Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Ineffectual Genomic Error Correction Under Environmental Perturbation Dynamically Regulates Mutational Supply and Robustness

Sashikanta Barik
National Institute of Technology Durgapur
Parthasarathi Sahu
National Institute of Technology Durgapur
Koushik Ghosh
National Institute of Technology Durgapur
Hemachander Subramanian
hemachander@gmail.com

National Institute of Technology Durgapur

Research Article

Keywords: Kinetic proofreading, Punctuated equilibrium, Coding region length, Mutation robustness, Replication fidelity, Dynamic Balance

Posted Date: May 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9543896/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Molecular Evolution on August 27th, 2026. See the published version at https://doi.org/10.1007/s00239-026-10341-4.

Ineffectual Genomic Error Correction Under Environmental Perturbation Dynamically Regulates Mutational Supply and Robustness

Sashikanta Barik[1], Parthasarathi Sahu[1], Koushik Ghosh[1], and Hemachander Subramanian*[1]

[1]Department of Physics, National Institute of Technology Durgapur, West Bengal, India 713209

Abstract

Adaptation requires a adequate supply of heritable variation, yet too much mutation can limit organismal viability by disrupting essential molecular functions. Our study shows that this trade-off arises inherently from the kinetic proofreading mechanism that dictates the replication fidelity. In our model, environmental shifts alter the optimal driving rate constant of proofreading enzymes, transiently elevating replication error rates and triggering rapid evolutionary change until a new fidelity optimum is reached. This produces alternating periods of stasis and rapid adaptation, the central signature of punctuated equilibrium. We further show that coding-region length and population size jointly determine whether adaptation succeeds or mutational collapse occurs, reflecting the balance between mutation supply and error tolerance predicted by the drift-barrier principle. These results provide a molecular mechanism underlying nearly neutral evolutionary dynamics and illustrate how genomic organization limits long-term evolutionary resilience.

Keywords: Kinetic proofreading, Punctuated equilibrium, Coding region length, Mutation robustness, Replication fidelity, Dynamic Bala nce

Significance Statement

What do genomic error correction mechanisms have to do with an organism's evolutionary trajectory? Does rapid environmental change increase organisms' adaptability? We show that genomic error correction mechanisms become ineffectual under rapid environmental change, because enzymes responsible for error correction function sub-optimally in altered environments. The resultant increased error rates boost adaptability/plasticity, allowing organisms to rapidly alter their genotype and phenotype to enhance survival. This analysis could be helpful in estimating the adaptive rates of organisms during the human-induced rapidly changing climate.

1 Introduction

Evolution proceeds as a complex and nonlinear process shaped by the continual production of heritable variation and the differential preservation of that variation across environments [1]. Because most genetic differences originate from replication errors, the rate at which new variants are introduced—and the extent to which organisms can tolerate their consequences—defines a fundamental axis of evolutionary dynamics [2, 3]. This tension is captured by the interplay between *mutation supply*, which regulates the emergence of new variants at the evolutionary level [4, 5], and *mutational robustness*, which mitigates or rectifies

*Corresponding author: hemachander@gmail.com

their potentially deleterious effects at the molecular level. Robustness is achieved via fidelity mechanisms, including network-level redundancy, and genomic error correction mechanisms [6, 7]. Too few mutations prevent adaptive variation, while too many mutations increase the genetic load and the risk of extinction [8, 9]. Robust systems can gather cryptic or almost neutral variants that become adaptive under the environmental changes [10, 7]. Our study aims to understand how organisms solve this conflict between simultaneous requirements of novelty and stability.

This study shows that kinetic proofreading, a tunable, energy-driven mechanism that ensures genomic replication fidelity, also governs the balance between novelty and stability. Proofreading improves accuracy through substrate discrimination via non-equilibrium energy expenditure, which lowers replication or translation errors that go beyond the limits set by equilibrium-binding constants alone [11, 12]. Proofreading efficiency depends on rate constants of substrate binding and dissociation with the proof-reading enzyme, which are functions of the environmental parameters such as temperature. This environmental dependence of proofreading rate constants function as a molecular throttle on mutation supply, enabling cells to adjust replication and translation fidelity in response to physiological and evolutionary pressures. Under stable conditions, proofreading keeps mutation rates low and maintains mutational robustness.

However, this balance between novelty and stability in replication and translation, maintained partly through kinetic proofreading, is disrupted due to changes in the environment, such as physical (e.g., radiation, temperature extremes, osmotic or hypoxic stress), chemical (e.g., biocides, heavy metals, antibiotics, reactive oxygen species), and biological stressors (e.g., viral infection, nutrient limitation). These stressors disrupt the cellular homeostasis and increase mutational supply [13, 14, 15, 16, 17, 18, 19, 20, 21]. In these situations, proofreading should act as the primary error-correcting mechanism that can minimize the high error rates. But proofreading enzymes are also encoded in the genome, so they can be changed (mutated) by stress-induced mutations. In this study, we show that the environmental perturbations can diminish the efficiency of proofreading precisely at the moment when fidelity is essential, thereby increasing the mutation supply, prior to the re-optimization of the underlying kinetic proofreading rate constants by selection, resulting in restoration of high fidelity replication. This initial disruption of proofreading due to environmental changes and subsequent evolutionary re-optimization causes a temporary spike in mutations, followed by a slow return to a state with fewer errors as the selection restores the proofreading performance.

The survival and adaptability of an organism does not just depend on the error rates, but total errors accumulated during replication. Longer coding regions increase the mutational target size, and therefore adversely affect survival, while smaller genomes limit the sequence space available for mutations, and therefore, the ability to adapt [22, 23, 8]. Some studies show that the genes, which are highly functionally constrained and under purifying selection, are typically short and compact [24, 25], while organisms experiencing relaxed selective pressure can accommodate enlarged coding and non-coding regions [23]. Therefore, selection for an "optimal" coding region length is important to achieve a balance between preserving stability and allowing for evolutionary change.

Population size too plays a role in balancing stability and novelty. Selection is more effective in large populations, where it can fine-tune the length of the coding region and therefore the total error, whereas the genetic drift dominates over these length variations in smaller populations, facilitating mildly deleterious expansions/contractions to persist [4, 26, 27, 28]. Therefore, the changes in fidelity caused by environmental stress have different effects on evolution depending on the size of the population. This determines whether genomes stay in a stable balance between mutation supply and robustness or move toward maladaptive paths.

To investigate the relationship between mutational supply and robustness, short and long coding regions, and small and large population sizes, we establish a theoretical framework that integrates replication fidelity with genome evolution through the Hopfield–Ninio kinetic proofreading mechanism [11]. In this framework, the proofreading enzyme's driving-rate constant sets the replication error rate and allows for an optimum that reduces mutational input when conditions are stable. Environmental changes, like temperature fluctuations, move the system away from this optimal state and transiently raise the mutation rates. This indicates that fidelity needs to be re-optimized through evolution.

By incorporating this altered proofreading landscape into a population-genetic model, we illustrate that the error-correcting systems can adjust to high-fidelity phenotypes even without the direction of a fitness function. The development of this phenotype via the evolution of coding region length is significantly influenced by population size. Intermediate-length coding regions (characterised by mutation rates [29, 30,

31]) demonstrate the highest adaptive potential, swiftly re-optimizing the driving-rate constant and the selection of coding region length. In contrast, very long regions accumulate excessive mutational burden and are prone to collapse under increased mutation supply, while short coding regions evolve slowly with restricted adaptive response, persisting only in sufficiently large populations. Therefore, our results show that the evolution of coding-region length and the fidelity is not regulated by molecular constraints alone but arises from the interaction between mutational supply and robustness, regulated by the population-size–dependent balance between selection and drift.

2 Kinetic Proofreading Model

The error correction mechanism in biochemical processes, like DNA replication, genetic code reading in protein synthesis, and tRNA charging with amino acids, plays a crucial role in preserving and transferring genomic information with high accuracy across generations. The mechanism relies on enzymes discriminating the correct substrates from similar molecules. In the early 1970s, J.J. Hopfield and J. Ninio independently proposed the kinetic proofreading model to explain this discriminatory mechanism [11, 12]. Hopfield's interpretation of error correction involves the examination of substrate flux (substrate turnover rate) throughout an enzyme-substrate reaction pathway. The enzyme-substrate reaction pathway can be characterized by a series of successive interactions between enzymes and substrates, involving the formation of intermediate complexes, which facilitate the conversion of substrates into products, as illustrated in Fig. (1). Following this reaction pathway, the enzyme effectively incorporates correct substrates while discerning and excluding incorrect substrates to ensure correct product formation.

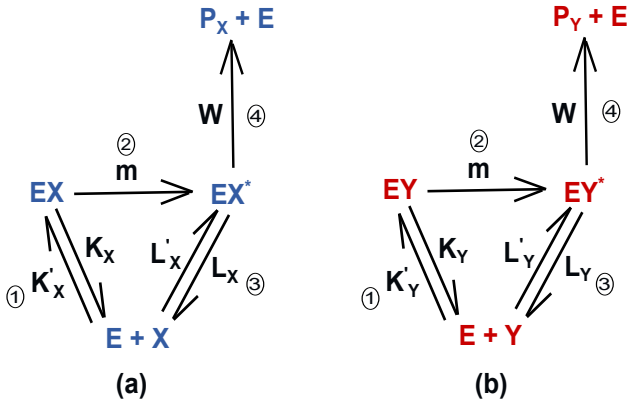


Figure 1: *The figure illustrates the correct and incorrect enzyme-substrate reaction pathways with blue colored and red colored chemical participants, respectively. This reaction scheme is identical to the one employed by Hopfield in his seminal kinetic-proofreading paper. E denotes enzyme, while X and Y denote correct and incorrect substrates, leading to formation of P_X and P_Y as correct and incorrect products, respectively. These pathways encompass multiple reaction steps with intermediate complexes resulting from enzyme and substrate interactions. EX (EY) and EX^* (EY^*) signify the correct (incorrect) enzyme-substrate complex and intermediate high-energy complex, respectively. $K'_{X/Y}$ and $L'_{X/Y}$ are the association rate constants of intermediate complexes, whereas $K_{X/Y}$ and $L_{X/Y}$ represent their corresponding dissociation rate constants. Note that the rate constants 'm' and 'w' are the same for both the reaction pathways. In step 2, 'm' signifies the driving rate constant (coupled with an external energy source, such as phosphate hydrolysis), while 'w' in step 4 facilitates the progression of the intermediate high-energy complex towards product formation before its dissociation through step 3.*

Fig. (1) illustrates the enzyme-substrate reaction network, where X and Y denote the correct and incorrect substrates, respectively. K'_X and K'_Y at step 1 represent rate constants for the association of substrates (X or Y) with the active site of enzymes, which leads to the formation of enzyme-substrate complex (EX or EY). K_X and K_Y denote the corresponding dissociation rate constants. The driving rate constant 'm' at step 2, associated with energy consumption via phosphate hydrolysis, drags the enzyme-substrate complex

to a high-energy transition state. This transition-state complex (EX^* or EY^*) can proceed to product formation through step 4 with rate constant w or it can reset the whole reaction by dissociating into free enzymes and substrates through step 3 with a rate constant L_X (or L_Y). The association rate constant, L'_X (or L'_Y), at step 3, is for the thermodynamically uphill, direct transition of enzymes and substrates to a high-energy intermediate complex (EX^* or EY^*), without energetic assistance.

In Fig. 1, the driving rate constant 'm' at step 2 and the product-formation rate w at step 4 are identical for both pathways and are treated as effectively irreversible transitions, assumptions made by Hopfield in his seminal article [11]. These assumptions are based on the energetics and kinetics of DNA replication. During the construction of the daughter strand, DNA polymerase incorporates a dNTP at the growth front of the daughter strand, that creates a phosphodiester bond with the previously incorporated nucleotide and releases pyrophosphate (PPi). Rapid PPi hydrolysis by cellular pyrophosphatases suppresses the reverse reaction effectively, making step 2 highly irreversible [32, 33, 34]. Step 4 is also an irreversible step, due to the release of the products into the surrounding medium. This indicates that once the product is released, the system proceeds to the next catalytic cycle, i.e., incorporation of the next nucleotide [11, 35].

The difference in the binding strength between the enzyme and the correct and incorrect nucleotides during the initial binding of a dNTP with the polymerase–DNA complex, prior to any conformational change, is relatively small [36, 37, 38]. Furthermore, the enzyme-nucleotide association is largely diffusion-limited, meaning that the rate at which the nucleotide binds is largely determined by how quickly molecules diffuse and collide in solution rather than by strong substrate-specific interactions. Because diffusion governs this step, different substrates (nucleotides) tend to associate with similar rates, and the enzyme cannot strongly discriminate between correct and incorrect substrates during the initial binding event. Consistent with this picture, experiments report only modest (≈2–5-fold) differences in initial binding rates [39, 40]. Larger apparent differences occur only when the "on-rate" reflects a conformational-closing step rather than diffusion-limited binding, in which case, the discrimination no longer arises from proofreading. Therefore, we have assumed the association rate constants to be equal, $K'_X = K'_Y$ and $L'_X = L'_Y$ and specificity arises only from the differences in off-rates. Nevertheless, we also examine the effect of moderate variations in association rates on our results in Fig. S10.

Therefore, the discriminatory capability of the reaction arises solely from the difference between kinetic parameters for the *dissociation* of correct and incorrect substrates from enzymes. For effective discrimination, the dissociation rate constants of EY and EY^* should be greater than those of EX and EX^*, i.e., $K_Y > K_X$ and $L_Y > L_X$. This rate difference impedes the conversion of incorrect substrates into products by trapping them within the enzyme-substrate reaction pathway, while allowing the correct substrates to proceed to product formation.

Within these reaction pathways, steps 1, 2, and 3 in Fig. 1 collectively compose a cyclic path, where step 1 initiates the reaction with the formation of the enzyme-substrate complex, step 2 drives the enzyme-substrate complex unidirectionally to a transition state and step 3 resets the reaction back to enzyme and substrate, which are inputs for state 1, prior to engagement in step 4 for product formation. This cyclic path facilitates the circulation of substrate flux, which is necessary for error correction. The cyclic path entraps the incorrect substrates for a longer time than the correct substrates, resulting in a comparative decrease in the rate of incorrect product formation. Consequently, the greater circulation frequency of incorrect substrates compared to correct ones results in enhanced error correction [41]. This entrapment process is enhanced when the substrates are cycled unidirectionally, enhancing the flux in the step 1 - step 2 - step 3 direction, and reducing the flux in the opposite step 3 - step 2 - step 1 direction, which is accomplished by employing an energetically driven stage between steps 2 and 3, which takes EX and EY to EX^* and EY^*, irreversibly. This thermodynamic drive increases the time delay between the correct and incorrect product formations by increasing the number of times the incorrect substrate cycles through states 1, 2, and 3, compared to the correct substrate, due to their dissociation rate differences.

3 Mathematical Model

As discussed in the previous section, the dynamics of substrate flux within the enzyme-substrate reaction pathway plays a crucial role in error correction. To investigate the dynamics of substrate flux, we apply the rate law to the enzyme-substrate reaction pathway, illustrated in Fig. (1), considering both correct and

incorrect substrates.

For each substrate $i \in \{X, Y\}$, the enzyme–substrate pathway can be expressed as,

$$\frac{d[P_i]}{dt} = w\,[Ei^*], \tag{1a}$$

$$\frac{d[Ei^*]}{dt} = m\,[Ei] + L_i'\,[E][i] - (L_i + w)\,[Ei^*], \tag{1b}$$

$$\frac{d[Ei]}{dt} = K_i'\,[E][i] - (m + K_i)\,[Ei]. \tag{1c}$$

Equation (1a) describes the rate of formation of the correct (and incorrect) product, while Equations (1b) and (1c) describe the formation rates of intermediate complexes for correct (and incorrect) substrates.

We apply the steady-state approximation to the above rate equations for both correct and incorrect enzymatic reaction pathways: $\frac{d[Ei^*]}{dt} = 0$ and $\frac{d[Ei]}{dt} = 0$. Simplifying the Equations (1b and 1c with the steady-state approximation and defining the error rate (f) as the ratio of the rate of incorrect product formation to the rate of correct product formation, we get

$$f = \left\{\frac{m \cdot K_Y' + (m + K_Y) \cdot L_Y'}{m \cdot K_X' + (m + K_X) \cdot L_X'}\right\} \cdot \frac{(K_X + m)(L_X + w)}{(K_Y + m)(L_Y + w)}. \tag{2}$$

Equation (2) describes the dependence of the error rate on the kinetic parameters of the correct and incorrect reaction pathways. All the kinetic parameters involved in both the correct and incorrect pathways are equal, except the dissociation rate constants.

Furthermore, we have considered the above reaction pathway (Fig. 1) in the absence of the driving step to find out the non-driven error rate.

$$\mathrm{E} + \mathrm{i} \underset{\mathrm{K}_i}{\overset{\mathrm{K}_i'}{\rightleftharpoons}} \mathrm{Ei} \xrightarrow{\mathrm{w}} \mathrm{P}_i + \mathrm{E} \tag{3}$$

By applying the steady-state approximation to the following rate equations that quantify the dynamics of the above reactions,

$$\frac{d[P_i]}{dt} = w[Ei] = constant \tag{4a}$$

$$\frac{d[Ei]}{dt} = K_i'[E][i] - (K_i + w)[Ei] = 0 \tag{4b}$$

$$\tag{4c}$$

We derive the non-driven error rate as

$$f_0 = \frac{K_X + w}{K_Y + w}. \tag{5}$$

Assuming $w \to 0$, the non-driven error rate is, $f_0 = \frac{K_X}{K_Y}$. The non-driven error rate shows the energetic discrimination between the correct and incorrect pathways with a single discriminating step.

The error rates (2) and (5) can alternatively be expressed in terms of free energy disparity between the correct and incorrect substrates at the enzyme's recognition site. The enzyme's ability to distinguish between correct and incorrect substrates relies on this free energy difference: A greater disparity in free energy enhances the enzyme's discrimination between correct and incorrect substrates, facilitating error correction and reducing the error rate.

For the reaction pathway with two-step discrimination, step 1 and step 3, illustrated in Fig. (1), the error rate in terms of free energy disparity is given by,

$$f \equiv \exp(\frac{-\Delta G_{XY}^*}{RT}) \tag{6}$$

Here, ΔG^*_{XY} represents the effective free energy difference between correct (X) and incorrect (Y) substrates at the enzyme's recognition sites throughout the whole reaction pathway [11, 42, 43, 44]. According to Equations (2) and (6), this free energy disparity depends on the rate of dissociation of intermediate complexes in the two discriminating paths. Therefore, an increase in the dissociation rate of the incorrect intermediate complexes amplifies the number of circulations of incorrect substrates in the cyclic pathway formed by steps 1, 2, and 3, leading to an apparent increase in ΔG^*_{XY}. Analogously, the free energy disparity between X and Y at the enzyme's recognition sites in a single-step discrimination path (Eq. 3) can be denoted as ΔG_{XY}, and the associated error rate f_0 can be alternately written as $f_0 = \exp(\frac{-\Delta G_{XY}}{RT})$.

In Hopfield's model [11], the total error rate of the enzyme-substrate biochemical pathway with two discriminating paths can attain a minimal value equal to the square of the error rate resulting from the pathway with single-step discrimination ($f_{min} = f_0^2$), which implies that the maximum value of ΔG^*_{XY} is twice that of ΔG_{XY}. To achieve this minimum error rate and maximum effective free energy disparity, consumption of external energy is required [11, 45, 46, 47], through the driving step 2 of Fig. (1).

Similarly, the external energy consumption can also be quantified in terms of the driving rate constant (m). According to the Arrhenius equation [48, 49], 'm' can be represented in terms of an activation energy barrier (kinetic barrier height E_a) between the intermediate complexes.

$$m \equiv A \exp(\frac{-E_a}{RT}) \tag{7}$$

or,

$$\log(\frac{m}{A}) = -(\frac{E_a}{RT}), \tag{8}$$

where E_a is the activation energy, A denotes the pre-exponential factor, R is the universal gas constant, and T is the absolute temperature. This energy is required to drag the enzyme-substrate complexes to a high energetic transition state during the driving step 2, which is obtained through phosphate hydrolysis, such as dNTP and GTP hydrolysis, in the case of DNA replication and translation process, respectively [50, 51].

4 Evolutionary framework

Our aim here to model phenotypic adaptation of error-correcting enzymes, driven by genetic variation that arises when genomic error correction becomes ineffective under environmental perturbation. We assume that mutations in the perturbed environment primarily affect the catalytic rate m of the enzyme, representing the energy-consuming transition step (e.g., phosphate hydrolysis at the active site) that drives proofreading. Although multiple kinetic parameters that characterize the error-correcting enzyme's function may vary, in reality, many of them do not vary substantially from environmental perturbation to have a noticeable effect on error correction, as we argue below. Moreover, varying a single representative rate captures the dominant adaptive response while avoiding over-parameterization. We therefore treat m as the evolving kinetic trait characterizing the enzyme phenotype.

4.1 Temperature-shift as environmental perturbation

Environmental perturbations are introduced via a shift in temperature, which predominantly influences the dissociation rate constants in biochemical processes. Higher temperatures destabilize hydrogen bonds between nucleotides, leading to their faster dissociation. The association rate constants ($K'_{x/y}$, $L'_{x/y}$) correspond to diffusion-limited encounters between a nucleotide and the polymerase–DNA complex, which therefore produces only modest variation with temperature. In contrast, hydrogen bond dissociation rates vary as $K \propto e^{-\Delta G/RT}$, rendering H-bond lifetimes strongly temperature dependent. In DNA hybridization, for instance, the dissociation rate increases exponentially with temperature, whereas the association rate shows only a weak negative temperature dependence [52]. Thus, association rate constants ($K'_{x/y}$, $L'_{x/y}$) remain comparatively insensitive, whereas dissociation rate constants ($K_{x/y}$, $L_{x/y}$) respond strongly to thermal fluctuations.

Although both dissociation rate constants, $K_{x/y}$ and $L_{x/y}$, appear in the reaction network, only K (K_x and K_y) are treated as temperature-dependent in our model. Variations in L (L_x and L_y) do not affect the

error rate as long as their ratio $\frac{L_x}{L_y}$ remains fixed. This follows from Eq. 2, where, considering only the step 3 dissociation, the relevant discrimination term appears as $\frac{L_x+w}{L_y+w}$; under the assumption $w \to 0$, the error rate depends solely on $\frac{L_x}{L_y}$. Consistently, the optimal driving constant m_0 is independent of L (see Eq. 14).

The temperature dependence of the dissociation rate constant between enzyme and substrate, K, follows the transition state theory:

$$K = \frac{K_B T}{h} \exp(-\frac{\Delta G^{\ddagger}}{RT}) \tag{9}$$

where $\Delta G^{\ddagger}$ is the free energy of activation for the dissociation reaction. Eq. 9 gives $K_x = 2s^{-1}$ at 293 K for the dissociation of correct nucleotides from polymerase during DNA replication [35, 53]. We estimate $\Delta G^{\ddagger} = 70.023\ kJmol^{-1}$, which also aligns with [53].

4.2 Genotype-Phenotype Mapping

We next establish a genotype–phenotype relationship to link molecular error rates (genotype) to the evolving driving rate constant (phenotype). The phenotypic change of the enzyme is defined as

$$\Delta m = \frac{1}{\gamma} \cdot l \cdot \langle f \rangle \tag{10}$$

Where, Δm is the change in the enzyme phenotype (which is the change in the driving rate constant in units of s^{-1}, as the result of mutations), l is the coding region length, $\langle f \rangle$ is the average mutations per base pair per generation (which is the average error rate of the population in the current generation defined in Eq. 2) and γ is a proportionality constant in units of $mutations/generation/sec$ (here, $\gamma = 1\ mutations/generation/sec$).

Following Eq. 10, $\gamma = \frac{l \cdot \langle f \rangle}{\Delta m}$, or,

$$\gamma = \frac{Total\ mutations\ per\ generation}{change\ in\ phenotype}.$$

In simpler terms, γ quantifies how many total mutations per generation are required to produce a unit change (here, $\Delta m = 1s^{-1}$) in the phenotypic driving rate constant.

Equation 10 is motivated by mutation-supply–limited evolutionary dynamics. The quantity $l\langle f \rangle$ represents the total number of mutations entering the population per generation (genomic mutation supply). The proportionality constant γ converts the number of mutational events into a change in the phenotypic parameter m, thereby providing a minimal genotype–phenotype mapping in which evolutionary change of the phenotype slows naturally as mutation supply decreases near the adapted state [54, 55, 56].

4.3 Adaptive Dynamics

To observe how mutations shape phenotypic traits in the course of adaptation, we implement the above genotype-to-phenotype mapping to an initial population of N_0 individuals (enzymes), whose trait values, that is, the driving rate constants, ($m_{individual}$) are normally distributed around the pre-shift optimum ($m_{0,old}$) with a small standard deviation ($\sigma = 0.001$) in the initial unperturbed environmental condition (Temperature, T_0). This indicates that individuals have driving rate constants very close to the optimum, thereby maintaining an error rate on the order of 10^{-8}. A temperature shift ΔT, introduced via Eq. 9, increases the replication error rate owing to the absence of immediate phenotypic adjustment. This elevated error rate (Fig. S6) drives phenotypic adaptation toward the new optimum ($m_{0,\text{new}}$) in the perturbed environment ($T_{\text{new}} = T + \Delta T$) through genomic mutations, as described by Eq. 10.

To link error rates to survival, we define the individual elimination probability:

$$P_{\text{kill}}(m_i) = \frac{f_i}{f_{\text{th}}}, \tag{11}$$

where $f_i(m_i)$ is the error rate of the i^{th} individual, determined by the equation (2), and f_{th} ($= 10^{-6}$) is the threshold error rate beyond which error catastrophe is assumed to occur and survival of the individual

is assumed to be improbable. The threshold value takes into account the fact that high mutation rates are typically deleterious [57], i.e., when mutation rates exceed the threshold, essential genes such as those encoding DNA polymerase would accumulate disruptive mutations, abolishing their activity. Eq. 11 stochastically dictates the elimination probability of an individual, which therefore allows for even unfit individuals to sometimes survive and reproduce. Individuals with high P_{kill} values have a higher probability of being selected out, whereas, low P_{kill} value individuals' evolutionary trajectories are dominated by drift.

We also define the probability of emergence of new individuals (mutated variants) with trait $m_i + \Delta m$, $m_i - \Delta m$, and m_i into the population. We define the offspring probability, P_{off}, which depends on the relative survivability of individuals. This offspring probability is therefore defined as a function of selection coefficient $s(m_i)$, below:

$$P_{\text{off}}(m_i) = \frac{\exp(as(m_i))}{1 + \exp(as(m_i))}, \qquad \text{where } s(m_i) = 1 - P_{\text{kill}}(m_i). \tag{12}$$

The logistic form was chosen because it provides a bounded, sigmoidal mapping between the selection coefficient and offspring probability, ensuring that individuals with higher survival likelihood ($s \to 1$) have a progressively greater probability of producing offspring, while even maladapted individuals ($s \to 0$) retain a baseline offspring probability of 0.5, thereby allowing drift to influence evolutionary outcomes. In addition to this baseline behavior, the steepness factor 'a' (exponent of Eq. 12) is chosen as 5 (see *Supplementary Information*, Fig. S4) to amplify the discrimination between high- and low-fitness individuals, thereby mimicking the strong selective pressures typically encountered under evolutionary constraints.

Such logistic mappings are widely used in population genetics and quantitative evolutionary models to capture nonlinear selective advantages under finite resources and competition [58, 59]. This offspring probability controls the additions of new mutated variants ($m_i + \Delta m$, $m_i - \Delta m$ and m_i) to the population. For instance, if $P_{off}(m_i)$ is high (≥ 0.75), then two offspring will be selected at random from the set of three (new variants), ensuring that each of the three possible pairwise combinations occurs with equal probability of $\frac{1}{3}$. If $P_{off}(m_i)$ is low (≤ 0.75), then only one offspring will be selected out of the three variants randomly with equal probability of $\frac{1}{3}$.

This offspring probability is stochastic but depends on the selection coefficient (s). However, this choice of the stochastic boundary does not affect our result (see *Supplementary Information*, Fig. S8). This stochasticity is introduced to increase unpredictability and to better match real simulations. Once offspring probabilities are assigned and new individuals are added, the overall population size must also be regulated. If the population exceeds the carrying capacity, $N_{\text{cap}} = 2N_0$, individuals are uniformly down-sampled at random, without bias toward higher-fitness individuals. This implements a simple resource constraint, ensuring that the population never exceeds N_{cap}, which is common in population genetic studies.

Following this simulation algorithm, we performed 100 independent simulation replicates to ensure statistical robustness and tracked the evolutionary trajectory of m_{avg} ($m_{avg} = \langle m_i \rangle$), population size and adaptation rate across each generation, averaged over all the independent replicates.

5 Parameter Regime

To connect molecular kinetics with evolutionary adaptation, we model DNA replication within the proofreading scheme originally proposed by Hopfield. As shown in Fig. 1, E represents the DNA polymerase (T7 polymerase), and X/Y denote the correct and incorrect nucleotides, respectively. The framework illustrates nucleotide incorporation into the template strand by the polymerase.

In step 1, an incoming deoxynucleotide triphosphate (dNTP) binds to the polymerase at the growing end of the template, forming enzyme–substrate complexes (EX and EY). In step 2, hydrolysis of the triphosphate releases pyrophosphate, driving the enzyme–substrate complexes into a transition state (EX^* and EY^*). From this transition state, the pathway can proceed either to product formation in step 4 (successful nucleotide incorporation) or to substrate dissociation in step 3. In DNA replication, the baseline error rate (f_0) is typically on the order of 10^{-4}, whereas the observed minimum error rate (f_{min}) lies between 10^{-8} and 10^{-9} [60, 61, 62, 36]. This substantial reduction in error rate is achieved through energy expenditure during step 2.

Here, we have assumed that $K_X = \alpha K_Y$ and $L_X = \alpha L_Y$, where, α represents the energetic bias between correct and incorrect pathways at steps 1 and 3. Furthermore, the association rate constant at step 1 is assumed to be much larger than that at step 3, i.e., $(K'_X = K'_Y) \gg (L'_X = L'_Y)$, where L'_X and L'_Y remain small due to the direct transition into the high-energy state ($E + X \rightarrow EX^*$ or $E + Y \rightarrow EY^*$). Under these constraints, correct nucleotides remain bound to the enzyme's active site longer than incorrect ones, yielding a higher turnover of correct products. Finally, we set $w \rightarrow 0$ so that incorrect substrates undergo repeated cycles through steps 1–3, thereby enhancing error correction.

The results presented in this work are robust across a wide range of kinetic parameter values, as long as they satisfy the constraints outlined above. For our analysis, we employed the following representative set of parameters: $K'_X = K'_Y = 250\ \mathrm{s}^{-1}$, $L'_X = L'_Y = 10^{-2}\ \mathrm{s}^{-1}$, $L_X = 0.2\ \mathrm{s}^{-1}$, and $K_X = 2\ \mathrm{s}^{-1}$ (Table. S1). These values are comparable to the parameters used in previous studies [35, 53], though direct adoption is not possible since those reaction networks differ slightly from ours, which is aligned with Hopfield's proofreading scheme. Finally, we set the energetic discrimination factor to $\alpha = 10^{-4}$ to ensure consistency with the baseline error rate, $f_0 = \frac{K_X}{K_Y} = 10^{-4}$.

6 Results

In this study, we examine how environmental perturbations influence the fidelity of genomic processes, such as replication and translation, and show that such perturbations lead to evolutionary dynamics resembling punctuated equilibrium. We analyze how sudden temperature shifts affect the efficiency of the error-correction of the proofreading enzyme. We observe that the temperature shifts disrupt the evolutionarily finely tuned discrimination steps of kinetic proofreading that help in maintaining the low error rates. This results in an increase in mutational supply, leading to rapid adaptive change until the enzyme re-optimizes its driving rate constant and restores a new error-minimizing configuration under the perturbed environment, thereby reestablishing evolutionary stasis.

We also investigate the evolutionary consequences of such perturbations on the coding region length. Because coding-region length determines both the mutational input available for adaptation and the vulnerability to accumulating harmful changes, evolutionarily dictated coding sequence lengths reflect a balance between mutational supply and mutational robustness. This balance is mediated by the competition between natural selection and genetic drift, which erodes weak refinements, particularly in small populations. Together, these two forces determine the conditions under which different coding-region lengths can support successful adaptation or collapse under elevated mutational pressure.

6.1 Variation of optimal proofreading with environmental change

The flux of substrates through the discriminating steps (steps 1 and 3 in Fig. 1) is regulated by the driving step. At very low driving rate m, progression from the enzyme–substrate complexes (EX and EY) to the transition-state complexes (EX^* and EY^*) is comparatively slower than their dissociation. Even the relatively stable EX complexes dissociate before advancing, while EY dissociates even more rapidly. As a result, both correct and incorrect complexes predominantly dissociate to become free enzyme and nucleotide, making discrimination ineffective through step 1 and suppressing the product formation. On the other hand, for a very high value of the driving rate m, EX and EY are rapidly driven into EX^* and EY^*, leaving virtually no complexes at step 1 for discrimination through the dissociation pathway. In this regime, discrimination again relies primarily on step 3, while step 1 contributes negligibly (see Eq. 13).

Thus, at both extremes of m, step 1 ceases to contribute to discrimination—either because complexes dissociate before reaching the transition state (low m) or because they are driven forward too rapidly for differential dissociation to occur (high m). In both cases, error correction is governed solely by step 3, yielding the limiting error rate:

$$f_{m\rightarrow 0} = f_{m\rightarrow \infty} = \frac{L_X}{L_Y} \tag{13}$$

For intermediate values of m, however, both steps 1 and 3 are actively engaged in discrimination, lowering the error rate below α, the value associated with single-step discrimination. Hence, an intermediate,

optimized driving rate constant is crucial for maximizing fidelity. The optimal m is obtained by minimizing the error rate (Eq. 2) with respect to the driving rate constant.

$$\frac{df}{dm} = 0$$

The optimized value for driving constant (M_0) is given by

$$M_0 = \Big(\frac{q-1}{q} K_x K_y\Big)^{1/2},$$

or,

$$M_0 = K_x \Big(\frac{q-1}{q\alpha}\Big)^{1/2}. \tag{14}$$

Here,

$$q = 1 + \frac{L'_x}{K'_x} = 1 + \frac{L'_y}{K'_y}.$$

From Equations (2) and (14), the minimum error rate can be approximated, in the limit $w \to 0$ and $M_0 \leq K_x$ (see Fig. 2c and supplementary information), as

$$f_{min} = f(M_0) \approx \frac{K_X}{K_Y}\frac{L_X}{L_Y}.$$

Therefore, $f_{min} \approx \alpha^2$. This result demonstrates that the error rate associated with the reaction pathway involving two discriminating steps is simply the square of the error rate observed in a single-step discrimination pathway. Such a reduction is attainable by investing *optimal* driving energy, provided the conditions of approximation above hold.

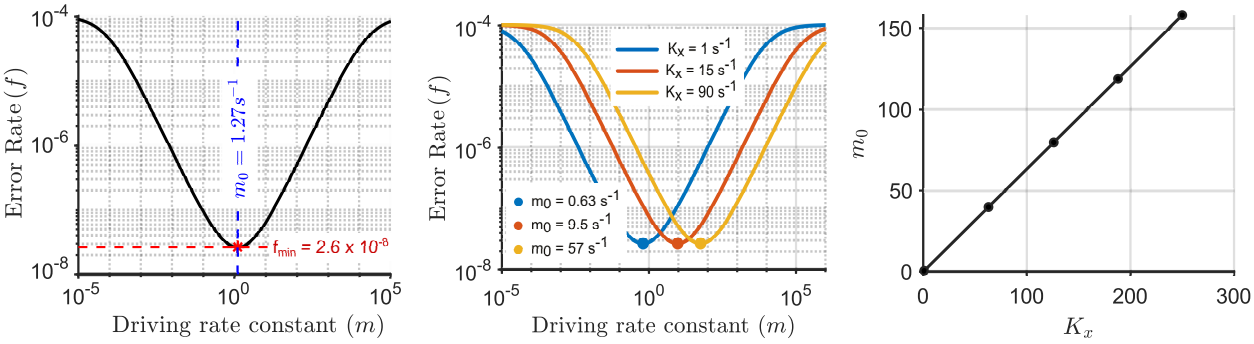


Figure 2: *(a) Interplay between the driving rate constant (m) and error rate (f) variables for the dissociation rate parameter $K_x = 2s^{-1}$, set by the environment. The red star marks the minimum value of error rate, reflecting the maximum fidelity achieved by the enzyme during proofreading. The associated driving rate constant denotes the optimal energy expenditure for maximal discrimination between correct and incorrect substrates. Saturation at both extremes (low and high m) reflects single-step discrimination ($f \approx 10^{-4}$), whereas at the optimum (red star), the error rate is reduced to $\sim 10^{-8}$, consistent with Hopfield's prediction that the two-step optimal error rate is the square of single-step discrimination. (b) Dependence of error rate on driving rate m across different dissociation rates (K_x), due to different environmental temperatures. The blue, orange, and yellow curves correspond to $K_x = 1s^{-1}$, $15s^{-1}, 90s^{-1}$, respectively. Although the minimum error rate remains constant, the optimal m_0 shifts to higher values as K_x increases due to increase in the environmental temperature, following Eq. 14. Blue, orange, and yellow dots mark the minima at $m_0 = 0.63s^{-1}$, $9.5s^{-1}$, $57s^{-1}$, respectively. (c) Relationship between K_x, dictated by the environment, and the optimal driving rate constant m_0, resulting from enzymatic adaptation. Increasing K_x at higher environmental temperatures alters the substrate-flux balance, necessitating higher m_0 values to restore optimal discrimination.*

The relationship between error rate (f) and the driving rate constant (m) is shown in Fig.2a. The red star denotes the optimal driving rate constant (m_0), corresponding to an optimum energy investment at which the error rate reaches its minimum. Deviations from this optimum increase error rate f, dividing the curve into two sub-optimal regimes. In the left sub-optimal regime ($m < m_0$), f decreases as the driving rate constant (m) increases, whereas in the right sub-optimal regime ($m > m_0$), f grows with further increase in driving rate constant beyond the optimum. At extreme m values, f saturates at $f_{\max} = \alpha$, as quantified in Eq. 13. This upper bound ($f \sim 10^{-4}$) is due to the discrimination between correct and incorrect substrates through the dissociation path of step 3 (single-step discrimination), since step 1 discrimination becomes ineffective. However, the error rate reaches its minimum ($f_{\min} \sim 10^{-8}$) at the optimum (m_0, red star), where both step 1 and step 3 contribute cooperatively (two-step discrimination) to error correction.

We also observe a systematic shift in the relationship between the error rate and the driving rate constant as the dissociation rate K_X changes, due to environmental temperature change, as shown in Fig. 2b. The upper and lower limits of the error rate stay the same as the dissociation rate K_X changes; however, the optimal driving rate constant (m_0) increases with the dissociation rate K_X, following Eq. 14 (see Fig. 2c), while the minimum error rate stays the same. This behavior can be understood in terms of the substrate flux along the enzymatic pathway. Changes in K_X alters the finely-tuned balance of the flux through the two discriminating steps (step 1 and step 3), thereby increasing the error rates. Increasing K_X while keeping m constant makes $[EX]$ and $[EY]$ dissociate more easily, which makes it harder to convert them to $[EX^*]$ and $[EY^*]$ and makes step 1 discrimination weaker. On the other hand, lowering K_X reduces the conversion rate of $[EX]$ and $[EY]$ to their constituent substrates, which again makes step 1 discrimination less accurate. To keep discrimination effective across both steps and make sure that wrong substrates stay entrapped for a long time, the driving constant m needs to be evolutionarily adjusted. Thus evolutionary optimization of the enzymatic energy expenditure, denoted by the variable m, enables preservation of replication fidelity f at the magnitude of 10^{-8}.

6.2 Adaptive Response to Environmental Perturbations

The optimized driving rate constant m_0 can be considered an evolutionary trait influenced by the kinetic parameters of genomic error correction mechanisms, such as K_X and K_Y. These kinetic parameters are not stable; they fluctuate with environmental conditions. For instance, temperature fluctuations alter dissociation rate constants K_X and K_Y in fig. 1, with higher temperatures generally accelerating the dissociation. In our framework, we model temperature variation as the environmental perturbation by adjusting the dissociation rate constants, particularly those of step 1 ($K_{X/Y}$), within the biochemical pathway. Because dissociation rates are environment-dependent (see Methods), the optimal driving rate constant m_0 required to minimize error is itself dynamic. To maintain a minimal error rate (on the order of 10^{-8} in DNA replication), enzymes must therefore recalibrate their driving rate constant under a changing environment. This selective pressure drives the evolutionary fine-tuning of the trait (m_0), ensuring that the driving rate constant remains optimally adapted to the prevailing environment.

To examine this process, we implement a mechanistic adaptive model *without imposing a predefined fitness function* (see *Evolutionary Framework section*). Starting from an initial population of 10^4 individuals, we expose the system to an abrupt temperature shift from $T_{\text{old}} = 293\text{K}$ to $T_{\text{new}} = 335\text{K}$. This perturbation is incorporated by varying the dissociation rate constant of step 1, from $K_{x,\text{old}} = 1\text{s}^{-1}$ to $K_{x,\text{new}} = 42\text{s}^{-1}$, following the Arrhenius relation in Eq. 9. The sudden variation increases the mean error rate of the population by about ten times its initial minimum value. This is because the enzyme cannot immediately readjust its driving rate to the new optimal value (from $m_{0,\text{old}} = 0.632\text{s}^{-1}$ to $m_{0,\text{new}} = 26.569\text{s}^{-1}$). This increased error rate causes genetic modifications in the coding region (assumed length 10^6 bp), which then shows up as changes in the enzyme's phenotype that could be beneficial or deleterious.

Mutations due to the ineffectual error correction create the genetic variation that fuels the adaptation process. The enzyme gradually lowers the high error rate and brings back functionality in the altered environment. The adaptation rate is regulated by the mutation rate (the average number of mutations per base pair per generation). Higher mutation rates drive the adaptation process faster by giving more options for selection, while lower mutation rates slow it down. As the error rate of the population becomes minimum, the mutational supply decreases, and enzymes are evolutionarily selected to function at the newly required optimal driving rate.

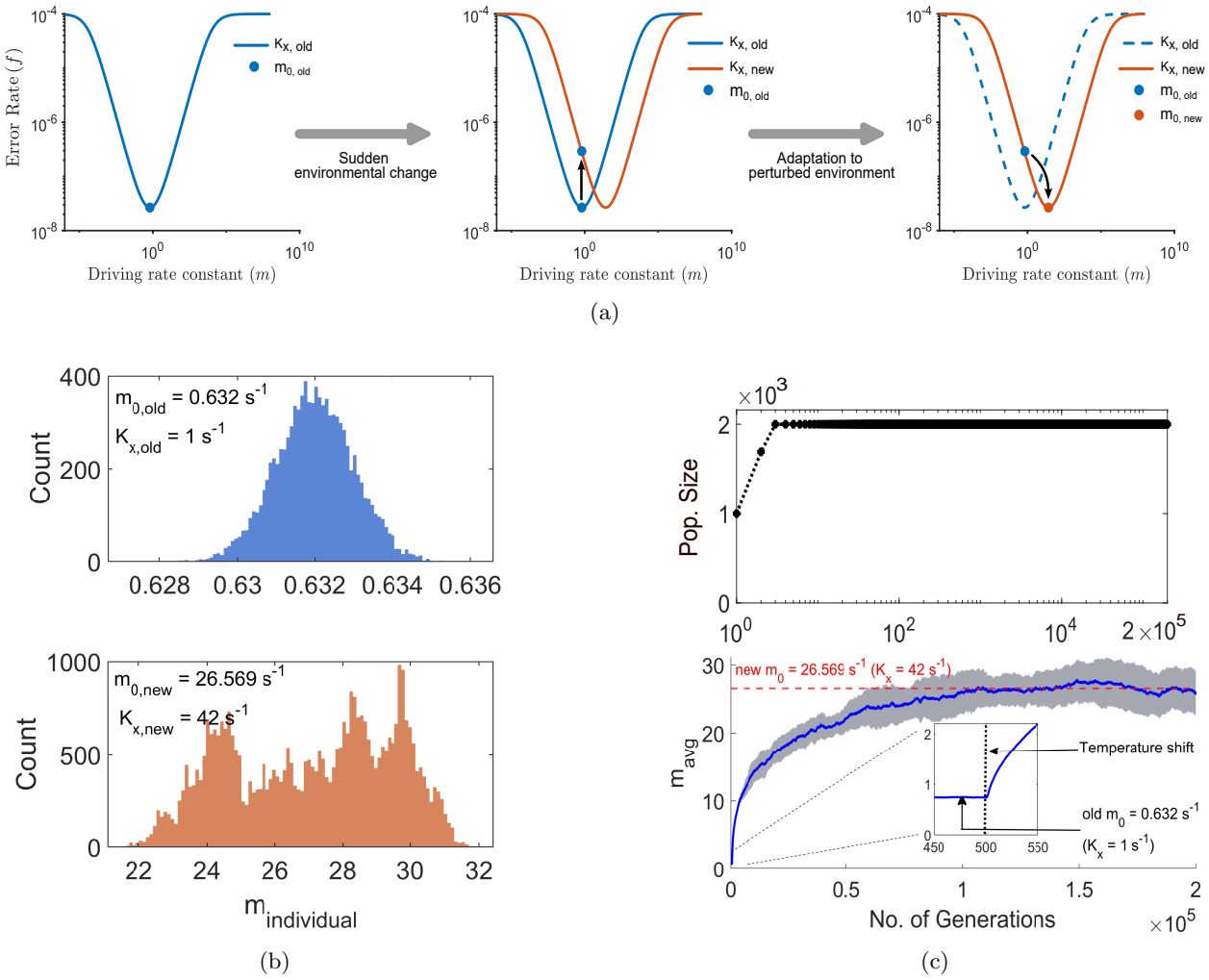


Figure 3: *The figure illustrates the adaptive dynamics of the enzyme to minimize the error rate in a perturbed environment. The proofreading enzymes should have an optimal driving rate constant tuned to the kinetic parameters dictated by prevailing conditions to achieve the minimum error rate. (a) The panels show the enzymatic adaptation due to an abrupt perturbation schematically. In the first panel, the blue dot on the blue curve marks the optimal driving rate constant and the corresponding value of the error rate has the minimum magnitude ($f_{\min} \approx 10^{-8}$) in the stable environmental condition ($K_{x,old} = 1\ s^{-1}$, $T_{old} = 293\ K$, $m_{0,old} = 0.632\ s^{-1}$). With a sudden environmental change ($K_{x,new} = 42\ s^{-1}$, $T_{new} = 335\ K$), the error rate increases about tenfold as the enzyme cannot immediately adjust to the new optimum due to sudden environmental perturbation, and remains at the old driving rate optimum, $m_{0,old} = 0.632\ s^{-1}$. This is shown in the second panel, where the blue dot shifts vertically from the blue curve to the orange curve (error rate vs. driving rate constant). The orange curve represents the perturbed environment ($K_{x,new} = 42\ s^{-1}$, $T_{new} = 335\ K$). Over successive generations, mutations drive adaptive evolution toward a new optimum, $m_{0,new} = 26.569\ s^{-1}$. The third panel depicts this transition, as the blue dot evolves into the orange dot on the orange curve, representing the adapted state of the enzyme. (b) Distribution of individual optimal driving rates ($m_{individual}$). The blue histogram represents the initial population of 10^4 individuals with mean $m_{0,old} = 0.632\ s^{-1}$. The orange histogram shows 4×10^4 individuals after 2×10^5 generations, with mean $m_{0,new} = 26.569\ s^{-1}$, indicating successful adaptation. (c) Evolutionary trajectory of adaptation. The blue curve (with 99.9% confidence interval) shows the mean driving rate constant over generations. The trajectory saturates at the new optimum (red horizontal line) without an explicit fitness function, reaching a steady state with minimal error rate in $\sim 10^5$ generations. The corresponding population size dynamics are shown in black (with population axis in logarithmic scale for visualization of the early rise).*

The enzymatic evolution, illustrated in Fig. 3a, shows how the error rate varies with environmental conditions. Before the environmental perturbation, the trait value of the enzyme, $m_{0,\text{old}} = 0.632\ \text{s}^{-1}$, is optimized for the environmentally determined parameter $K_{x,\text{old}} = 1\ \text{s}^{-1}$ at a temperature $T_{\text{old}} = 293$ K. The blue dot on the blue curve represents this state. When the enzyme's fidelity is disturbed due to a sudden environmental change, the old driving rate $m_{0,\text{old}} = 0.632\ \text{s}^{-1}$ of the enzyme ceases to be optimal for this new environment. The resultant increase in the error rate is shown by the blue dot moving to the orange curve corresponding to the perturbed environment ($K_{x,\text{new}} = 42\ \text{s}^{-1}$) in the above fig. 3a. This increased mutation rate drives evolutionary change. Adaptation moves the enzyme toward the new optimum ($m_{0,\text{new}} = 26.569\ \text{s}^{-1}$) over successive generations and restores high fidelity under the new environmental conditions. This is shown in the third panel by the movement of the orange dot towards the minimum of the orange curve.

The quantification of this adaptive process is demonstrated in Fig. 3c and Fig. 3b through Monte Carlo simulations of trait evolution under mutational bias *without an explicit fitness function*. The distribution of the driving rates of individuals within the initial population ($N = 10^4$, mean $m_{0,\text{old}}$) under the initial stable environment is shown in the blue histogram, while the orange histogram shows the distribution of the individual error rates of the adapted population in the perturbed environment ($N = 4 \times 10^4$, mean $m_{0,\text{new}}$) (Fig. 3b). Fig. 3c illustrates the evolutionary dynamics of this adaptation process, where the adaptation of the mean driving rate (m_{avg}) of the population across generations is depicted by the blue curve. The evolutionary trajectory, beginning from the old optimum before the temperature change, exhibits a convergence toward the new optimum after the temperature change, marked by the red horizontal line, reaching a steady state with minimum error rate after 10^5 generations. The corresponding population size dynamics is shown by the black curve, above.

We define the adaptation rate as the change in phenotype per generation, which is computed directly from simulations as the change in m_{avg} of the population per generation. As shown in Fig. 4, the adaptation rate shows an initial sharp rise reflecting the sudden error increase imposed by the temperature shift, followed by a gradual decline as the mutation rate decreases and the mean driving rate of the population converges toward the new optimum ($m_{0,\text{new}}$). This observation is in agreement with the core assumptions of Fisher's geometric model, in which Δm is taken to be uniform across the population within a generation and the mutational step size diminishes as the phenotype approaches its optimum. A comparison of evolutionary dynamics under uniform Δm versus individual-specific Δm is provided in Fig. S1 of the Supplementary Information.

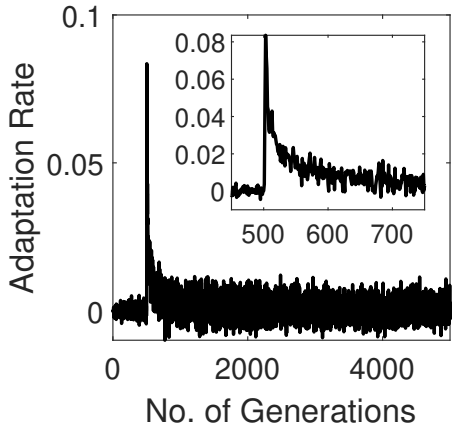


Figure 4: *Evolution of the adaptation rate, defined as the change in phenotype m per generation, across generations. The adaptation rate rises sharply in the early generations just after the perturbation at* 500^{th} *generation, reflecting the sudden increase in replication error following the environmental (temperature) shift. As the mean trait of the population approaches the new optimum, the adaptation rate progressively declines, driven by the reduced mutation rate at steady state.*

The imposed temperature shift temporarily increases error frequencies in replication and protein syn-

thesis, and errors generated in one process propagate into others, elevating organism-level mutation supply [63]. The resulting surge in mutational input produces rapid phenotypic change, corresponding to the high initial adaptation rate observed in Fig. 4. The adaptation rate decreases due to the decline in the mutational supply as the enzymes readjust their driving rate constant to the new optimum value ($m_{0,\mathrm{new}}$) and restore replication fidelity, causing the population to enter a state of evolutionary stasis. Subsequent environmental perturbations reproduce this cyclical pattern of rapid adaptation followed by relative stabilization, thereby generating alternating intervals of evolution and stasis that are broadly consistent with punctuated equilibrium–like dynamics [64]. These signatures of punctuated equilibrium are illustrated in Fig. S13.

6.3 Influence of Coding-Region Length and Population Size on Adaptive Dynamics

To examine how genome size influences evolutionary dynamics under changing environments, we here investigate the fate of populations differing only in the length of their coding region. Using the same evolutionary framework described earlier, we simulate populations with coding-region lengths of 10^3, 10^6, and 10^8 bp, each exposed to environmental perturbations that alter the optimal driving rate constant of the proofreading enzyme (Fig. 5). The perturbation was introduced at the 500^{th} generation by shifting the environmental temperature, which increases the replication error rate by displacing the error-correcting enzyme from its optimal configuration.

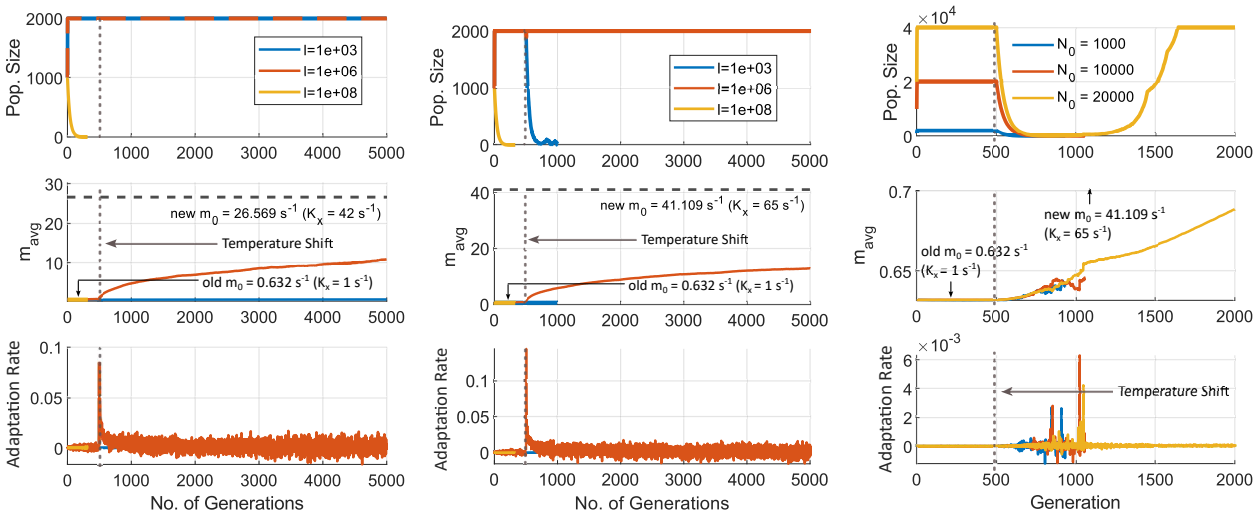


Figure 5: *The figure illustrates the evolutionary dynamics of populations with varying coding region lengths and initial sizes under environmental temperature perturbation. Each of the three panels shows population size, mean driving rate constant (m_{avg}) of the error-correcting enzymes, and adaptation rate.* ***(a)*** *Populations with initial size $N_0 = 10^3$ and coding-region lengths of 10^3, 10^6, and 10^8 bp (blue, orange, and yellow curves, respectively) subjected to a temperature shift of $\Delta T = 42$ K at generation 500. Populations with very long coding regions (10^8 bp) collapse prior to perturbation due to rapid accumulation of deleterious mutations. Intermediate-length populations (10^6 bp) retain sufficient mutational input to re-optimize efficiently after perturbation. Short coding-region populations (10^3 bp) survive the perturbation but adapt slowly because of limited variation.* ***(b)*** *Evolutionary dynamics of the same populations under a stronger perturbation of $\Delta T = 47.68$ K. Under this harsher shift, populations with short coding region length (10^3 bp) fail to persist, as the increased error rate overcomes their restricted adaptive potential. Intermediate-length populations (10^6 bp) continue to re-optimize, while very long coding regions (10^8 bp) collapse even before the perturbation.* ***(c)*** *Evolutionary dynamics of populations with a fixed coding-region length of 10^3 bp but different initial population sizes ($N_0 = 10^3$, 10^4, and 2×10^4; blue, orange, and yellow curves) under the strong perturbation ($\Delta T = 47.68$ K). Larger initial populations provide sufficient mutational input and buffer against drift, enabling recovery and re-optimization. When $N_0 = 2 \times 10^4$, the population successfully adapts despite its small coding-region length, whereas smaller populations collapse.*

Our results reveal a clear selection for specific coding lengths, disfavoring smaller and larger lengths and selecting an intermediate length, depending on the mutational rate and population size. Populations with very long coding regions (10^8 bp) accumulate deleterious mutations so rapidly that they collapse even before any perturbation occurs (*Muller's ratchet*), illustrated in Fig. 5a and Fig. 5b. The longer genomes have a higher probability of deleterious mutations, pointing to the fact that excessive mutational supply does not necessarily increase the adaptive potential. Instead, it speeds up mutational meltdown in small or moderately sized populations.

On the other hand, populations with intermediate length of the coding-region (10^6 bp) maintain low error rates under stable conditions and exhibit high adaptation rate under perturbation (Fig. 5a). The average driving rate constant m_{avg} of these populations evolves rapidly toward the new optimum value just after the environmental shift, which is shown as a transient sharp rise in the adaptation rate curve (3^{rd} panel of Fig. 5a). This adaptation indicates rapid and efficient acquisition of beneficial mutations. This coding-region length sets up a sufficient mutational input to search through adaptive variants while holding down the build-up of deleterious changes, thereby balancing stability and novelty, and allowing for a robust evolutionary response.

Populations with the shortest coding region (10^3 bp) show up with a distinct pattern of evolutionary dynamics. They remain viable under a temperature shift of 42 K, but evolve at an extremely slow pace. The small coding length produces very limited mutational variation, leading to an almost flat trajectory of the

average driving rate constant with negligible phenotypic change (Fig. 5a). When the perturbation ramps up to $\Delta T = 47.68$ K (Fig. 5b), the replication errors produced are not sufficient for the genetic variation necessary to survive such a large perturbation, and as a result, the population dies out.

Nevertheless, this short coding region becomes selectively favored in large populations, where a subset of individuals survives by maintaining error rates well below the threshold (f_{th}), effectively lowering their elimination probability (P_{kill}) and persisting long enough to accumulate beneficial mutations. These survivors gradually move away from the critical error boundary and adapt slowly toward the new optimal driving rate (Fig. 5c). The small coding region becomes selectively favored not because it inherently promotes adaptation, but because a sufficiently large population compensates for its limited mutational input. This demonstrates that slow, nearly neutral adaptation can be deleterious and lead to extinction in small populations, whereas the same trait can be naturally selected in large populations, where selection effectively amplifies even subtle fitness advantages. Thus, both coding region length and population size jointly determine evolutionary resilience and the likelihood of extinction under environmental stress.

Together, these results (including Fig. S7 and S9) demonstrate that the evolutionarily favored coding-region length is not fixed but emerges from the balance between mutation supply and mutational robustness, as modulated by population size. Very long coding regions fail because they overload the population with deleterious mutations; intermediate coding regions support rapid and stable adaptation; and very short coding regions survive only when large population sizes provide enough mutational input to enable evolutionary response. This interplay between coding region length, mutation rate, and population size directly reflects the competition between natural selection and genetic drift that shapes molecular evolution.

7 Discussion

In this study, we show how living systems dynamically regulate mutational supply, needed for adaptation, with robustness to mutations, to avoid error catastrophe, without foresight, that is, without the help of an explicit fitness function. Error-correcting systems maintain robustness by lowering the mutational input. However, environmental perturbations disturb this balance by increasing the mutational supply through destabilization of the error-correcting systems, which helps accelerate adaptation to the said environmental perturbation. Once adapted to the new environment, the error-correcting system (polymerase) with altered driving rates again achieves low error rates, thereby regaining robustness.

We model this dynamic regulation of mutational supply, where the error-correcting enzymes autonomously increase the mutation rates when the environment changes, and decrease the rates during stable times, by explicitly linking replication fidelity to environmental perturbation through kinetic proofreading. In the kinetic proofreading scheme of Hopfield, environmental perturbations, such as temperature fluctuations, change the kinetic parameters that govern substrate binding, dissociation, and discrimination, thereby pushing the enzyme away from its optimal regime, resulting in increased mutation rates. This increased mutation rate gives rise to abundant genetic variation, allowing for rapid adaptive change under the new environmental conditions. This adaptive change happens at the genomic locations encoding the enzyme as well, resulting in the enzyme evolving toward a new optimal state, which minimizes the error rate in the new perturbed regime. As a result, the mutation rate declines and the system settles into evolutionary stasis. This pattern of sudden perturbation, rapid adaptation, and subsequent convergence to evolutionary stasis is consonant with the dynamics of punctuated equilibrium observed across microbial, molecular, and phenotypic evolution [65, 66, 67, 68, 69].

The balance between mutational supply and robustness is also regulated by the length of the coding region and the population size. The larger the coding region, larger the mutational target space, which therefore has a higher probability of generating mutations, both beneficial and deleterious, whereas, smaller coding regions have limited adaptive potential. In finite populations, the length of the coding region is evolutionarily chosen such that rapid adaptation occurs when the environment changes, while avoiding catastrophic error accumulation. Our results in Fig. 3 reflect this evolutionary dynamics clearly, where populations with very long coding regions build up deleterious mutations so rapidly that they die out even before any environmental perturbation sets in. The populations with intermediate-length coding regions maintain low error rates under stable conditions and adjust to perturbations efficiently. Populations with very short coding regions move along evolutionary trajectories at an extremely slow, nearly neutral pace, as

their limited mutational input prevents rapid adaptation. When an environmental perturbation is applied, mutation rates increase, and these small-coding-region populations die out unless they are backed up by sufficiently large population sizes, which provides the necessary beneficial mutations required for adaptation (see Fig. 5). Such population dynamics behavior highlights the role of the population size in competition between natural selection and drift, following the drift barrier hypothesis [8, 70].

This mechanism can be used to explain the observed cases of rapid evolutionary change, such as the emergence of antibiotic resistance, adaptive melanism, and phenotypic shifts in invasive or climate-stressed species. Because environmental perturbations directly influence molecular fidelity mechanisms, they can modulate evolutionary rates across ecological and geological timescales alike. Thus, this framework provides not only a mechanistic understanding of how selection, drift, and molecular fidelity interact but also a predictive foundation for assessing the adaptive capacity and long-term resilience of evolving populations under fluctuating environments.

Author Contributions

HS conceived the idea. SB, PS, and KG, together with HS, researched and wrote the article.

Declaration of Interests

The authors declare no competing interests.

Funding

Support for this work was provided by the Science & Engineering Research Board (SERB), Department of Science and Technology (DST), India, through a Core Research Grant with file no. CRG/2020/003555 and a MATRICS grant with file no. MTR/2022/000086.

Data Availability

All data necessary to reproduce the results are contained in the manuscript, and the associated MATLAB codes are available at: `https://github.com/Sashik05/kinetic-proofreading-mutation-robustness`.

References

[1] Thomas BL Kirkwood. "Evolution of ageing". In: *Nature* 270.5635 (1977), pp. 301–304.

[2] Peter J Bowler. "Evolution". In: *The History of Science and Religion in the Western Tradition*. Routledge, 2003, pp. 549–557.

[3] John Maynard Smith. *The theory of evolution*. Cambridge University Press, 1993.

[4] Motoo Kimura. "The neutral theory of molecular evolution". In: *Scientific American* 241.5 (1979), pp. 98–129.

[5] John H Gillespie. *The causes of molecular evolution*. Vol. 2. Oxford University Press, 1991.

[6] Andreas Wagner. "Robustness, evolvability, and neutrality". In: *FEBS letters* 579.8 (2005), pp. 1772–1778.

[7] Andreas Wagner. "Robustness and evolvability: a paradox resolved". In: *Proceedings of the Royal Society B: Biological Sciences* 275.1630 (2008), pp. 91–100.

[8] Michael Lynch. "Evolution of the mutation rate". In: *TRENDS in Genetics* 26.8 (2010), pp. 345–352.

[9] John W Drake. "A constant rate of spontaneous mutation in DNA-based microbes." In: *Proceedings of the National Academy of Sciences* 88.16 (1991), pp. 7160–7164.

[10] Andreas Wagner. "The role of robustness in phenotypic adaptation and innovation". In: *Proceedings of the Royal Society B: Biological Sciences* 279.1732 (2012), pp. 1249–1258.

[11] John J Hopfield. "Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity". In: *Proceedings of the National Academy of Sciences* 71.10 (1974), pp. 4135–4139.

[12] Jacques Ninio. "Kinetic amplification of enzyme discrimination". In: *Biochimie* 57.5 (1975), pp. 587–595.

[13] Errol C Friedberg et al. *DNA repair and mutagenesis*. American Society for Microbiology Press, 2005.

[14] Alanna R Kaplan and Peter M Glazer. "Impact of hypoxia on DNA repair and genome integrity". In: *Mutagenesis* 35.1 (2020), pp. 61–68.

[15] Nimrat Chatterjee and Graham C Walker. "Mechanisms of DNA damage, repair, and mutagenesis". In: *Environmental and molecular mutagenesis* 58.5 (2017), pp. 235–263.

[16] Jean Cadet et al. "Radiation-induced DNA damage: formation, measurement, and biochemical features". In: *Journal of Environmental Pathology, Toxicology and Oncology* 23.1 (2004).

[17] Charlotte Auerbach. "Chemical mutagenesis". In: *Biological reviews* 24.3 (1949), pp. 355–391.

[18] Mark D Evans and Marcus S Cooke. *Oxidative damage to nucleic acids*. Springer Science & Business Media, 2009.

[19] Patricia L Foster. "Stress responses and genetic variation in bacteria". In: *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 569.1-2 (2005), pp. 3–11.

[20] Patricia L Foster. "Stress-induced mutagenesis in bacteria". In: *Critical reviews in biochemistry and molecular biology* 42.5 (2007), pp. 373–397.

[21] Evelyn M Witkin. "Ultraviolet mutagenesis and inducible DNA repair in Escherichia coli". In: *Bacteriological reviews* 40.4 (1976), pp. 869–907.

[22] Michael Lynch. "The origins of eukaryotic gene structure". In: *Molecular biology and evolution* 23.2 (2006), pp. 450–468.

[23] Michael Lynch and Bruce Walsh. *The origins of genome architecture*. Vol. 98. Sinauer associates Sunderland, MA, 2007.

[24] Eduardo PC Rocha and Antoine Danchin. "Gene essentiality determines chromosome organisation in bacteria". In: *Nucleic acids research* 31.22 (2003), pp. 6570–6577.

[25] You S Rao et al. "Selection for the compactness of highly expressed genes in Gallus gallus". In: *Biology Direct* 5.1 (2010), p. 35.

[26] Tomoko Ohta. "Slightly deleterious mutant substitutions in evolution". In: *Nature* 246.5428 (1973), pp. 96–98.

[27] Tomoko Ohta. "The nearly neutral theory of molecular evolution". In: *Annual review of ecology and systematics* (1992), pp. 263–286.

[28] Brian Charlesworth and Deborah Charlesworth. *Elements of Evolutionary Genetics*. Greenwoord Village, Colorado, USA: Roberts and Company Publishers, 2010. ISBN: 9780981525.

[29] Way Sung et al. "Drift-barrier hypothesis and mutation-rate evolution". In: *Proceedings of the National Academy of Sciences* 109.45 (2012), pp. 18488–18492.

[30] Calvin P Sjaarda et al. "Temporal dynamics and evolution of SARS-CoV-2 demonstrate the necessity of ongoing viral genome sequencing in Ontario, Canada". In: *Msphere* 6.3 (2021), pp. 10–1128.

[31] José M Cuevas et al. "Experimental evolution and population genetics of RNA viruses". In: *Open Evol. J* 3 (2009), pp. 9–16.

[32] Jithesh Kottur and Deepak T Nair. "Pyrophosphate hydrolysis is an intrinsic and critical step of the DNA synthesis reaction". In: *Nucleic Acids Research* 46.12 (2018), pp. 5875–5885.

[33] Erik Johansson and Nicholas Dixon. "Replicative DNA polymerases". In: *Cold Spring Harbor perspectives in biology* 5.6 (2013), a012799.

[34] Jessica LE Wimmer, Karl Kleinermanns, and William F Martin. "Pyrophosphate and irreversibility in evolution, or why PPi is not an energy currency and why nature chose triphosphates". In: *Frontiers in microbiology* 12 (2021), p. 759359.

[35] Kinshuk Banerjee, Anatoly B Kolomeisky, and Oleg A Igoshin. "Elucidating interplay of speed and accuracy in biological error correction". In: *Proceedings of the National Academy of Sciences* 114.20 (2017), pp. 5183–5188.

[36] Kenneth A Johnson. "Conformational coupling in DNA polymerase fidelity". In: *Annual review of biochemistry* 62.1 (1993), pp. 685–713.

[37] Manfred Eigen. "Selforganization of matter and the evolution of biological macromolecules". In: *Naturwissenschaften* 58.10 (1971), pp. 465–523.

[38] Mary E Craig, Donald M Crothers, and Paul Doty. "Relaxation kinetics of dimer formation by self complementary oligonucleotides". In: *Journal of molecular biology* 62.2 (1971), pp. 383–401.

[39] Kenneth A Johnson. "The kinetic and chemical mechanism of high-fidelity DNA polymerases". In: *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1804.5 (2010), pp. 1041–1048.

[40] Michelle P Roettger, Marina Bakhtina, and Ming-Daw Tsai. "Mismatched and matched dNTP incorporation by DNA polymerase β proceed via analogous kinetic pathways". In: *Biochemistry* 47.37 (2008), pp. 9718–9727.

[41] Arvind Murugan, David A Huse, and Stanislas Leibler. "Speed, dissipation, and error in kinetic proofreading". In: *Proceedings of the National Academy of Sciences* 109.30 (2012), pp. 12034–12039.

[42] Linda K Clayton et al. "Error induction and correction by mutant and wild type T4 DNA polymerases. Kinetic error discrimination mechanisms." In: *Journal of Biological Chemistry* 254.6 (1979), pp. 1902–1912.

[43] Daniel E Koshland Jr. "Application of a theory of enzyme specificity to protein synthesis". In: *Proceedings of the National Academy of Sciences* 44.2 (1958), pp. 98–104.

[44] Richard Wolfenden. "Analog approaches to the structure of the transition state in enzyme reactions". In: *Accounts of Chemical Research* 5.1 (1972), pp. 10–18.

[45] C Blomberg, M Ehrenberg, and CG Kurland. "Free-energy dissipation constrainst on the accuracy of enzymatic selections". In: *Quarterly Reviews of Biophysics* 13.2 (1980), pp. 231–254.

[46] Pankaj Mehta, Alex H Lang, and David J Schwab. "Landauer in the age of synthetic biology: energy consumption and information processing in biochemical networks". In: *Journal of Statistical Physics* 162 (2016), pp. 1153–1166.

[47] CG Kurland and Måns Ehrenberg. "Growth-optimizing accuracy of gene expression". In: *Annual review of biophysics and biophysical chemistry* 16.1 (1987), pp. 291–317.

[48] Keith J Laidler. "The development of the Arrhenius equation". In: *Journal of chemical Education* 61.6 (1984), p. 494.

[49] Joseph Crapse et al. "Evaluating the Arrhenius equation for developmental processes". In: *Molecular systems biology* 17.8 (2021), e9895.

[50] Rolf R Freter and Michael A Savageau. "Proofreading systems of multiple stages for improved accuracy of biological discrimination". In: *Journal of theoretical biology* 85.1 (1980), pp. 99–123.

[51] M Ehrenberg and C Blomberg. "Thermodynamic constraints on kinetic proofreading in biosynthetic pathways". In: *Biophysical journal* 31.3 (1980), pp. 333–358.

[52] Stefan Howorka et al. "Kinetics of duplex formation for individual DNA strands within a single protein nanopore". In: *Proceedings of the National Academy of Sciences* 98.23 (2001), pp. 12996–13001.

[53] Isaac Wong, Smita S Patel, and Kenneth A Johnson. "An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics". In: *Biochemistry* 30.2 (1991), pp. 526–537.

[54] H Allen Orr. "Theories of adaptation: what they do and don't say". In: *Genetica* 123.1 (2005), pp. 3–13.

[55] H Allen Orr and Jerry A Coyne. "The genetics of adaptation: a reassessment". In: *The American Naturalist* 140.5 (1992), pp. 725–742.

[56] Ronald Aylmer Fisher. *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press, 1999.

[57] Ern Loh, Jesse J Salk, and Lawrence A Loeb. "Optimization of DNA polymerase mutation rates during bacterial evolution". In: *Proceedings of the National Academy of Sciences* 107.3 (2010), pp. 1154–1159.

[58] John H Gillespie. *Population genetics: a concise guide*. JHU press, 2004.

[59] Reinhard Bürger. *The mathematical theory of selection, recombination, and mutation*. John Wiley & Sons, 2000.

[60] Pablo Sartori and Simone Pigolotti. "Thermodynamics of error correction". In: *Physical Review X* 5.4 (2015), p. 041039.

[61] Lawrence A Loeb and Thomas A Kunkel. "Fidelity of DNA synthesis". In: *Annual review of biochemistry* 51.1 (1982), pp. 429–457.

[62] Thomas A Kunkel and Katarzyna Bebenek. "DNA replication fidelity". In: *Annual review of biochemistry* 69.1 (2000), pp. 497–529.

[63] D Allan Drummond and Claus O Wilke. "The evolutionary consequences of erroneous protein synthesis". In: *Nature Reviews Genetics* 10.10 (2009), pp. 715–724.

[64] Stephen Jay Gould and Niles Eldredge. "Punctuated equilibria: the tempo and mode of evolution reconsidered". In: *Paleobiology* 3.2 (1977), pp. 115–151.

[65] Santiago F Elena, Vaughn S Cooper, and Richard E Lenski. "Punctuated evolution caused by selection of rare beneficial mutations". In: *Science* 272.5269 (1996), pp. 1802–1804.

[66] Paul D Sniegowski, Philip J Gerrish, and Richard E Lenski. "Evolution of high mutation rates in experimental populations of E. coli". In: *Nature* 387.6634 (1997), pp. 703–705.

[67] Stuart T Nichol, Joan E Rowe, and Walter M Fitch. "Punctuated equilibrium and positive Darwinian evolution in vesicular stomatitis virus." In: *Proceedings of the National Academy of Sciences* 90.22 (1993), pp. 10424–10428.

[68] Lydia R Heasley, Nadia MV Sampaio, and Juan Lucas Argueso. "Systemic and rapid restructuring of the genome: a new perspective on punctuated equilibrium". In: *Current genetics* 67 (2021), pp. 57–63.

[69] Niles Eldredge. *Time frames: the evolution of punctuated equilibria*. Vol. 1001. Princeton University Press, 2014.

[70] Michael Lynch. "Mutation and human exceptionalism: Our future genetic load". In: *Genetics* 202.3 (2016), pp. 869–875.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf