Biology Direct

Research

The look-ahead effect of phenotypic mutations

Dion J Whitehead*[†1], Claus O Wilke*[†2], David Vernazobres[1] and Erich Bornberg-Bauer[1]

Address: [1]Institute for Evolution and Biodiversity, The Westphalian Wilhelms University of Muenster, 48149 Muenster, Germany and [2]Center for Computational Biology and Bioinformatics and Institute for Cell and Molecular Biology, Section of Integrative Biology, University of Texas, Austin, Texas 78712, USA

Email: Dion J Whitehead* - dion@uni-muenster.de; Claus O Wilke* - cwilke@mail.utexas.edu; David Vernazobres - dv@uni-muenster.de; Erich Bornberg-Bauer - ebb@uni-muenster.de

* Corresponding authors   †Equal contributors

Abstract

Background: The evolution of complex molecular traits such as disulphide bridges often requires multiple mutations. The intermediate steps in such evolutionary trajectories are likely to be selectively neutral or deleterious. Therefore, large populations and long times may be required to evolve such traits.

Results: We propose that errors in transcription and translation may allow selection for the intermediate mutations, if the final trait provides a large enough selective advantage. We test this hypothesis using a population based model of protein evolution.

Conclusion: If an individual acquires one of two mutations needed for a novel trait, the second mutation can be introduced into the phenotype due to transcription and translation errors. If the novel trait is advantageous enough, the allele with only one mutation will spread through the population, even though the gene sequence does not yet code for the complettrait. Thus, errors allow protein sequences to "look-ahead" for a more direct path to a complex trait.

Reviewers: This article was reviewed by Eugene Koonin, Subhajyoti De (nomimated by Madan Babu), and David Krakauer.

1 Introduction

According to a central principle of molecular evolution, the likelihood that a given mutation occurs is independent of the mutation's phenotypic consequences. Organisms cannot choose specific mutations. This tenet was challenged by [1], who observed that under a certain selective pressure, *E. coli* cells appeared to acquire an excess of beneficial mutations. The idea that cells can somehow 'direct' evolution was thought provoking, and stimulated many investigations (for reviews see [2-6]). While the notion that cells can directly decide in which genomic regions to increase their mutation rate has been mostly abandoned [4,7], the original observations by [1] have been corroborated (see above reviews).

If mutations arise independently of their phenotypic consequences, then how can adaptations occur that require multiple amino acid mutations and for which the intermediate stages are either selectively neutral or disadvantageous? Large populations can climb multiple fitness peaks, even with disadvantageous intermediate alleles [8,9]. Although no new mechanisms are therefore

required to explain the evolution of complex proteins [10], we propose that errors in transcription and translation (*phenotypic* mutations) allow the selection of the intermediate mutations of a multiple-mutation requiring trait, and can thus speed up the evolution of complex traits.

Studies on the phenotypic mutation rate indicate that it is orders of magnitude larger than the genotypic mutation rate [11,12]: the misreading error rate during protein synthesis is estimated to be between 10^{-3} to 10^{-4} misreadings per codon [13], compared with a genotypic mutation rate of between $\sim 10^{-7}$ to 10^{-11} [14]. Consequently, for a protein of 300 residues, on average more than 1 in 10 copies of the protein will contain a mutation. Using mutation rates derived from the literature and conservative biological assumptions, we show via mathematical modeling and simulations that phenotypic mutations allow evolution to select for neutral intermediate alleles of a multi-mutation trait, actually selecting for proteins whose exact DNA sequence is not in the organism under selection. Evolution is then able to look ahead for evolutionary jackpots in sequence space.

Our theory is based on the following hypothetical scenario. A protein can increase the fitness of an individual if it evolved a specific trait. This trait requires two mutations, for example a disulphide bridge between two cysteine residues. A modification of only two residues can result in large structural changes [15]. Having only one of the required mutations is either selectively neutral or deleterious, however when an individual has only one mutation, small amounts of the protein with both mutations will be produced due to phenotypic mutations. If the presence of both mutations at low concentrations provides even a small fitness improvement then the allele with one mutation will spread though the population. As the frequency of the intermediate allele increases, there is a greater probability that if the second mutation occurs, it will be in the presence of the first mutation, and thus provide the full fitness benefit.

Our hypothesis is similar to an effect proposed in 1896 by J.M. Baldwin [16], known now as the "Baldwin Effect" or "Organic Selection". The core idea is that the *probability* of a trait occuring can be selected for, not just the trait itself. If the phenotypic plasticity of an organism allows it to learn a trait, then during the course of evolution, the organism's descendants may get better at aquiring the trait, and may ultimately acquire genes that code for the trait directly. Using a genetic-algorithm based model, Hinton and Nowlan [17] tested this idea, and concluded that organisms with phenotypic plasticity evolved faster, even though the learned traits did not become hereditary in their study. Subsequent studies futher explored the Bald-

win effect in different fitness landscapes in the context of machine learning (see [18] for a review). This work clarified two aspects of the Baldwin effect: that lifetime learning can accelerate evolution in certain contexts, and that this learning usually comes with a cost. Organisms that have to acquire a beneficial trait by learning will generally have a selective disadvantage over organisms that genetically encode the trait. That evolution can select for the ability to learn has been demonstrated experimentally in fruit flies [19]. Our model differs in a subtle but important way from the Baldwin effect. The Baldwin effect describes learning from the perspective of the individual, meaning that an organism starts its life without the trait, but then later has a chance of acquiring the trait. In our model, phenotypic mutations occur at a given rate, although some individuals are more predisposed to the highly beneficial phenotypic mutations than others. The traits are not learned, because individuals that have the neutral, intermediate mutation will always express the beneficial phenotypic mutation at a low rate. The organisms in our model are not exactly phenotypically variable, rather phenotypically *diffuse*. They do not learn, but rather possess a small part of the many phenotypes close to their genotype. We call this effect the "look-ahead effect" as opposed to the Baldwin effect to highlight that no learning takes place.

Our work is more closely related to works on phenotypic plasticity [20] and random or noisy phenotypes [21]. The aim of this article is to derive explicit analytic expressions for the fixation process of genes whose fitness is modulated by phenotypic mutations, and to show that adaptive phenotypic mutations can undergo positive selection under biologically plausible conditions.

2 Model assumptions

We model the scenario of a protein evolving a trait that requires two mutations. The model is based on a population-genetics framework where a single gene can evolve into different alleles. We do not consider duplication and divergence of genes. In addition, the process described here will likely only occur for proteins with sufficiently long half-lives, as the protein must persist for some time to exert a phenotypic effect. As we model only a single gene, we expect our results to be more relevant for single-celled organisms and viruses than for multicellular organisms, which tend to have larger genomes and smaller effective population sizes than microorganisms.

The model consists of the evolution of three non-recombining haploid genotypes, where each genotype contains one of the three alleles shown in Figure 1. The three different alleles are named according to number of relevant mutations, corresponding to zero mutations (allele 0), a single mutation (allele 1), and both mutations (allele 2)
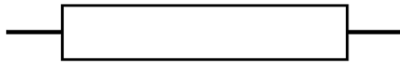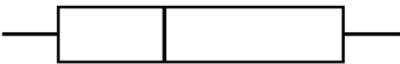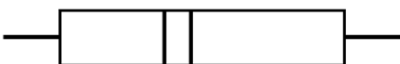
Figure 1
The three alleles. The three alleles (or genotypes). The vertical lines in the genes indicate the number of key mutations required for the novel two-residue function. The fitness of the allele 1 increases if phenotypic mutations are taken into consideration.

required for the adaptive feature. Having both mutations of the adaptive feature provides a selective advantage *s*. We assume that the intermediate allele (allele 1) is selectively neutral *if transcribed and translated without error*. We specifically take into consideration errors in transcription and translation, that is, *phenotypic mutations*.

In the model, the population initially consists of one individual carrying allele 1 and *N* - 1 individuals carrying allele 0. So long as allele 1 is present, allele 2 can be generated by mutations. The population evolves for a fixed time period, during which allele 2 can be generated by mutation and go to fixation. In each generation, selection increases the frequency of the alleles according to their corresponding fitness values. Allele 0 has a fitness of 1. Allele 2 has a fitness of 1 + *s*, where *s* is the selection coefficient provided by the adaptive feature. The fitness of allele 1, the intermediate allele with only a single mutation, depends on the phenotypic error rate. Most phenotypic errors will be neutral or deleterious, however some will be beneficial. For simplicity, we assume that the length of the protein and the expression level are both constant. In addition, we do not explicitly model deleterious phenotypic mutations. As long as the spectrum of deleterious phenotypic mutations does not change substantially among alleles 0, 1, and 2, we can treat its fitness effect as a common factor which we divide out of all fitness values. This assumption becomes invalid if, for example, phenotypic mutations for allele 1 are significantly more deleterious than those for either allele 0 or allele 2. What happens when we relax these assumptions will be the subject of future work.

If there are no phenotypic mutations, allele 1 has the same fitness as allele 0. However, if phenotypic mutations occur, allele 1 can produce a small number of allele 2 pro-

teins due to phenotypic errors. The fitness of allele 1 is therefore dependent on the number of such errors. We assume, for the sake of simplicity, that fitness is a linear sum of individual proteins, meaning that if some phenotypic variants of a protein have a higher fitness, then the overall fitness of that allele is proportionally increased.

We let *r* be the number of residues that can potentially complement the first mutation to provide the full two-residue adaptive feature. These *r* residues represent, e.g., the sites at which the second cysteine of a cysteine bridge could arise; other similar two-residue mutations that significantly improve functionality can be proposed. Two residues that comprise an adaptive trait are likely to co-evolve, because if a mutation occurs in one of the residues, selection strongly favors a compensatory mutation in the other. Based on a large data set, [22] found that co-evolving residues are spatially near. Co-evolving residues were, on average, 98.6 amino-acids apart along the sequence, but had a mean spatial distance of 6.9 Å. This spatial distance can be compared to the width of the van der Waals volume of an amino-acid (5–6 Å), showing that most co-evolving residues are effectively in contact proximity. Therefore, *r* is mostly independent of the size of the protein, as long as the protein is of sufficient length. [23] calculated the mean contact density (the mean number of residues in contact with a given residue) for 194 yeast proteins, and found that most residues have a mean contact density of seven to eight residues. In this work we use *r* = 8. Given *r* possible positions for the second residue, and assuming that each position requires a specific residue, the fraction of proteins of allele 1 containing the second (now highly beneficial) mutation is $\beta = \frac{r}{19}\lambda$, where λ is

the per codon non-synonymous phenotypic mutation rate. In this model, we use $\lambda = 4.5 \times 10^{-4}$ mistranslations per codon [24,25]. The fraction β of allele-2 proteins contribute to the fitness, giving allele 1 a fitness of $1 + s\beta$.

When considering genetic (i.e. inherited) mutations, for simplicity we neglect back mutations (e.g. from allele 1 to allele 0), and assume there are no recurrent mutations of allele 1 from allele 0 (the model starts with a single copy of allele 1). Allele 2 arises via a mutation from allele 1. We ignore the possibility of a double mutation directly from allele 0 to allele 2, as this probability is extremely small in the parameter range we are interested in. The genetic mutation rate for allele 1 mutating into allele 2 is derived as follows: For microbes, the rate of mutations per nucleotide per generation is between $\sim 10^{-7}$ to 10^{-11} [14]. Here we use 10^{-8} as the non-synonymous mutation rate per codon per generation. The resulting mutation rate for changing allele 1 into allele 2 is $U = \frac{r}{19} 10^{-8} = \frac{8}{19} 10^{-8}$.

Genes can also acquire null mutations, rendering the gene non-functional and therefore eliminating the organism. The null mutation rate for protein-encoding genes is on the order of 10^{-6} per generation [14]. However, this rate will depend on the length (L) of the protein. Assuming an average protein length of 300 residues, the per-residue null mutation rate is given by $10^{-6}/300 = \sim 3.3^{-9}$. For a protein of length L, the null mutation rate is given by $\mu = 3.3^{-9}L$.

3 Results
3.1 Analytical fixation rate of allele 2
To calculate the fixation rate of allele 2 we have to consider the two fates of allele 1. Firstly, allele 1 can become lost. In this case allele 2 can only be generated during the period of drift of allele 1. The alternative fate of allele 1 is fixation. Then allele 2 can be generated either while allele 1 drifts to fixation or after allele 1 is already fixed. We would like to know how many mutation events from allele 1 to allele 2 are expected for either fate of allele 1. We let $n(s\beta)$ be the expected number of mutation events for when allele 1 is eventually fixed, and $n_{loss}(s\beta)$ be the expected number of mutation events for the case when allele 1 is lost. We can calculate $n(s\beta)$ and $n_{loss}(s\beta)$ from diffusion theory, by integrating over the sojourn times of allele 1. The corresponding calculations are cumbersome but straightforward, and for the sake of brevity we present the details in the Appendix (A.4 and A.5). For $n(s\beta)$, allele 2 can be generated as allele 1 drifts to fixation, and also after allele 1 has already reached fixation. For $n_{loss}(s\beta)$, allele 2 can only be generated while allele 1 drifts.

Assuming that m is the expected number of times allele 2 is generated, what is the probability that at least one copy goes to fixation? The probability of fixation of a single copy of allele 2 is $u(s)$ [26]. (In Appendix A.1, we reproduce the exact expression for $u(s)$, as well as approximations for large and small s.) Thus, if allele 2 is generated k times, its probability of fixation is $1 - [1 - u(s)]^k$. Since the probability that allele 2 is generated k times follows a Poisson distribution with mean m, we find for the probability v that at least one of the mutations to allele 2 goes to fixation

$$v = 1 - \sum_k \frac{m^k}{k!} e^{-m} [1 - u(s)]^k \qquad (1)$$
$$= 1 - e^{-mu(s)}.$$

We calculate this probability separately for $n(s\beta)$ and $n_{loss}(s\beta)$, setting m equal to either of these values. We assume that T is sufficiently large so that allele 1 has time to reach fixation within this interval (we assume $T \gtrsim 2N$). Then the probability $u_2(s, \beta)$ that allele 2 is generated and goes to fixation (starting with a single copy of allele 1) is

$$u_2(s, \beta) = u(s\beta)(1 - e^{-n(s\beta)u(s)}) + (1 - u(s\beta))(1 - e^{-n_{loss}(s\beta)u(s)}). \qquad (2)$$

The first half of the equation stems from the case when allele 1 eventually reaches fixation, where the probability that allele 1 becomes fixed, $u(s\beta)$, is multiplied by the probability v that at least one copy of allele 2 is generated and fixed. The second half corresponds to the case of loss of allele 1 from the population, where the probability of loss of allele 1, $(1 - u(s\beta))$, is multiplied by the probability of at least one mutation from allele 1 to allele 2 and subsequent fixation of allele 2. Taking into account allele 2 mutations during allele 1 loss is important especially for small s. Allele 1 is more likely to be lost than fixed for small s, but can occasionally drift for long times before being lost.

In the limit $\beta \to 0$, i.e., in the absence of phenotypic mutations, we find with Eqs. (A2), (A27), and (A35)

$$u_2(s, 0) = \frac{N+1}{N} - e^{-NU(T-N)u(s)} / N - e^{-NUu(s)}. \qquad (3)$$

(We assume that $N \gg 1$, and neglect corrections of order 1 compared to N. Note that we cannot simplify $(N + 1)/N$ to 1, because for small U, $1 - e^{-NUu(s)}$ and $(1 - e^{-NU(T-N)u(s)})/N$ are of the same order in N.) As we are interested in the effect of phenotypic mutations ($\beta > 0$) compared to the case without phenotypic mutations ($\beta = 0$), we define the increase in the probability of fixation from advantagous phenotypic mutations (the look-ahead effect) as

$$\xi = \frac{u_2(s,\beta)}{u_2(s,0)}. \qquad (4)$$

We can broaden the assumption of $T \gtrsim 2N$ to $T \to \infty$ with good accuracy. For $T \to \infty$, if allele 1 is destined to reach fixation, then the probability of generating at least one copy of allele 2 that goes to fixation approaches 1. Therefore, $1 - e^{-n(s\beta)u(s)} \to 1$, in this limit, and thus

$$\xi \approx \frac{u(s\beta)+(1-u(s\beta))(1-e^{-n_{\text{loss}}(s\beta)u(s)})}{(N+1)/N-e^{-NUu(s)}}. \qquad (5)$$

Apart from a correction for the case when allele 2 occurs while allele 1 is destined for extinction, Equation (5) is just the ratio of the probability of allele-1 fixation in the presence and absence of phenotypic mutations, $u(s\beta)/u(0) = Nu(s\beta)$.

To first order in $s\beta$, Eq. (5) simplifies to (Appendix A.6)

$$\xi \approx 1 + Ns\beta + O(s^2\beta^2). \qquad (6)$$

We can see from this equation that the look-ahead effect becomes important when N is on the order of $1/(s\beta)$.

For $Ns\beta \gg 1$, only the first term contributes to the numerator in Eq. (5), and we obtain (Appendix A.7)

$$\xi \approx \frac{(1-e^{-2s\beta})}{(N+1)/N-\exp[-NU(1-e^{-2s})]}. \qquad (7)$$

3.2 Simulations

We confirmed our analytic results for the fixation probabilities $u_2(s, \beta)$ and $u_2(s, 0)$ by numerical simulation, for different values of s (Figure 2). With a population size $N = 10^4$, the effect of phenotypic mutations can be seen for $s > 0.1$, and increases for larger s. For $s < 0.1$, the effect is too small and the intermediate allele is effectively neutral, meaning the fixation of allele 2 depends on the random fixation of the neutral allele 1. The look-ahead effect, ξ, shows the simulation results compared to Equations (5), (6) and (7). Figure 3 shows the magnitude of the look-ahead effect for the same parameter settings. For large s, the look-ahead effect can inflate the probability of fixa-
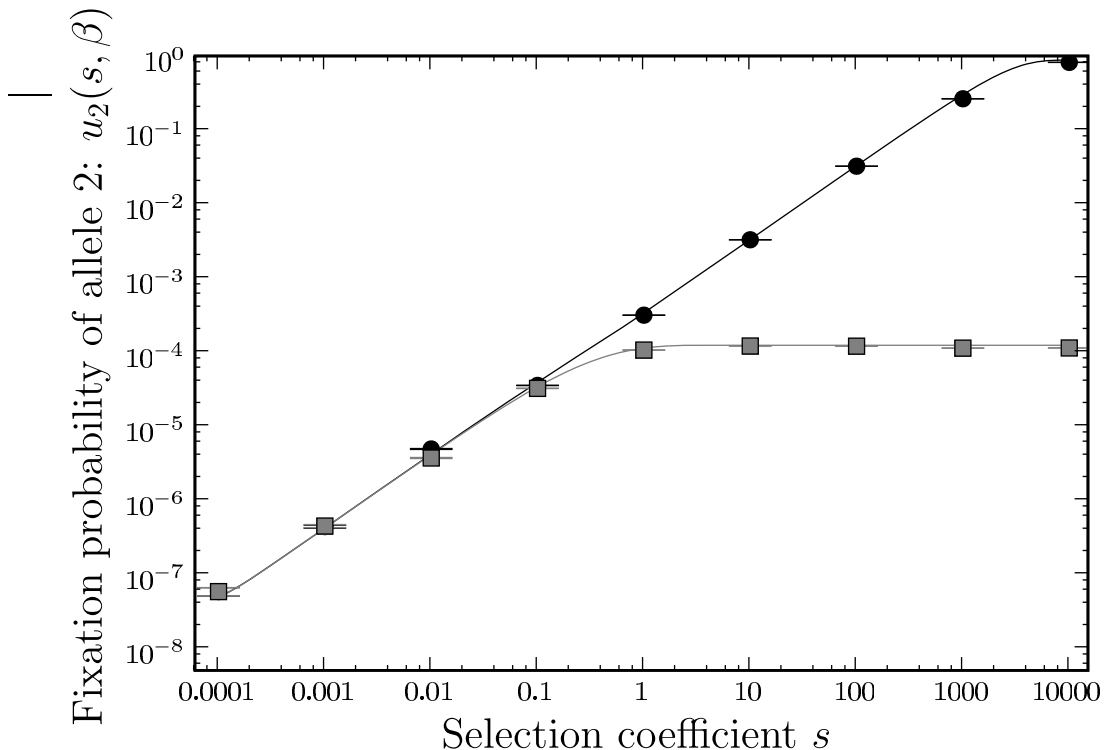

Figure 2

Fixation probability of allele 2 (u_2) vs. the selection coefficient s. Black is for $u_2(s, \beta)$, grey is for $u_2(s, 0)$. Solid lines are predictions according to Eq. (2) and (3), data points are for simulations with 10^9 repeats. $N = 10^4$, $U = \frac{8}{19}10^{-8}$, $\beta = 0.00019$, $T = 5 \times 10^5$. Error bars are standard errors.

tion of allele 2 by several orders of magnitude. We also display the different analytic expressions for ξ in Figure 3. The approximation (5), derived in the limit $T \to \infty$, works well for all values of s. The approximation (6), derived for small $s\beta$, captures correctly the magnitude of s at which the look-ahead effect starts to operate, i.e., $s \gtrsim 1/(N\beta)$. Similarly, approximation (7), valid for $Ns\beta \gg 1$, approximates ξ well for larger s.

Figure 4 shows ξ for different population sizes. As expected from the condition $s \gtrsim 1/(N\beta)$, the look-ahead effect will work with smaller selection coefficients s in larger populations. For large s, ξ saturates at approximately N.

We studied the effect of different values of the phenotypic error rate β (Fig. 5). As the error rate β increases, the look-ahead effect ξ increases by the same order of magnitude. For a very high phenotypic error rate of $\beta = 0.019$, the look-ahead effect is present for very small values of s. However, such a high error rate is likely to be severely detrimental, and in our model we do not take into account the loss of overall fitness for increasing phenotypic error

rates. Conversely for smaller β, the look-ahead effect is restricted to large s.

4 Discussion

We have described a model demonstrating the consequences of positive phenotypic mutations on the evolution of a single gene. We have compared numerical simulations with the analytical approximations and found them to be in good agreement. When phenotypic mutations exert an effect on fitness, selection can operate on the intermediate allele of a complex trait, which otherwise (without phenotypic mutations) would be neutral. We refer to selection for the intermediate allele as the *look-ahead effect*, because this effect allows evolution to select for sequences not yet in the genome.

The approximation for small $s\beta$, Eq.(6), shows most clearly the relationship between the parameters. The look-ahead effect is proportional to N, s, and β, and sets in when N is on the order of $1/(s\beta)$. For large $Ns\beta$, the look-ahead effect saturates. The asymptotic value of ξ is approximately N for $NU \ll 1$.
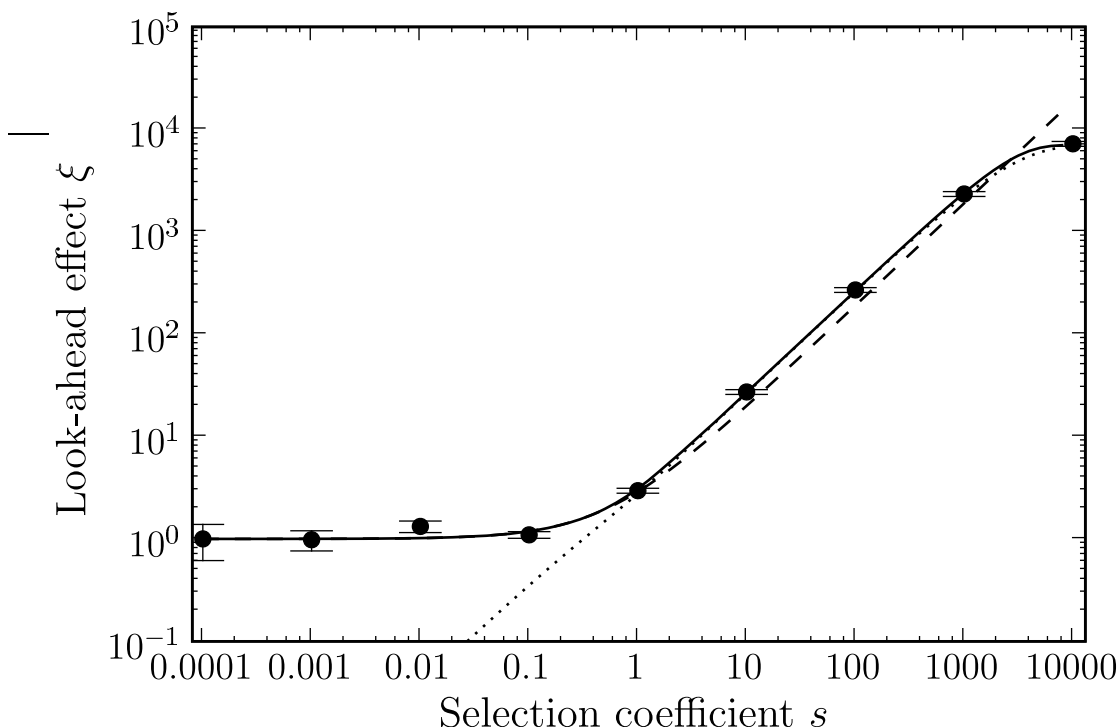

Figure 3

Look-ahead effect (ξ) due to phenotypic mutations vs. the selection coefficient s. The solid line is for Eq. (5), dashes are for Eq. (6), dots are for Eq. (7), and data points are for simulations with 10^9 repeats. $N = 10^4$, $U = \frac{8}{19}10^{-8}$, $\beta = 0.00019$, $T = 5 \times 10^5$. Error bars are standard errors.
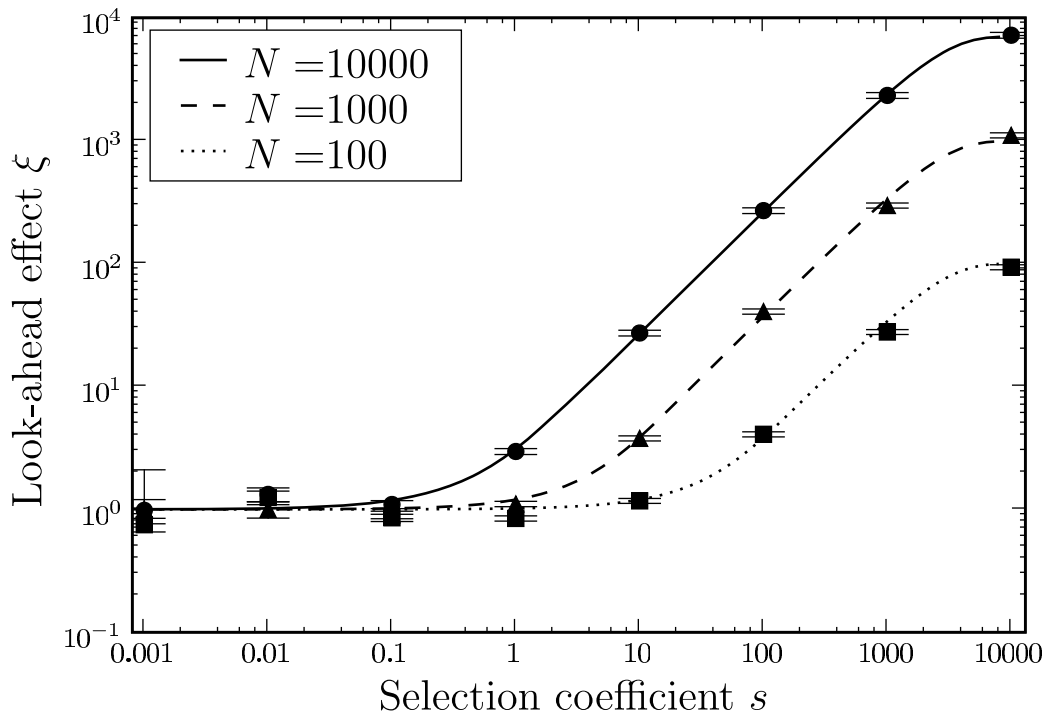
Figure 4

Look-ahead effect (ξ) due to phenotypic mutations vs. the selection coefficient *s* for different population sizes (N). Solid lines are from Equation (5), data points are for simulations with 10^8 repeats. $U = \frac{8}{19}10^{-8}$, $\beta = 0.00019$, $T = 5 \times 10^5$. Error bars are standard errors.

Therefore, large populations have two advantages over small populations in terms of the look-ahead effect: the effect sets in for smaller values of *s*, and saturates at a larger asymptotic value ξ. Of course, even in the absence of the look-ahead effect, larger populations can more easily traverse multiple local fitness peaks [9]. Because the selection coefficient *s* depends on the environment, a valid question is how often does *s* reach sufficiently high levels so that the look-ahead effect can operate. For microbial species such as bacteria, sufficiently large *s* should be reasonably common. Many bacteria experience highly fluctuating [27] and structured [28] environments, where growth is limited by the lack of a key trait. An obvious and extreme example is antibiotic resistance. Evolving a defense against an antibiotic molecule can involve only a few amino acids [29], and those individuals that can generate an enzyme capable of degrading the antibiotic, even if briefly or weakly, will have a very large fitness increase. In fact, if the efficacy of the antibiotic is 100% on susceptible genotypes, a mutation providing only moderate resistance has an infinite selective advantage. And even for very small antibiotic concentrations, mutants differing by only two amino acids at a single *β*-lactamase gene can be

selected effectively [30,31]. Thus, bacteria may frequently experience environments in which a large fitness increase (large *s*) is only a few mutations away. Similarly, in bacteriophages, selective coefficients *s* of 10 or more are not uncommon, even for individual mutations [32]. Our work is entirely theoretical, but we expect that it will be possible to experimentally verify our predictions in future work. For experimentally observing the look-ahead effect, we would need a system where *s* and *N* are both large, while *β* (the phenotypic mutation rate) can be modified. The values of both *N* and *s* used in this work are well within biologically realistic ranges achievable in a microbiological laboratory. Conditions for large *s* may be created with e.g. antibiotic resistance, which is a common laboratory workhorse. Unfortunately, many antibiotics function by reducing translation fidelity [33], and thus would conflate *s* and *β*. Changing *β* could involve a mutated ribosome. Ribosomes appear to be optimized for accurate and efficient translation of mRNA [34], and several examples of altered ribosome fidelity exist, both increasing [35] and decreasing fidelity [36]. Specific regions of the ribosome rRNA sequence have been identified as influencing fidelity [37], and various agents can
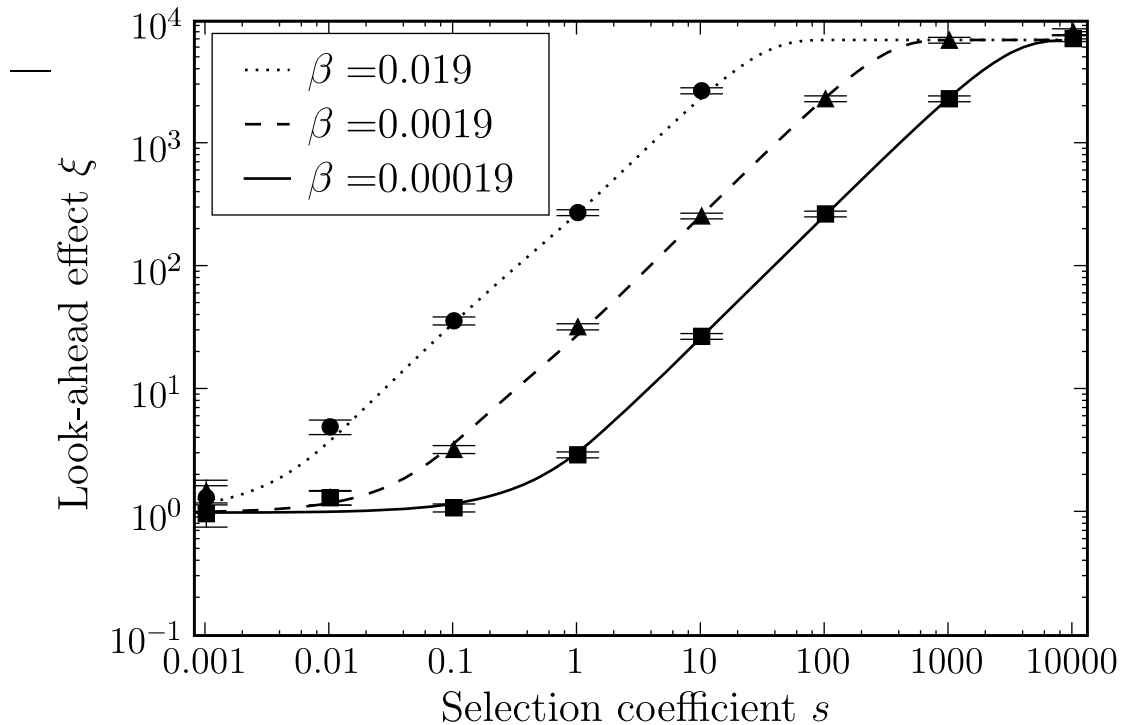
Figure 5

Look-ahead effect (ξ) due to phenotypic mutations vs. the selection coefficient *s* for different phenotypic error rates (β). Solid lines are from Equation (5), data points are for simulations with 10^8 repeats. $N = 10^4$, $U = \frac{8}{19} 10^{-8}$, $T = 5 \times 10^5$. Error bars are standard errors.

reduce fidelity, e.g., streptomycin, magnesium, and ethanol [36]. Few mutations may be sufficient to alter the fidelity of a ribosome, for example, a single mutation in the S5 ribosomal protein in *E. coli* increases frameshifting and nonsense mutations [38]. In yeast, mutations in the 18S RNA have been found that both increase and decrease translational fidelity [39].

In fact, a similar system to what we propose was already used to estimate the effects of tRNA competition on misreading error rates [13]. Here, several firefly luciferases were constructed with inactivating point mutations at an essential active-site lysine residue. Mistranslations of the transcript occur, rescuing the mutant and restoring the wild-type function. This system demonstrated a possible evolutionary constraint for the system presented in this work, in that some codons are more likely to be misread than others, depending on the relative amounts of tRNAs.

In this work, we have calculated the look-ahead effect from a comparison between the two cases of $\beta > 0$ and $\beta = 0$. The latter may not be experimentally possible; any

experiment will likely compare two different positive values of β. Nevertheless, Figure 5 shows that a larger look-ahead effect can be achieved with a higher β, where increasing β by one order of magnitude both increases the look-ahead effect by an order of magnitude and lowers the smallest *s* where an effect is observed. Of course, our model does not take into account the loss of fitness or other confounding effects from a higher phenotypic mutation rate. Thus, a balance must be found in having two different values of β that are different enough to measure, while at the same time minimizing the confounding effects. The most obvious consequence of increasing the phenotypic mutation rate is that overall fitness may be reduced, for example in *E. coli*, where a higher translational error rate activates stress responses [40], or in mouse, where such errors are implicated in neurodegeneration caused by misfolded proteins that aggregate [41]. Increasing translational fidelity may not come without fitness cost either. The hyperaccurate mutations in the 18S RNA in yeast [39] cause an increase in oxidative stress. This observation suggests that cells consume more energy to achieve hyperaccuracy. It may also partially explain

why the phenotypic error rate is much higher than the genotypic error rate, as there is possibly a direct disadvantage in reducing the phenotypic error rate, rather than only reducing the selective advantage that occurs if the phenotypic error rate is reduced, as discussed in [42].

Buerger *et. al* [42] asked whether evolution has selected for the current phenotypic error rate, which does not differ significantly between eukaryotes and prokaryotes [24,25] even though the source of errors is different. They suggested that the increase in fitness becomes incrementally smaller for improvements to transcription and translation fidelity. We would like to speculate that the phenotypic error rate is on the border between minimal costs (of e.g. misfolded proteins) and maximum payoff (via the look-ahead effect). The goal of our analysis was to demonstrate that the look-ahead effect is theoretically possible, and as such, we intentionally excluded confounding factors for the sake of clarity. There are several aspects not considered in our model that may play important roles. For example, in this work we did not consider the expression level. For low expressed genes, the mutation from allele 1 to allele 2 will occur less frequently compared to highly expressed genes. However, if allele 2 is produced it will be at a higher concentration (of allele 2 mutant proteins in a population of allele 1 proteins), as the overall copy number of allele 1 is low. This difference in expression levels is likely reduced in a large population, where beneficial mutations occur with sufficient frequency. Another factor related to the expression level is translational robustness. It has been proposed that highly expressed genes are under selection to properly fold despite phenotypic mutations, and consequently evolve slower [43,44]. If a gene is robust to translational errors, then it can tolerate a larger variety of mutations, of which some may be intermediates to a new adaptive multi-residue trait. Thus, translational robustness may increase the sequences available for experimentation at the phenotypic level. However, if the intermediate allele is itself not robust to errors in translation, then it will not be neutral, and may be selected against. The location of the protein trait will also influence the viability of the intermediate allele: mutations near the surface of the protein are less likely to disrupt the protein compared to mutations in the core [45].

In the presence of noise, phenotypic mutations may also help purge negative mutations [46]. If we have a system similar to the one described in this work but the final two-mutation trait is deleterious, then the phenotypic errors will lead to a selective disadvantage of the intermediate genotype. To give a concrete example, consider the case of prions, where an intermediate mutation favouring the formation of prions would be expressed at a small rate and would increase the liklihood of forming the misfolded proteins [47]. Since the majority of mutations are delete-

rious, the negative look-ahead effect is probably more common than the positive look-ahead effect on which we focused here.

In this work we use a single fitness optimum, and do not take into consideration multiple local optima as done by Borenstein *et al.* [21], who studied the effect of learning and of noisy phenotypes on evolution. Borenstein *et al.* considered varying learning rates, and showed that there is a trade-off between the amount of phenotypic plasticity and both the speed of reaching a local optimum and the genetic stability of the evolving individuals. It would be of interest to see if the same conclusions apply using our model, replacing phenotypic plasticity by a diffuse phenotype and learning rate by the phenotypic mutation rate. Such an analysis will require, however, that we explicitly model the deleterious spectrum of phenotypic mutations, and allow for different distributions of phenotypic mutations for allele 0, 1, and 2.

In conclusion, we propose that organisms can experiment with protein sequences that are mutationally close to the current sequence, but not yet in the genome. This effect allows selection for intermediates of complex traits, opening up a more direct route to the trait and thus reducing the time needed for fixation in the population.

5 Materials and methods

The numerical simulations were written in Java using the Colt scientific library [48] for the generation of random numbers. The analytic expressions were evaluated using both Mathematica and Python, the latter in conjunction with the SciPy package [49]. Source code for the numerical simulations is available on request from DJW.

The population in each simulation is represented by three numbers, corresponding to the abundance of each of the three alleles. As described, the initial abundances are $N - 1$, 1, 0 for alleles 0, 1, 2, respectively. The simulation runs for a specified number of generations T. We used $T = 5 \times 10^5$ throughout this work. Strictly speaking, T is the number of generations in which allele 1 can mutate into allele 2; for later generations this possibility of mutation is disabled. If allele 2 is present at time T, then the simulation is continued until allele 2 is either lost or has reached fixation. Generations are discrete, with mutations, selection, and drift occurring at each generation. During each generation we perform the following steps. First we check if either allele 0 or allele 2 has reached fixation; if so, we stop the simulation, as both cases are absorbing states. Next, for each allele we check for null mutations by drawing a random number from the Poisson distribution where the expected number of events is the null mutation rate μ multiplied by the total number of individuals with the given allele. Mutations from allele 1

to allele 2 are computed in a similar manner, where the expected number of events is *U* multiplied by the number of allele 1 individuals. Then, after the possible production of the mutant allele 2, selection acts on the fitness of the alleles, where the frequency of each allele is multiplied by its corresponding fitness, [1, 1 + *sβ*, 1 + *s*] for alleles [0, 1, 2], giving the new number of alleles in a possibly larger population. Finally, the next population of *N* individuals is chosen by recursively sampling from the binomial distribution, representing random genetic drift. Allele 0 is first sampled with the mean = (frequency of allele 0), and the (number of trials) = *N*. Allele 1 is then sampled from the combined allele 1 and 2 individuals. The number of simulations where allele 2 becomes fixed is divided by the total number of simulations, giving an estimate of the fixation probability. The number of simulations for each parameter set was between 10^8 and 10^9.

6 Authors' contributions

DJW, DV, COW, and EBB developed the original idea, DJW and COW performed the simulations and analysis, all authors contributed to the writing.

7 Reviewers' comments

7.1 Reviewers report 1

Eugene V. Koonin, NCBI, NLM, NIH, Bethesda, MD 20894, United States

The idea of this paper is as brilliant as it is pretty obvious...in retrospect. A novel solution is offered to the old enigma of the evolution of complex features in proteins that require two or more mutations (emergence of a disulphide bond is a straightforward example). Whitehead et al. propose that selection for such traits could be facilitated by phenotypic mutations (errors of transcription and, especially, translation). Due to phenotypic mutations, rare variants of proteins will emerge that are "pre-adapted" to accommodate the second, beneficial mutation, yielding the complex, adaptive trait, even if transiently. Simply put, for the case of a disulfide bond, one cysteine appears as a result of a phenotypic mutation and the other one due to a genotypic mutation. The result will be that, for a while, the cell will have in its possession the protein molecule with a disulfide bond. Thus, "pre-adaptation" owing to phenotypic mutation would promote fixation of the second mutation which will be beneficial even without the first one – if the selective advantage of the complex trait is high enough (the ultimate situation that helps understanding is that this trait is essential for survival). The actual fixation of the complex trait, then, requires only one (the first) mutation and is thus greatly facilitated. Mathematical modeling described in the paper shows that, if the selective advantage of the complex trait, i.e., the selection coefficient for the second mutation, is high enough, this look-ahead effect becomes realistic

under the experimentally determined mistranslation rates. Obviously, the realization of the look-ahead effect will depend on a variety of factors including the overall translation fidelity, the local context of the codon involved, the stability of the protein etc. This allows a number of rather straightforward experimental tests of the model.

From my perspective, this is a genuinely important work that introduces a new and potentially major mechanism of evolution and, in a sense, overturns the old adage of evolution having no foresight. It seems like, even if non-specifically and unwittingly, some foresight might be involved. At a more general conceptual level, this work is important in that it puts together, within a single conceptual framework, the evolutionary effects of genotypic and phenotypic mutations. There is much more to investigate here!

I would like to mention a rather general biological implication. It seems obvious enough that, under conditions of stress (e.g., amino acid starvation, heat shock etc), when translation fidelity drops, the look-ahead effect will be enhanced. Thus, this could be a general and crucial mechanism of adaptation during evolution.

Eugene Koonin

Author response: *We would like to thank Eugene Koonin for his enthusiastic and positive review.*

7.2 Reviewers report 2

Subhajyoti De, MRC Laboratory of Molecular Biology Hills Road, Cambridge CB2 2QH, United Kingdom

I have read the revised manuscript, and have found that all points raised by the referees were fully addressed. The work is rigorous and very interesting, and I believe, will make a significant contribution in the field. I'll be happy to consider it for publication.

Subhajyoti De

Author response: *We would like to thank Subhajyoti De for feedback that improved the original manuscript, and the subsequent positive review.*

7.3 Reviewers report 3

David Krakauer, Santa Fe Institute, United States

In this paper the authors demonstrate how phenotypic variation arising through errors in development (e.g. transcription and translation), can, when building on (amplifying) genetic variation, accelerate the fixation rate of neutral alleles. By assuming that neutral alleles are genet-

ically closer to an optimum genotype than a mutation-free wild-type, this can also reduce the time required to reach the optimum. The result is illustrated through stochastic simulation and some limiting-case analytical approximations.

This is an interesting paper that is technically rigorous, and correct in many of the conclusions that it reaches. The paper is now much improved as it now includes specific reference to the almost identical, Baldwin effect. As the authors correctly state, many of papers on the Baldwin effect emphasize learning, but a significant fraction explore the role of random ontogenetic variation on evolutionary dynamics, and a few, explicitly consider the adaptive value of errors in transcription and translation on the exploration of fitness landscapes. It is not yet clear how important the differences are between treating Baldwin effects in terms of individual ontogenetic programs versus population level dynamics. In both cases, the key insight is that random variation is capable of generating a more effective gradient for population dynamics.

I think it worthwhile therefore to give a brief review of this mechanism and a little of its literature.

A Synoptic Outline Of the Baldwin-Morgan-Osborn Effect

1. The essential insight of Baldwin and several other 19th century biologists (listed above) was to understand that phenotypic plasticity can have a direct effect on genetic evolution. In some cases, this can give rise to the appearance of Lamarckian inheritance, as selection on plastic phenotypes derived from a single genotype, can lead to the fixation of polymorphic sequences generating these phenotypes without plasticity.

2. The modern investigation of this effect is associated with the work of Hinton and Nowlan (1987) who showed that ontogenetic variability or plasticity, could lead to effective genetic optimization in neutral fitness landscapes.

3. This has been followed by numerous papers exploring complex landscapes, diverse models of plasticity, including learning, homeostasis, diffusion, and combinatorial sampling. See Turney (1996) for a review with an emphasis on computational approaches.

4. Ancel and Fontana (2000) (building on some more theoretical work by Ancel) demonstrated for RNA secondary structure, the crucial requirement that phenotypic plasticity and genetic polymorphism should exhibit a particular correlational structure for the Baldwin effect to be effective.

5. The most recent, and somewhat exhaustive analysis of the Baldwin effect has been conducted by Borenstein et al (2006) in fluctuating landscapes, exploring both directed and random phenotypic variation.

6. Krakauer and Sasaki (2002) demonstrated a "negative Baldwin effect" whereby developmental errors could amplify mildly deleterious mutations in finite populations, thereby leading to their effective purging.

Certainly the paper by Krakauer and Sasaki does not consider learning explicitly, but something much closer to the so called "look ahead effect" described by Whitehead et al, as it treats the ensemble of variant proteins generated by a single underlying sequence as a result of errors in transcription or translation. In both the Baldwin effect and the "look ahead" effect, genetically identical organisms generate phenotypically diverse populations. I think it an interesting subject for future work to establish the precise nature of any differences manifesting at the level of population dynamics, rather than at the incidental level, of mechanism.

Author response: *We appreciate this correction of a large hole in our background literature. We have cited relevant literature about the Baldwin effect, and discussed the main differences between the look-ahead effect and the Baldwin effect. While on the surface the look-ahead effect is very similar to the Baldwin effect, crucially the Baldwin effect is about individual learning, whereas the look-ahead effect is about errors that always produce different proteins from a single gene, at a given rate. Thus, in our model there is little difference between individuals with the same genotype, as no learning is involved, as opposed to the Baldwin effect, where, due to learning, two organisms with identical genotypes can have very different phenotypes. Therefore, we believe that it is important to distinguish clearly between the cases with and without learning, and to use different terminology to emphasize this distinction.*

I was somewhat confused by the remark that double mutations are neglected because they are very rare.

Firstly, double mutations should be allowed within the binomial model presented by the authors. Secondly, the statement is empirically false for many haploid genomes. Bonhoeffer and Nowak (1997) showed that in large populations double mutants are likely to exist at fairly high abundance.

Author response: *We agree that for RNA-based viral genomes, which often have genomic mutation rates 1000 times greater than DNA-based organisms, double mutations occur frequently. Our model focused on DNA-based organisms, where double mutations are rare. If we wanted to apply our model to RNA viruses, we would have to include double muta-*

tions. *However, the results from such a modification are obvious: If double mutations are frequent, the organism will happen upon the beneficial double mutation quickly and not require the look-ahead effect at all.*

The treatment of deleterious mutations remains a little confusing. Presumably developmental noise can both amplify existing deleterious effects (e.g. cryptic genetic variation, sensu Gibson & Dworkin 2004) and contribute novel pathologies, orthogonal to those of the underlying transcript (e.g. gain of function mutations). This should be made an explicit, distributional property of the model rather than assuming a fixed background cost.

Author response: *The explanation of how we treat deleterious mutations was extremely brief in our original draft, and we have expanded and clarified the respective paragraph. We believe that a more explicit, complex treatment of deleterious effects would detract from the main message the model in this work was meant to convey. We have added to the discussion how phenotypic mutations can amplify deleterious genotypic mutations. A more complete treatment of deleterious phenotypic mutations will be a topic of future work.*

A Appendix
Here, we present the details of our analytic derivations.

A.1 Probability of fixation
According to [26], the probability of fixation $u(s)$ of a single allele with selection coefficient s is given by

$$u(s) = \frac{1-e^{-2s}}{1-e^{-2Ns}}. \qquad (A1)$$

For $s \gtrsim 1/N$, this expression simplifies to

$$u(s) = \frac{1}{N} + \frac{N-1}{N}s + O(s^2), \qquad (A2)$$

whereas for $Ns \gg 1$, this expression simplifies to

$$u(s) \approx 1 - e^{-2s}. \qquad (A3)$$

A.2 A single allele drifting to fixation or loss
We first consider a single allele with selective advantage s drifting to fixation or extinction, and ask how many mutations this allele generates until it is either fixed or lost. We will treat these two cases separately. Let $n_{\text{fix}}(s)$ be the expected number of mutations generated while the allele drifts to fixation, and let $n_{\text{loss}}(s)$ be the expected number of mutations generated while the allele drifts to extinction. We calculate these two quantities using diffusion theory, by integrating the sojourn times of the allele over all frequencies.

For an allele with selective coefficient s and starting at frequency $p = 1/N$, [50] calculated its mean sojourn time $\tau(\gamma)$ between frequencies γ and $\gamma + d\gamma$ as

$$\tau(\gamma) = 2[V(\gamma)G(\gamma)]^{-1}[u_{\text{loss}}(1/N)g(0, \gamma)\theta(1/N - \gamma) + u_{\text{fix}}(1/N)g(\gamma, 1)\theta(\gamma - 1/N)]. \qquad (A4)$$

Here,

$$V(\gamma)G(\gamma) = \gamma(1 - \gamma)e^{-2Ns\gamma} / N, \qquad (A5)$$

$$g(a, b) = \frac{e^{-2Nsa} - e^{-2Nsb}}{2Ns}, \qquad (A6)$$

$$u_{\text{loss}}(p) = \frac{e^{-2Nsp} - e^{-2Ns}}{1 - e^{-2Ns}}, \qquad (A7)$$

$$u_{\text{fix}}(p) = 1 - u_{\text{loss}}(p) = \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}}, \qquad (A8)$$

and $\theta(z)$ is the Heaviside step function. We want to integrate expressions involving $\tau(\gamma)$ from $\gamma = 0$ to $\gamma = 1$. Since $\gamma = 1/N$ corresponds to a single copy of the allele that drifts to fixation, values of γ less than $1/N$ are not relevant for our analysis. Therefore, we discard the term proportional to $\theta(1/N - \gamma)$ in Eq. (A4), and use in what follows

$$\tau(\gamma) = 2u_{\text{fix}}(1/N)g(\gamma, 1)/[V(\gamma)G(\gamma)] \quad \text{for } \gamma > 1/N. \qquad (A9)$$

A.3 Number of mutations conditional on fixation
For the sojourn time conditional on fixation, $\tau_{\text{fix}}(\gamma)$, [50] finds

$$\tau_{\text{fix}}(\gamma) = \tau(\gamma)u_{\text{fix}}(\gamma)/u_{\text{fix}}(p). \qquad (A10)$$

Using this expression, we have

$$n_{\text{fix}}(s) = NU\int_{1/N}^{1} \tau_{\text{fix}}(\gamma)\gamma d\gamma. \qquad (A11)$$

Plugging the expressions for $V(\gamma)G(\gamma)$, $g(a, b)$, $u_{\text{fix}}(p)$, and $\tau(\gamma)$ into $\tau_{\text{fix}}(\gamma)$, we arrive at

$$\tau_{\text{fix}}(\gamma) = \frac{1}{s(1-e^{-2Ns})} \frac{(1-e^{-2Ns\gamma})(1-e^{-2Ns(1-\gamma)})}{\gamma(1-\gamma)}. \qquad (A12)$$

This expression corresponds to the one by [51]. Note that $\gamma\tau_{\text{fix}}(\gamma) \to 0$ for $\gamma \to 0$. Therefore, we can extend the lower limit of integration to 0 in Eq. (A11), and rewrite $n_{\text{fix}}(s)$ as

$$n_{\text{fix}}(s) = \frac{NU}{s(1-e^{-2Ns})} I(2Ns) \qquad (A13)$$

with

$$I(a) = \int_0^1 \frac{(1-e^{-a\gamma})(1-e^{-a(1-\gamma)})}{1-\gamma} d\gamma. \qquad (A14)$$

The integral $I(a)$ can be rewritten as

$$I(a) = \gamma - \text{Ei}(-a) + \ln(a) + e^{-a}[\gamma - \text{Ei}(a) + \ln(a)], \qquad (A15)$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant and $\text{Ei}(z)$ is the exponential integral,

$$\text{Ei}(z) = -\int_{-z}^{\infty} \frac{e^{-t}}{t} dt. \qquad (A16)$$

For $s \lesssim 1/N$, we find

$$n_{\text{fix}}(s) = N^2 U + O(s^2). \qquad (A17)$$

For $Ns \gg 1$, we obtain the asymptotic expansion

$$n_{\text{fix}}(s) \approx \frac{NU}{s}[\ln(2Ns) + \gamma], \qquad (A18)$$

using [52] 5.1.51,

$$\text{Ei}(-z) \sim -\frac{e^{-z}}{z}\left(1 - \frac{1}{z} + \frac{2}{z^2} - \frac{6}{z^3}\right) \quad \text{for large } z. \qquad (A19)$$

A.4 Number of mutations conditional on extinction
For the sojourn time conditional on extinction, $\tau_{\text{loss}}(\gamma)$, [50] finds

$$\tau_{\text{loss}}(\gamma) = \tau(\gamma)u_{\text{loss}}(\gamma)/u_{\text{loss}}(p). \qquad (A20)$$

Using this expression, we have

$$n_{\text{loss}}(s) = NU\int_{1/N}^1 \tau_{\text{loss}}(\gamma)\gamma d\gamma. \qquad (A21)$$

Plugging the expressions for $V(\gamma)G(\gamma)$, $g(a, b)$, $u_{\text{loss}}(p)$, and $\tau(\gamma)$ into $\tau_{\text{loss}}(\gamma)$, we find

$$\tau_{\text{loss}}(\gamma) = \frac{1}{s(1-e^{-2Ns})} \frac{e^{2s}-1}{1-e^{-2(N-1)s}} \frac{(e^{-2Ns\gamma}-e^{-2Ns})(1-e^{-2Ns(1-\gamma)})}{\gamma(1-\gamma)}. \qquad (A22)$$

We rewrite n_{loss} as

$$n_{\text{loss}} = \frac{NU}{s(1-e^{-2Ns})} \frac{e^{2s}-1}{1-e^{-2(N-1)s}} J(N,s) \qquad (A23)$$

with

$$J(N,s) = \int_{1/N}^1 \frac{(e^{-2Ns\gamma}-e^{-2Ns})(1-e^{-2Ns(1-\gamma)})}{1-\gamma} d\gamma. \qquad (A24)$$

The integral can be rewritten as

$$J(N, s) = -2e^{-2Ns}(\gamma - \text{Chi}[2(N-1)s] + \ln[2(N-1)s]), \qquad (A25)$$

where $\text{Chi}(z)$ is the hyperbolic cosine integral,

$$\text{Chi}(z) = \gamma + \ln(z) + \int_0^z \frac{\cosh(t)-1}{t} dt. \qquad (A26)$$

For $s \lesssim 1/N$, we find

$$n_{\text{loss}}(s) = (N-1)U + O(s^2). \qquad (A27)$$

For $Ns \gg 1$, we obtain the asymptotic expansion

$$n_{\text{loss}}(s) \approx \frac{U}{2s^2}(1 - e^{-2s}), \qquad (A28)$$

using

$$\text{Chi}(z) \approx \frac{\text{Ei}(z)}{2} \approx \frac{e^z}{2z} \quad \text{for large } z. \qquad (A29)$$

[This expansion follows directly from the definitions of $\text{Chi}(z)$, $\cosh(z)$, and $\text{Ei}(z)$.]

A.5 Number of mutations within a given time interval
We now extend the derivations in Section A.3 to calculate the number of mutations to allele 2 generated within a certain time interval T, conditional on fixation of allele 1. We assume that T is sufficiently large so that allele 1 has time to reach fixation within this interval. We only consider the case conditional on fixation because no new mutations are generated once allele 1 has gone extinct.

We calculate $n(s) = n_{\text{fix}}(s) + n_T(s)$, where $n_T(s)$ is the total number of mutations generated once the first mutation has reached fixation. We have

$$n_T(s) = NU[T - t_{\text{fix}}(s)], \qquad (A30)$$

where $t_{\text{fix}}(s)$ is the time to fixation of a mutation with selective advantage s. This time is given by the integral over all sojourn times,

$$t_{\text{fix}}(s) = \int_0^1 \tau_{\text{fix}}(\gamma)d\gamma = \frac{I_2(2Ns)}{s(1-e^{-2Ns})} \quad (A31)$$

with

$$I_2(a) = \int_0^1 \frac{(1-e^{-a\gamma})(1-e^{-a(1-\gamma)})}{\gamma(1-\gamma)}\,d\gamma. \quad (A32)$$

A partial fraction decomposition of the integrand reveals that $I_2(a) = 2I(a)$, and thus we have

$$t_{\text{fix}}(s) = \frac{2I(2Ns)}{s(1-e^{-2Ns})} \quad (A33)$$

Combining this result with Eqs. (A13) and (A30), we find

$$n(s) = n_{\text{fix}}(s) + n_{\text{T}}(s) = NU\left[T - \frac{I(2Ns)}{s(1-e^{-2Ns})}\right]$$
$$= NUT - n_{\text{fix}}(s). \quad (A34)$$

Note that $n(s) = n_{\text{fix}}(s)$ for $T = t_{\text{fix}}(s)$.

For $s \lesssim 1/N$, we find

$$n(s) = NU(T - N) + O(s^2). \quad (A35)$$

For $Ns \gg 1$, using Eqs. (A15) and (A19), we obtain the asymptotic expansion

$$n(s) \approx NU\left(T - \frac{\ln(2Ns)+\gamma}{s}\right). \quad (A36)$$

A.6 ξ for sβ << 1
From Eq. (4), using Eqs. (A27), (A35), and (A2), we obtain to first order in $s\beta$

$$\xi \approx 1 + \frac{e^{-NUu(s)}-e^{-NU(T-N)u(s)}}{u_2(s,0)}s\beta + O(s^2\beta^2). \quad (A37)$$

If further $NU(T - N)u(s) << 1$, we obtain

$$\xi \approx 1 + N(1 - N/T)s\beta + O(s^2\beta^2), \quad (A38)$$

and for $T \to \infty$,

$$\xi \approx 1 + Ns\beta + O(s^2\beta^2). \quad (A39)$$

A.7 ξ for Ns$\beta \gg 1$
For $Ns\beta \gg 1$, only the first term contributes to Eq. (2), and we obtain from Eqs. (A36) and (A3)

$$u_2(s,\beta) = (1 - e^{-2s\beta})\left[1 - \exp\left(-NU\left[T - \frac{\ln(2Ns\beta)+\gamma}{s\beta}\right]\left(1 - e^{-2s}\right)\right)\right]. \quad (A40)$$

Likewise, in this limit we can simplify Eq. (3) to

$$u_2(s,0) = \frac{N+1}{N} - \exp[-NU(T-N)(1 - e^{-2s})]\,/\,N - \exp[-NU(1 - e^{-2s})], \quad (A41)$$

giving

$$\xi \approx \frac{(1-e^{-2s\beta})\left[1-\exp\left(-NU\left[T-\frac{\ln(2Ns\beta)+\gamma}{s\beta}\right]\left(1 - e^{-2s}\right)\right)\right]}{(N+1)/N-\exp[-NU(T-N)(1-e^{-2s})]/N-\exp[-NU(1-e^{-2s})]}. \quad (A42)$$

Furthermore, for $T \to \infty$, this expression simplifies to

$$\xi \approx \frac{(1-e^{-2s\beta})}{(N+1)/N-\exp[-NU(1-e^{-2s})]}. \quad (A43)$$

If $NU << 1$, then $\xi \to N$ in the limit $s \to \infty$.

References
1. Cairns J, Overbaugh J, Miller S: **The origin of mutants.** *Nature* 1988, **335(6186):**142-5.
2. Bridges BA: **The role of DNA damage in stationary phase ('adaptive') mutation.** *Mutat Res* 1998, **408:**1-9.
3. Foster PL: **Mechanisms of stationary phase mutation: a decade of adaptive mutation.** *Annu Rev Genet* 1999, **33:**57-88.
4. Cairns J: **Mutation and cancer: the antecedents to our studies of adaptive mutation.** *Genetics* 1998, **148(4):**1433-1440.
5. Hall BG: **Adaptive mutagenesis: a process that generates almost exclusively beneficial mutations.** *Genetica* 1998, **102–103(1–6):**109-125.
6. Rosenberg SM: **Evolving responsively: adaptive mutation.** *Nat Rev Genet* 2001, **2(7):**504-15.
7. Foster PL: **Adaptive mutation: has the unicorn landed?** *Genetics* 1998, **148(4):**1453-1459.
8. Behe MJ, Snoke DW: **Simulating evolution by gene duplication of protein features that require multiple amino acid residues.** *Protein Sci* 2004, **13(10):**2651-64.
9. Weinreich DM, Chao L: **Rapid evolutionary escape by large populations from local fitness peaks is likely in nature.** *Evolution Int J Org Evolution* 2005, **59(6):**1175-1182.
10. Lynch M: **Simple evolutionary pathways to complex proteins.** *Protein Sci* 2005, **14(9):**2217-25. discussion 2226–7
11. Springgate CF, Loeb LA: **On the fidelity of transcription by Escherichia coli ribonucleic acid polymerase.** *J Mol Biol* 1975, **97(4):**577-91.
12. Edelmann P, Gallant J: **Mistranslation in E. coli.** *Cell* 1977, **10:**131-7.

13. Kramer EB, Farabaugh PJ: **The frequency of translational misreading errors in E. coli is largely determined by tRNA competition.** *RNA* 2007, **13:**87-96.
14. Drake JW, Charlesworth B, Charlesworth D, Crow JF: **Rates of spontaneous mutation.** *Genetics* 1998, **148(4):**1667-86.
15. Chothia C, Lesk AM: **The relation between the divergence of sequence and structure in proteins.** *EMBO J* 1986, **5(4):**823-826.
16. Baldwin JM: **A new factor in evolution.** *American Naturalist* 1896, **30:**441-451.
17. Hinton GE, Nowlan SJ: **How Learning Can Guide Evolution.** *Complex Systems* 1987, **1:**495-502 [http://htpprints.yorku.ca/archive/00000172/01/hinton-nowlan.htm].
18. Turney P: **Myths and legends of the Baldwin Effect.** *Proceedings of the ICML-96 (13th International Conference on Machine Learning)* 1996 [http://citeseer.ist.psu.edu/turney96myths.html].
19. Mery F, Kawecki TJ: **Experimental evolution of learning ability in fruit flies.** *Proc Natl Acad Sci USA* 2002, **99(22):**14274-14279.
20. Ancel LW, Fontana W: **Plasticity, evolvability, and modularity in RNA.** *J Exp Zool* 2000, **288(3):**242-283.
21. Borenstein E, Meilijson I, Ruppin E: **The effect of phenotypic plasticity on evolution in multipeaked fitness landscapes.** *J Evol Biol* 2006, **19(5):**1555-1570.
22. Martin LC, Gloor GB, Dunn SD, Wahl LM: **Using information theory to search for co-evolving residues in proteins.** *Bioinformatics* 2005, **21(22):**4116-4124.
23. Bloom JD, Drummond DA, Arnold FH, Wilke CO: **Structural determinants of the rate of protein evolution in yeast.** *Mol Biol Evol* 2006, **23(9):**1751-1761.
24. Ellis N, Gallant J: **An estimate of the global error frequency in translation.** *Mol Gen Genet* 1982, **188(2):**169-72.
25. Shaw RJ, Bonawitz ND, Reines D: **Use of an in vivo reporter assay to test for transcriptional and translational fidelity in yeast.** *J Biol Chem* 2002, **277(27):**24420-6.
26. Kimura M: **On the probability of fixation of mutant genes in a population.** *Genetics* 1962, **47:**713-9.
27. Smit E, Leeflang P, Gommans S, Broek J van den, van Mil S, Wernars K: **Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods.** *Appl Environ Microbiol* 2001, **67(5):**2284-2291.
28. Baquero F, Negri MC: **Selective compartments for resistant microorganisms in antibiotic gradients.** *Bioessays* 1997, **19(8):**731-736.
29. Palzkill T, Le QQ, Venkatachalam KV, LaRocco M, Ocera H: **Evolution of antibiotic resistance: several different amino acid substitutions in an active site loop alter the substrate profile of beta-lactamase.** *Mol Microbiol* 1994, **12(2):**217-229.
30. Baquero F, Negri MC, Morosini MI, Blázquez J: **The antibiotic selective process: concentration-specific amplification of low-level resistant populations.** *Ciba Found Symp* 1997, **207:**93-105. discussion 105-11
31. Baquero F, Negri MC, Morosini MI, Blázquez J: **Selection of very small differences in bacterial evolution.** *Int Microbiol* 1998, **1(4):**295-300.
32. Bull JJ, Badgett MR, Wichman HA: **Big-benefit mutations in a bacteriophage inhibited with heat.** *Mol Biol Evol* 2000, **17(6):**942-950.
33. Ogle JM, Ramakrishnan V: **Structural insights into translational fidelity.** *Annu Rev Biochem* 2005, **74:**129-177.
34. Baxter-Roshek JL, Petrov AN, Dinman JD: **Optimization of ribosome structure and function by rRNA base modification.** *PLoS ONE* 2007, **2:**e174.
35. Vila-Sanjurjo A, Ridgeway WK, Seymaner V, Zhang W, Santoso S, Yu K, Cate JHD: **X-ray crystal structures of the WT and a hyper-accurate ribosome from Escherichia coli.** *Proc Natl Acad Sci USA* 2003, **100(15):**8682-8687.
36. Friedman SM, Berezney R, Weinstein IB: **Fidelity in protein synthesis. The role of the ribosome.** *J Biol Chem* 1968, **243(19):**5044-5048.
37. O'Connor M, Thomas CL, Zimmermann RA, Dahlberg AE: **Decoding fidelity at the ribosomal A and P sites: influence of mutations in three different regions of the decoding domain in 16S rRNA.** *Nucleic Acids Res* 1997, **25(6):**1185-1193.
38. Kirthi N, Roy-Chaudhuri B, Kelley T, Culver GM: **A novel single amino acid change in small subunit ribosomal protein S5 has**

39. **profound effects on translational fidelity.** *RNA* 2006, **12(12):**2080-2091.
40. Konstantinidis TC, Patsoukis N, Georgiou CD, Synetos D: **Translational fidelity mutations in 18S rRNA affect the catalytic activity of ribosomes and the oxidative balance of yeast cells.** *Biochemistry* 2006, **45(11):**3525-3533.
40. Fredriksson A, Ballesteros M, Peterson CN, Persson O, Silhavy TJ, Nyström T: **Decline in ribosomal fidelity contributes to the accumulation and stabilization of the master stress response regulator sigmaS upon carbon starvation.** *Genes Dev* 2007, **21(7):**862-874.
41. Lee JW, Beebe K, Nangle LA, Jang J, Longo-Guess CM, Cook SA, Davisson MT, Sundberg JP, Schimmel P, Ackerman SL: **Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration.** *Nature* 2006, **443(7107):**50-55.
42. Buerger R, Willensdorfer M, Nowak MA: **Why are phenotypic mutation rates much higher than genotypic mutation rates?** *Genetics* 2006, **172:**197-206.
43. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH: **Why highly expressed proteins evolve slowly.** *Proc Natl Acad Sci USA* 2005, **102(40):**14338-14343.
44. Wilke CO, Drummond DA: **Population genetics of translational robustness.** *Genetics* 2006, **173:**473-481.
45. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS: **The stability effects of protein mutations appear to be universally distributed.** *J Mol Biol* 2007, **369(5):**1318-1332.
46. Krakauer DC, Sasaki A: **Noisy clues to the origin of life.** *Proc Biol Sci* 2002, **269(1508):**2423-2428.
47. Weissmann C: **The state of the prion.** *Nat Rev Microbiol* 2004, **2(11):**861-871.
48. Colt Project: **Colt: Open Source Libraries for High Performance Scientific and Technical Computing in Java.** 2007 [http://acs.lbl.gov/~hoschek/colt/].
49. SciPy: **Scientific Tools for Python.** 2007. http://www.scipy.org/
50. Nagylaki T: **The moments of stochastic integrals and the distribution of sojourn times.** *Proc Natl Acad Sci USA* 1974, **71(3):**746-749.
51. Ewens WJ: **Conditional diffusion processes in population genetics.** *Theor Popul Biol* 1973, **4:**21-30.
52. Abramowitz M, Stegun IA: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* New York: Dover, ninth dover printing, tenth gpo printing edition; 1964.