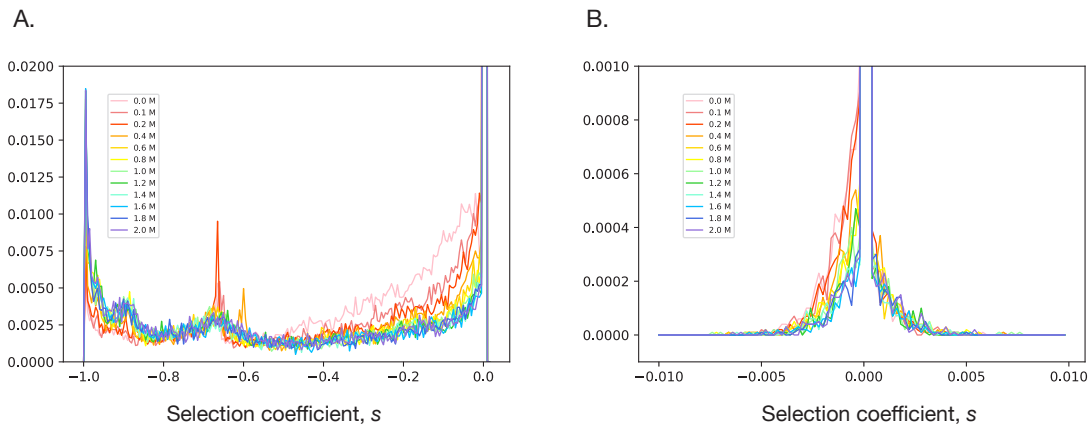
Although there does not appear to be a simple rule that determines which genes assume these roles most frequently, a pattern is discernable at the extreme ends of the complexity spectrum. One measure of the complexity of a vector protein is the number of strokes in the Chinese character it depicts. By this measure, genes 9E7C.01 and 9E7C.02 encode the most complex vector proteins in this vector proteome, both depicting a 24-stroke character (鹼) that means *alkaline*. At the other end of the spectrum, genes 4E00.01, 4E00.02 and 4E8C.01 encode the simplest vector proteins—ones that depict the 1-stroke character for the number 1 (一) and the 2-stroke character for the number 2 (二). The most complex genes (9E7C.01 and 9E7C.02, represented at the far right of Figure 5) are among the most frequent fitness limiters, whereas the simplest genes never limit the genomic fitness. A reasonable interpretation is that the most complex vector proteins are hardest to keep in good working order and are therefore frequently found to be the poorest performers, whereas the simplest vector proteins



**Figure 4:** Histograms showing the distribution of gene scores within the evolving genome at different times (measured by fixation events). The inset key gives the number of fixations (in millions) for each time slice. doi:10.5048/BIO-C.2023.2.f4



**Figure 5:** Number of times each gene served as the mutation carrier (top) or the fitness limiter (bottom) over the first 100,000 fixations. doi:10.5048/BIO-C.2023.2.f5



**Figure 6: Distributions of  $s$  values of incident mutations and fixed mutations for runs of 100,000 fixation events.** A) Frequency of incident mutations with selection coefficient  $s$ , calculated as counts in bins of size 0.005 divided by the total number of scorable incident mutations in 1000-fixation samplings that started with the genome at the evolutionary extent indicated in the inset key. B) Frequency of fixed mutations with selection coefficient  $s$ , calculated as counts in bins of size 0.0002 divided by 100,000. The large spikes at  $s = 0$  are caused by silent mutations (changes to a gene sequence that don't change the vector sequence). These are clipped in order to visualize the non-silent mutations. [doi:10.5048/BIO-C.2023.2.f6](https://doi.org/10.5048/BIO-C.2023.2.f6)

readily retain better-than-average scores because they have modest information requirements.

It should be pointed out that selection greatly slows the genomic decline, even though it is unable to prevent it. The dramatic difference between the distribution of  $s$  values for incident mutations (Figure 6A) and the distribution of  $s$  values for fixed mutations (Figure 6B) is entirely due to selection. Although the shapes of these distributions change somewhat over the course of the experiment, the strong drop in the likelihood of fixation caused by lower  $s$  values is consistent throughout.

## DISCUSSION

### Interpretation of model-population behavior

Figure 7 shows how fitness change in our model population results from an imbalance between fitness-increasing and fitness-decreasing fixation events. Solid horizontal lines represent the scores of the 223 genomic genes (most omitted for simplicity) on a vertical scale. If we ignore the tiny effects on fitness caused by fixation events that do not change the worst gene score (as discussed above, these can change the total cost slightly), then any trends in fitness are caused by trends in the fixation events that do change the worst score. These would typically either move a gene into that bottom position from a higher position or move the bottom gene to a higher position, the latter thereby causing what was previously the next-to-worst gene to become the worst gene.<sup>8</sup> The first scenario, which we call *demotion*, causes fitness decline. The second, which we call *promotion*, causes fitness improvement.

The dashed green line in Figure 7 represents the new lowest score after a demotion event. This event reduces the lowest score by  $\Delta^-$ , reducing the fitness by the same proportion. The rate at which these demotion events contribute to reduction of the worst score is the product of the rate of the events

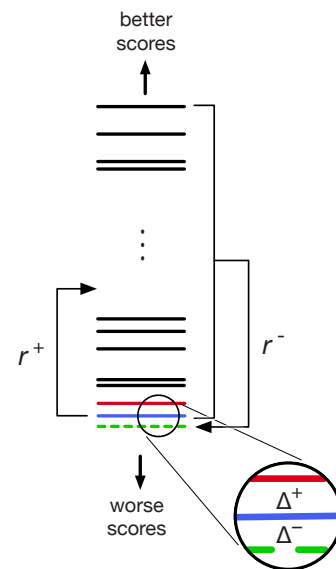
<sup>8</sup> Although it is possible for a mutation in the worst gene to leave it in that position with a slightly altered score, this will be a rare occurrence. Moreover, this does not change our analysis.

themselves ( $r^-$ ) and the size of their effect (the mean value of  $\Delta^-$ ). Likewise, the rate at which promotion events contribute to elevating the worst score is  $r^+$  times the mean of  $\Delta^+$ . The worst score  $S_-$  therefore changes at a rate given by the following differential equation:

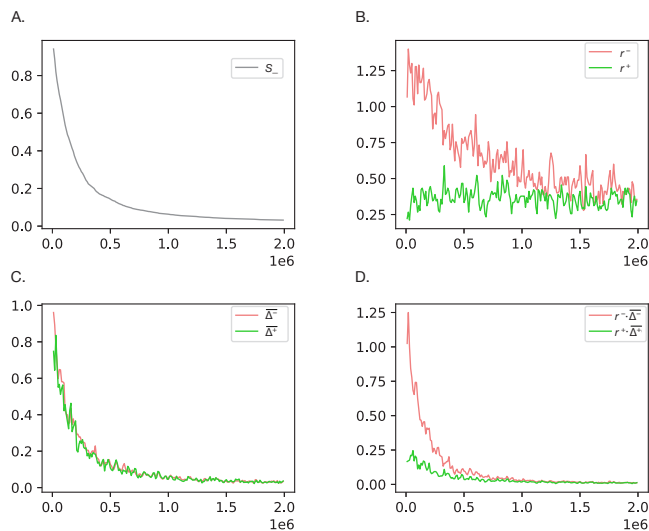
$$\frac{dS_-}{dt} = r^+ \cdot \overline{\Delta^+} - r^- \cdot \overline{\Delta^-} \tag{2}$$

where overbars signify mean values.

Evidently, the second term on the right side of Equation 2 overpowers the first term initially, causing the rate of change to



**Figure 7: Schematic illustration of the competing processes during long-term evolution.** The vertical positions of the horizontal lines represent the scores of the genes in the genome. The blue line represents the current worst-scoring gene. The red line represents the next worst gene, which becomes the worst if a mutation shifts the blue line upward. The dashed green line represents a new worst-scoring gene, which would be caused by a mutation impairing a gene that was higher in the stack. [doi:10.5048/BIO-C.2023.2.f7](https://doi.org/10.5048/BIO-C.2023.2.f7)



**Figure 8: Rates and effects of beneficial and non-beneficial fixation events over long-term evolution.** Horizontal scales represent time as measured in fixation events. Plots are of normalized parameters as follows: A) worst score averaged over 16,000 consecutive fixations and divided by initial worst score; B)  $r^-$  and  $r^+$  (counts over 16,000 fixations divided by 16,000) divided by the initial value of  $r^-$ ; C)  $\Delta^-$  and  $\Delta^+$  averaged over 16,000 consecutive fixations and divided by initial value of  $\Delta^-$ ; D)  $r^- \cdot \Delta^-$  and  $r^+ \cdot \Delta^+$  divided by the initial value of  $r^- \cdot \Delta^-$ .

[doi:10.5048/BIO-C.2023.2.f8](https://doi.org/10.5048/BIO-C.2023.2.f8)

be negative. As evolution proceeds, the two terms come increasingly close to balancing, causing  $S_-$  to stabilize at a much lower value than the initial one. Figure 8 shows how this unfolds in detail. As seen in Figure 8A,  $S_-$  falls in the same way that fitness does in Figure 3A, which is expected in light of the fact that the drop in  $S_-$  causes the drop in fitness. Figure 8B shows that the rate at which mutations that decrease  $S_-$  are fixed drops substantially over the course of the experiment. By contrast, the rate at which mutations that increase  $S_-$  are fixed remains stable. A likely explanation of this is that as  $S_-$  decreases, the slim score range that needs to be hit for a mutation to drop a new gene into the lowest scoring position becomes ever slimmer, making these fixation events ever rarer. Mutations that lift the lowest scoring gene out of that position do not face the same narrowing likelihood.

However, as Figure 8C shows, the magnitude of the changes to  $S_-$  (whether increases or decreases) falls dramatically, following the same trajectory as the fall in  $S_-$  itself. This fits with the fact that selection sifts fitness changes not with respect to any absolute scale, but rather in proportion to the current fitness. A 1% fitness drop is comparably hard to fix at any time in the experiment, but the absolute magnitude of that drop in terms of the change in  $S_-$  is tenfold smaller once  $S_-$  has dropped to a tenth of its initial value.

Figure 8D shows that the disparity between the two terms on the right-hand side of Equation 2 does indeed collapse as the experiment progresses.

Readers familiar with computational genetic algorithms, which are useful for various optimization problems, will have noticed that the de-optimization we observe in this work is the

reverse of what happens when genetic algorithms are successfully applied. Without claiming expertise in those computational methods, we think two differences may explain the different outcomes. The first is that our algorithm is more biology-like than most of the ones used to solve practical optimization problems. In cases where people actually have a problem to solve, they are free to augment aspects of biology with any non-biological features that might help. Related to this, the second difference is that we started with a highly refined solution (the original genome), whereas genetic algorithms tend to be applied in order to refine very crude initial solutions. In light of these differences, it makes sense that we see a downward trend, whereas practically applied genetic algorithms often see an upward trend.

### Implications for real populations

Although some important aspects of bacterial populations have been captured in our model, others have not. The value of this work is that it enables informed comparison of the fully known model system to the less well understood real systems, and this is as true when the behaviors of the two are dissimilar as it is when they are similar.

Kondrashov's paradox, discussed in the introduction, raises a question about survival: How do species avoid extinction in the face of mildly harmful mutations? But the results presented here, together with previous studies of the kind we have referred to, also raise a question about *origins*. The evolutionary understanding of life presupposes that mutation and selection operating on small offshoots of past populations drove the production of more complex and more highly fit forms. The paths thought to have been traversed in this way are characterized by abundant radical innovation—bacteria becoming sponges, among countless other things, and sponges (or something like them) becoming humans, among countless other things. Yet, when we apply mutation and selection to a population of a thousand individuals carrying a small genome, we find that selection is unable even to retain those simple model organisms, much less generate more sophisticated ones.

Moreover, the problem is independent of the rate at which mutations occur. The *timescale* of genomic decline is proportional to the mutation rate, but the mere fact of decline is a simple consequence of two things: 1) the imbalance in incident mutations—slightly harmful ones well outnumbering helpful ones—and 2) the inability of selection to eliminate slightly harmful mutations in populations of modest size.

How would real organisms avoid the same pitfall? If the initial imbalance on the right-hand side of Equation 2 results from there being fewer opportunities for the current worst gene to improve than for the 222 other genes (in our model) to become slightly worse than the current worst, surely that imbalance is even more extreme when the number of essential genes is much larger, as is the case for all eukaryotic species. Among the simplest of these is the well-studied yeast species *Saccharomyces cerevisiae*, which is estimated to have 1,107 essential genes [23]. We do not expect the weakest-link fitness model to apply to complex genomes, but if it applies even to a core set of essential

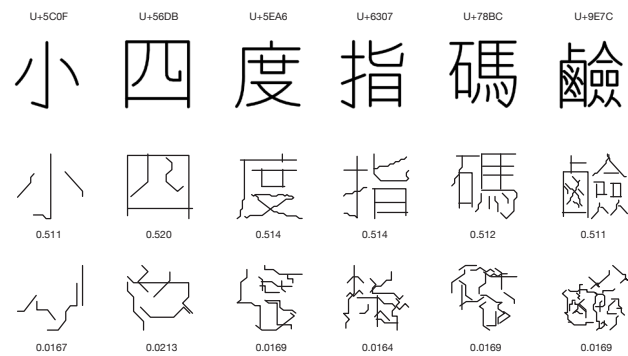
genes and this core is as large or larger than it is for *S. cerevisiae*, then loss of core information seems likely in small populations (in agreement with previous model studies [13–14]).

Could the initial state of the genome in this study have been unrealistically optimal? If so, then what we have described as degradation might be better described as relaxation to a more realistic state. Figure 8B shows the definitive way to answer this. If  $r^-$  and  $r^+$  are balanced in real populations, then the initial state of our model population was unrealistic. On the other hand, if real populations show the kind of imbalance seen at the left side of Figure 8B, then the model is realistic in this regard.

These are not easy measurements to make on real populations, however. As a more accessible alternative, Figure 9 shows that the initial genome in our model was not unrealistically optimal in terms of legibility. Rather, the final genome is unrealistically *suboptimal*, as seen by the illegibility of the Chinese characters it encodes. However uncertain the parallel to biological proteins may be, it is clear that selection was unable to preserve adequate functioning of the vector proteins encoded by our model genome in this legibility-based world.

Population size is another important consideration. If natural populations never have to persist for extended periods with something like 1,000 (or fewer) individuals, then the challenge we put to our model population was unrealistic. The concept of *effective* population size  $N_e$  (see “Introduction”) comes into play here. Because natural habitats are characterized by highly nonuniform distributions of resources and threats, the chances of survival and reproduction are highly nonuniform for reasons that have nothing to do with genetic fitness. And this is typical across the kingdoms of life. A small fraction of the members of a wild population happen to be in the right places at the right times to avoid the threats to survival and reproduction faced by many of their peers. Consequently, in terms of their ability to carry genetic diversity, real populations behave as though they were much smaller than they are. This is reflected in effective sizes being much smaller than actual (or census) sizes. Estimates of the ratio of the two,  $N_e/N$ , vary considerably. High-end estimates of  $N_e/N$  for wildlife populations are roughly 0.1 [24,25]. Low-end estimates are much lower:  $<10^{-4}$  for *Drosophila melanogaster* [26] and  $<10^{-10}$  for *Escherichia coli* [27].

It seems, then, that many wild populations may face a similar threat to genomic stability to that faced by our model population. This would occur when census populations are restricted for prolonged periods—bacteria to, perhaps, billions or trillions; insects to (roughly) tens or hundreds of millions; and larger animals to tens of thousands. With respect to the last of these, the International Union for Conservation of Nature maintains a Red List with over a 100,000 species categorized as *vulnerable*, *endangered*, or *critically endangered*, based primarily



**Figure 9: Sample of gene products before and after long-term evolution.** Top row: ideal forms of five characters. Middle row: vector proteins encoded by genes in the initial genome, with scores. Bottom row: vector proteins encoded by genes in the final genome, with scores. doi:10.5048/BIO-C.2023.2.f9

on population size or decline. The criterion<sup>9</sup> for inclusion in the *vulnerable* category (the least threatened of the three categories) is an adult population below 10,000, which corresponds roughly to  $N_e < 1,000$ .

Interestingly, many of these species persist in that category for extended periods without extinction, or recover by growing in number. The fact that all of them have much larger and more complex genomes than our tiny model genome only deepens the mystery.

The mystery also deepens if we turn from extinctions to origins because tiny initial populations would be the rule then, not the exception. Indeed, when all the factors that make naturalistic biological innovation improbable are brought together [28], the reasonable conclusion is that it should not happen at all. That is, the combinations of genetic circumstances required to account for much of what needs to be accounted for are so unlikely that we cannot expect them ever to occur in one individual. And if we suppose that they did, then this genetic oddity would have struggled to produce descendants that carried the new genes in homozygous form. In that context, a population of 1,000 pure-breeding individuals sharing the new innovation is a distant prospect, made all the more so by the finding that this hypothetical tiny population would be suffering genetic loss.

Considering the difficulty of direct confirmation, we judge this problem of genomic degradation to be less certain than the evolutionary problems that have been described before [28]. It does, however, constitute a paradox that—because of what is at stake—should be given careful attention.

<sup>9</sup> <https://www.iucnredlist.org/resources/summary-sheet>

1. Kratz A, Kim M, Kelly MR, Zheng F, Koczor CA, et al. (2023) A multi-scale map of protein assemblies in the DNA damage response. *Cell Systems* 14: 1–17. doi:10.1016/j.cels.2023.04.007
2. Groothuizen FS, Sixma TK (2016) The conserved molecular machinery in DNA mismatch repair enzyme structures. *DNA Repair* 38: 14–23. doi:10.1016/j.dnarep.2015.11.012
3. White MF, Allers T (2018) DNA repair in the archaea—an emerging picture. *FEMS Microbiol Rev* 42(4): 514–526. doi:10.1093/femsre/fuy020
4. Liu D, Keijzers G, Rasmussen LJ (2017) DNA mismatch repair and its many roles in eukaryotic cells. *Mutation Research* 773: 174–187. doi:10.1016/j.mrrev.2017.07.001
5. Ishino Y, Narumi I (2015) DNA repair in hyperthermophilic and hyperthermostable microorganisms. *Curr Opin Microbiol* 25: 103–112. doi:10.1016/j.mib.2015.05.010
6. Cvijović I, Good BH, Desai MM (2018) The effect of strong purifying selection on genetic diversity. *Genetics* 209: 1235–1278. doi:10.1534/genetics.118.301058
7. Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10: 195–205. doi:10.1038/nrg2526
8. Eigen M, Schuster P (1977) A principle of natural self-organization. *Naturwissenschaften* 64: 541–565. doi:10.1007/BF00450633
9. Muller HJ (1964) The relation of recombination to mutational advance. *Mutation Research* 1: 2–9. doi:10.1016/0027-5107(64)90047-8
10. Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press (Cambridge).
11. Ohta T (1992) The nearly neutral theory of molecular evolution. *Annu Rev Ecol* 23: 263–286. doi:10.1146/annurev.es.23.110192.001403
12. Kondrashov AS (1995) Contamination of the genome by very slightly deleterious mutations: Why have we not died 100 times over? *J Theor Biol* 175(4): 583–594. doi:10.1006/jtbi.1995.0167
13. Lynch M, Conery J, Bürger R (1995) Mutation accumulation and the extinction of small populations. *Am. Naturalist* 146(4): 489–518. doi:10.1086/285812
14. Lynch M, Conery J, Bürger R (1995) Mutational meltdowns in sexual populations. *Evol* 49(6): 1067–1080. doi:10.1111/j.1558-5646.1995.tb04434.x
15. Lynch M, Blanchard J, Houle D, Kibota T, Schultz S, et al. Perspective: Spontaneous deleterious mutation. *Evol* 53(3): 645–663. doi:10.1111/j.1558-5646.1999.tb05361.x
16. Nelson CW, Sanford JC (2011) The effects of low-impact mutations in digital organisms. *Theor Biol Med Modelling* 8:9. doi:10.1186/1742-4682-8-9
17. Gibson P, Baumgardner JR, Brewer WH, Sanford JC (2013) Can purifying natural selection preserve biological information? In: Marks II RJ, Behe MJ, Dembski WA, Gordon BL, Sanford JC, eds. *Biological Information—New Perspectives*. World Scientific (London) pp 232–263. doi:10.1142/9789814508728\_0010
18. Nelson CW, Sanford JC (2013) Computational evolution experiments reveal a net loss of information despite selection. In: Marks II RJ, Behe MJ, Dembski WA, Gordon BL, Sanford JC, eds. *Biological Information—New Perspectives*. World Scientific (London) pp 338–368. doi:10.1142/9789814508728\_0014
19. Axe DD, Dixon BW, Lu P (2008) *Stylus*: A system for evolutionary experimentation based on a protein/proteome model with non-arbitrary functional constraints. *PLoS ONE* 3(6): e2246. doi:10.1371/journal.pone.0002246
20. Axe DD, Ewert W (2016) *Stylus* experiments made easy—A free app for personal computers. *BIO-Complexity* 2016 (1): 1–4. doi:10.5048/BIO-C.2016.1
21. Axe DD, Lu P, Flatau S (2011) A *Stylus*-generated artificial genome with analogy to minimal bacterial genomes. *BIO-Complexity* 2011(3): 1–15. doi:10.5048/BIO-C.2011.3
22. Axe DD, Gauger AK (2015) Model and laboratory demonstrations that evolutionary optimization works well only if preceded by invention—Selection itself is not inventive. *BIO-Complexity* 2015 (2): 1–13. doi:10.5048/BIO-C.2015.2
23. Ben-Aroya S, Coombes C, Kwok T, O'Donnell KA, Boeke JD, et al. (2008) Toward a comprehensive temperature-sensitive mutant repository of the essential genes of *Saccharomyces cerevisiae*. *Mol Cell* 30(2): 248–258. doi:10.1016/j.molcel.2008.02.021
24. Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genet Res Camb* 66: 95–107. doi:10.1017/S0016672300034455
25. Palstra FP, Ruzzante DE (2008) Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol Ecol* 17(15): 3428–3447. doi:10.1111/j.1365-294X.2008.03842.x
26. Shapiro JA, Huang W, Zhang C, Wu CI (2007) Adaptive genic evolution in the *Drosophila* genomes. *Proc Natl Acad Sci USA* 104(7): 2271–2276. doi:10.1073/pnas.0610385104
27. Charlesworth J, Eyre-Walker A (2006) The rate of adaptive evolution in enteric bacteria. *Mol Biol Evol* 23: 1348–1356. doi:10.1093/molbev/msk025
28. Axe DD, Gauger AK (2013) Explaining metabolic innovation: Neo-Darwinism versus design. In: Marks II RJ, Behe MJ, Dembski WA, Gordon BL, Sanford JC, eds. *Biological Information—New Perspectives*. World Scientific (London) pp 489–507. doi:10.1142/9789814508728\_0022