

Reviewed Preprint

Revised by authors after peer review.

About eLife's process

Reviewed preprint version 2

April 4, 2024 (this version)

Reviewed preprint version 1

January 10, 2024

Posted to preprint server

November 6, 2023

Sent for peer review

November 2, 2023

Stochastic parabolic growth promotes coexistence and a relaxed error threshold in RNA-like replicator populations

Mátyás Paczkó, Eörs Szathmáry , András Szilágyi

HUN-REN Centre for Ecological Research, Institute of Evolution, Budapest, Hungary • Doctoral School of Biology, Institute of Biology, ELTE Eötvös Loránd University, Budapest, Hungary • Center for the Conceptual Foundations of Science, Parmenides Foundation, Pöcking, Germany • Department of Plant Systematics, Ecology and Theoretical Biology, Eötvös Loránd University, Budapest, Hungary

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

The RNA world hypothesis proposes that during the early evolution of life, primordial genomes of the first self-propagating evolutionary units existed in the form of RNA-like polymers. Autonomous, non-enzymatic and sustained replication of such information carriers presents a problem, because product formation and hybridization between template and copy strands reduces replication speed. Kinetics of growth is then parabolic with the benefit of entailing competitive coexistence, thereby maintaining diversity. Here, we test the information-maintaining ability of parabolic growth in stochastic multispecies population models under the constraints of constant total population size and chemostat conditions. We find that large population sizes and small differences in the replication rates favor the stable coexistence of the vast majority of replicator species (“genes”), while the error-threshold problem is alleviated relative to exponential amplification. In addition, sequence properties (GC content) and the strength of resource competition mediated by the rate of resource inflow determine the number of coexisting variants, suggesting that fluctuations in building block availability favored repeated cycles of exploration and exploitation. Stochastic parabolic growth could thus have played a pivotal role in preserving viable sequences generated by random abiotic synthesis and providing diverse genetic raw material to the early evolution of functional ribozymes.

eLife assessment

This study provides a **valuable** theoretical exploration of non-enzymatic sustained replication of RNA systems, in the parabolic growth regime of the evolution of putative primordial replicators. It provides **convincing** evidence that parabolic growth mitigates the error threshold catastrophe, thus demonstrating another way in which this regime contributes to the maintenance of genetic diversity. The findings shed light on relevant evolutionary regimes of primordial replicators, with potential applicability to our understanding of the origin of life.

Introduction

Current knowledge of nucleotide chemistry as well as a large body of indirect evidence from recent organisms support the RNA world hypothesis; a concept that once RNA fulfilled the role of an evolvable primordial informational polymer and biochemical reaction catalyst at the same time (Woese, 1965 [↗](#); Crick, 1968 [↗](#); Orgel, 1968 [↗](#), 1992 [↗](#); Gilbert, 1986 [↗](#); Benner et al., 1989 [↗](#); Joyce, 2002 [↗](#); Lee et al., 2023 [↗](#)). However, the initial weaknesses of the original RNA world hypothesis, such as the lack of a reliable replication mechanism and the consequential loss of heritable information (Bernhardt, 2012 [↗](#); Szostak, 2012 [↗](#); Le Vay & Mutschler, 2019 [↗](#)), have prompted scientists studying the origin of life to devise a wide variety of physicochemically refined models of the RNA world. A potentially promising suggestion is that prior to template replication of complex polymers catalyzed by an RNA polymerase ribozyme (i.e., replicase), genetic information and catalytic functions were initially distributed among short sequence modules that could occasionally be ligated to increase molecular complexity in a stepwise manner (Vlassov et al., 2005 [↗](#); Manrubia & Briones, 2007 [↗](#); Briones et al., 2009 [↗](#)). Replication of these short sequence modules (perhaps with some catalytic properties) could have occurred in the absence of a replicase ribozyme, driven by some template directed, non-enzymatic replication mechanism (von Kiedrowski, 1986 [↗](#); Zielinski & Orgel, 1987 [↗](#); Patzke & von Kiedrowski, 2007 [↗](#); Leu et al., 2011 [↗](#), 2013 [↗](#)). Thus, templated, enzyme-free replication in pools of nucleic acid oligomers that resulted from random abiotic polymerization may have contributed to the emergence of structured oligonucleotides during prebiotic evolution (Le Vay & Mutschler, 2019 [↗](#); Rosenberger et al., 2021 [↗](#)). For example, it has been experimentally demonstrated that activated oligonucleotides can act as catalysts for non-enzymatic replication of RNA containing all four nucleotides, with the fidelity sufficient to sustain a genome size large enough to encode active ribozymes (Prywes et al., 2016 [↗](#)).

If enzyme-free replication of oligomers with a high degree of sequence variability was indeed attainable from prebiotic chemistry, then it becomes a vital issue how a critical level of diversity of the associated genetic information (Szostak, 2011 [↗](#)) could be preserved under competition among distinct replicator types with different competitive abilities (e.g., replicabilities) (Eigen, 1971 [↗](#); Szathmáry, 1991 [↗](#); Maynard Smith & Szathmáry, 1995 [↗](#); Jiménez et al., 2013 [↗](#); Ruiz-Mirazo et al., 2017 [↗](#)). This recognition calls for elucidation of the coexistence mechanisms that could have alleviated the competition among independently replicating information carrying modules. For instance, parabolic growth dynamics, a kinetic behavior observed in non-enzymatic self-replicating systems (von Kiedrowski, 1986 [↗](#)), has been proposed as an ideal candidate mechanism to sustain a large amount of prebiotic genetic information (Szathmáry & Gladkih, 1989 [↗](#); Scheuring & Szathmáry, 2001 [↗](#)). In this kinetics, the growth order p is equal or close to 0.5 (i.e., the dynamics is sub-exponential) because increased stability of the template-copy complex (rate of association divided by dissociation) promotes parabolic growth (von Kiedrowski et al., 1991 [↗](#); von Kiedrowski & Szathmáry, 2001 [↗](#)). This is in sharp contrast to exponential growth, where $p = 1$ implies maximum dissociation that allows for fast autocatalysis (Tjivikua et al., 1990 [↗](#); von Kiedrowski et al., 1991 [↗](#)). Thus, replicating individuals in populations displaying parabolic growth kinetics are inherently prone to self-inhibition by duplex formation; a feature that can efficiently damp competition and therefore promoting coexistence. Indeed, parabolic population growth can sustain an unlimited number of competing replicator species in the infinite population size limit (Szathmáry & Gladkih, 1989 [↗](#); Varga & Szathmáry, 1997 [↗](#)). However, molecular evolution of the Darwinian type is thought to necessitate exponential, rather than parabolic amplification, because in the latter case robust coexistence of the competing replicator species generally limits the efficiency of natural selection (Sievers & von Kiedrowski, 1998 [↗](#); Szilágyi et al., 2017 [↗](#); Strazewski, 2019 [↗](#)). Therefore, a major current focus in models of parabolic replication is how to incorporate Darwinian selection into its dynamics (Lifson & Lifson, 1999 [↗](#); Scheuring & Szathmáry, 2001 [↗](#); von Kiedrowski & Szathmáry, 2001 [↗](#)).

While the latter models considerably improved the parabolic replication model framework in terms of its potential to incorporate evolvability, they did not account for the copying error threshold of their respective systems; i.e., an error rate of the replication mechanism at which the system's ability to stably propagate genetic information from generation to generation is prohibited, resulting in irreversible loss of sequence-coded information (Eigen, 1971 [↗](#); Joyce, 2002 [↗](#); Szostak, 2012 [↗](#)). The possibility that high mutation rates could lead to such an “error catastrophe” is one of the major caveats surrounding the RNA world hypothesis, given that the first RNA replication mechanisms have probably been inherently error-prone (Eigen, 1971 [↗](#); Manrubia & Briones, 2007 [↗](#); Leu et al., 2011 [↗](#)).

Moreover, practically no study is known to have treated parabolic replication so that physicochemical details and ecological constraints, together with stochastic effects (that supposedly also have considerable impact on finite systems), would have been taken into consideration. The present study aims to remedy these deficiencies by quantifying the heritable information-maintaining potential of finite populations of unlinked, RNA-like template replicators displaying parabolic growth modelled as stochastic dynamics. We represent the replicator sequences individually, and copying error as well as energetic constraints in strand separation are also considered explicitly. Using a constant population model version of this framework, we first investigate how the diversity-maintaining mechanisms operate in the parabolic regime in a finite population of replicators. We further demonstrate that parabolic coexistence is resistant to mutation rates assumed for template-directed, enzyme-free replication, thereby proposing a simple biochemical mechanism to relax the error threshold in hypothetical information storage systems of the RNA world. In order for our stochastic simulation results to be comparable with the theoretical error threshold of parabolic dynamics, we analytically calculate the equilibrium master (i.e., fittest sequence) concentrations as a function of the copying fidelity and the reproductive superiority of the master sequences relative to the mutants. We also compare the error threshold in case of parabolic dynamics to the case of the exponential dynamics. Finally, in order to validate that Darwinian selection by competitive exclusion is also feasible in this framework we consider a chemostat model of parabolic replication, in which monomer building block resources are additionally taken into account. Using this approach, we demonstrate that increased competition, mediated by a decrease in the resource influx rate, can induce a switch from the coexistence ensured by parabolic dynamics to survival of the fittest.

Methods

Parabolic replicator model framework

We consider populations of oligomer molecules based on nucleotide-like monomer constituents (to preserve generality for potential informational polymers other than RNA). Individual molecules are represented by their primary (monomeric sequence) structures and are assumed undergo template-directed, non-enzymatic replication. This replication mode converts a single-stranded plus-, or minus-strand template into a template-copy duplex. Strand separation rates of these duplexes are relatively low compared to strand association (e.g., von Kiedrowski et al., 1991 [↗](#)). Therefore, replication cycles are strongly inhibited and as a result, the population dynamics displays parabolic characteristics. We define a set of replicators with the highest fitness relative to all mutant genotypes and refer to them as “master” types (cf. Eigen, 1971 [↗](#)), representing the genetic information to be maintained. To analyze the diversity-maintaining ability of parabolic dynamics in a competitive setting, the master replicator types are characterized by different replicabilities. This assumption reflects the fact that, at least in case of RNA, the rate of copying of short templates by polymerization of activated monomers inherently varies with template composition (Szostak, 2012 [↗](#)). In order to make our model framework to be more suitable for focusing on different aspects of parabolic dynamics, we considered two different model versions of parabolic replication (**Figure 1A** [↗](#)).

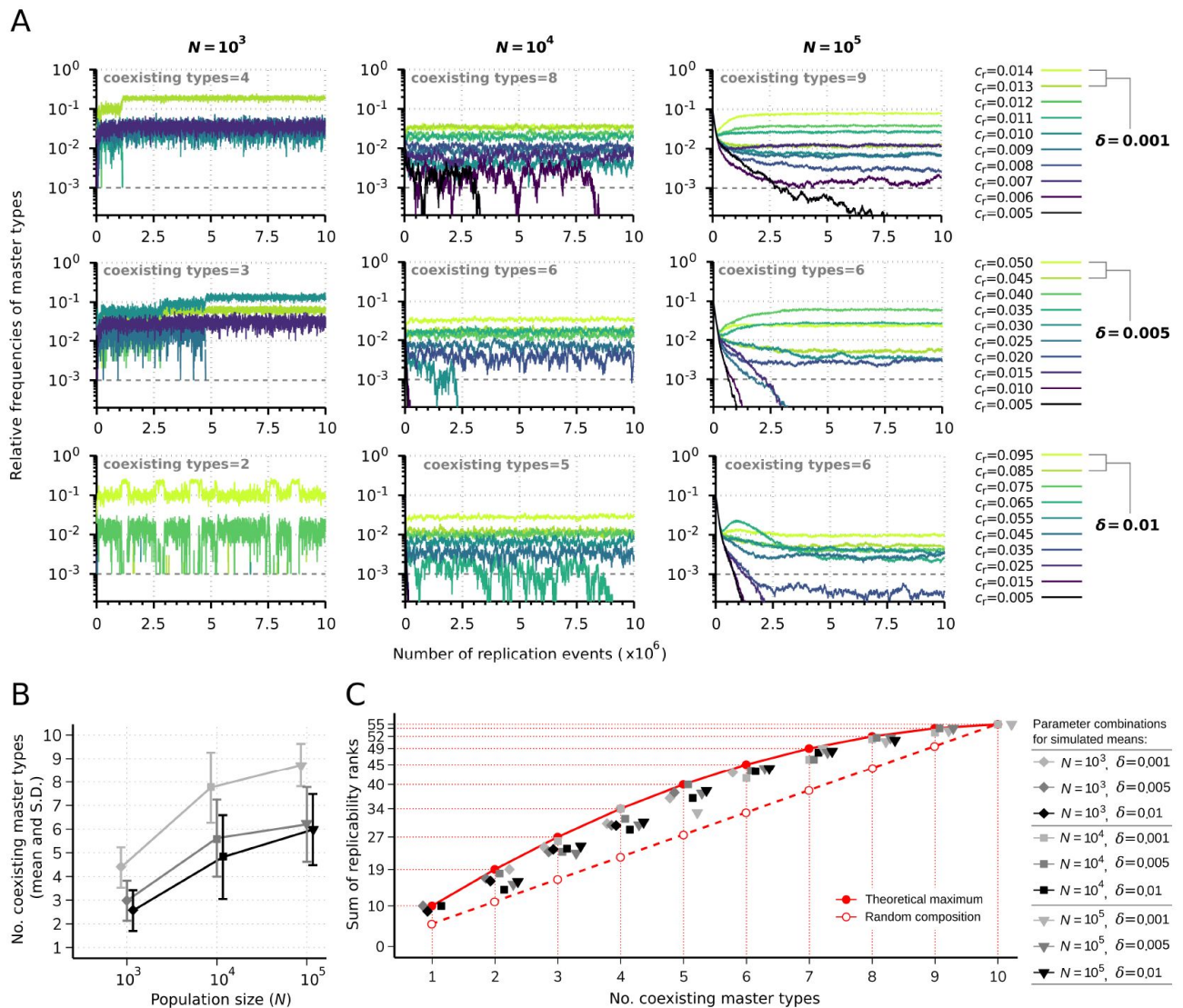


Figure 2.

Replicator coexistence in the constant population model. **(A)** Time series of the relative frequencies of master types. Panels corresponding to different rows demonstrate distinct replicability distances (δ), panels corresponding to different columns demonstrate distinct population sizes (N). Line colors indicate different replicabilities of the master types (as explained in the legends). The cut-off relative frequency for survival is indicated by horizontal dashed lines. **(B)** Mean and standard deviation of the number of sustained master types as the function of N and δ calculated from 50 independent runs (the corresponding legend is shown next to panel **C**). **(C)** Sum of replicability ranks of the survived master types at different number of coexisting types, corresponding to the simulation results shown in **(B)**. The red curve denotes the maximum of the sum of replicability ranks ($\rho_{\max}$), the dashed red line shows the expected values of the sum of replicability ranks when the surviving master types are random (ρ_{rand}) with respect to their replicabilities. The default parameter set (see [Table 1](#)) was used, unless otherwise indicated.

Constant population model of parabolic replication

In the constant population model (**Figure 1A**), there is no exchange of materials with the environment (monomer resources are not represented explicitly) and a constant total population size is ensured by implementing a Moran process into the replication mechanism (Moran, 1958). Accordingly, each newly synthesized strand replaces a sequence randomly chosen from the population (that can be either in simplex or duplex state; in the latter case the population size decreases by one until the next replication). The dynamics is based on the following three reaction categories: replication of a single strand S_i (Equation 1A); association of a single strand S_i with another strand S_j (Equation 2A), and dissociation of a double strand $S_i S_j$ (Equation 3A). The rate constants corresponding to these three reactions are c_r , c_a , c_d , respectively. Mutation events are taken into account during replication. These simple assumptions are in line with the kinetics of template-directed enzyme-free nucleic acid replication (for an experimental system, see e.g., Sievers & von Kiedrowski, 1998). The $n \cdot R$ expression in Equation 1A and Equation 1B denotes the amount of monomer building blocks needed for replication.



Chemostat model of parabolic replication

In case of the chemostat model (**Figure 1B**), a constant inflow of monomer building blocks is assumed to supply the synthesis of the new strands, i.e., the dynamics describes an open system (Schuster, 2011). This inflow component, together with type-specific replicability rate constants and a decay component on the replicators as well as outflow of the excess production, determines the time-dependent total number of replicators during the dynamics. The three basic reaction categories are the same as in the constant population model, namely: replication (Equation 1B), association (Equation 2B), and dissociation (Equation 3B), however, the dynamics of the resources is explicitly represented. The common inflow rate of all the four RNA nucleobases is c_{in} (Equation 4B). In contrast to the constant population model, here we assume decay of both single and double strands, the decay rates are denoted by $c_{\delta_{ss}}$ and $c_{\delta_{ds}}$, respectively, see Equation 5B and Equation 6B. The type independent outflow of molecular species is characterized by c_{out} (Equation 7B).



Parameter	Definition	Values or range (default)
N^*	Total replicator population size	$\{10^3; 10^4; \mathbf{10^5}\}$
δ^*	Replicability distance	$\{0.001; \mathbf{0.005}; 0.01\}$
p_{mut}^*	Per base mutation rate	$\{\mathbf{0.01}; 0.05; 0.1; 0.15\}$
$f^{*\ddagger}$	Duplex decay factor	$\{\mathbf{0.01}; 0.1; 0.5; 1\}$
$c_{\text{in}}^{*\ddagger}$	Kinetic constant for resource influx	$\{0.01; 0.1; 1\}$
$c_{\delta_{\text{ss}}}^{\ddagger}$	Kinetic constant for simplex decay reactions	10^{-4}
$c_{\delta_{\text{ds}}}^{\ddagger}$	Kinetic constant for duplex decay reactions	$[1.15 \cdot 10^{-8}, 6.4 \cdot 10^{-5}]$
$c_{\text{out}}^{\ddagger}$	Kinetic constant for outflux of molecular species	$[1.75 \cdot 10^{-6}, \mathbf{10^{-5}}]$
c_{r}	Kinetic constant for replication reactions	$[0.005, 0.095]$
c_{a}	Kinetic constant for association reactions	1
c_{d}	Kinetic constant for dissociation reactions	$[0.01, 0.6309]$
T	Number of master replicator types	10
L	Sequence length	10
ε	Error factor for the Hamming distance $(\Delta_{ij}) \in \{2; 1; 0\}$ mutant classes	$\{0.05; 0.2; 1\}$
ϕ_{min}	Minimum master replicability	0.005
κ	Scale factor for duplex dissociation probability distribution	0.2302

*: model parameters involved in screening, $\ddagger$: parameter used only in the chemostat system model

Table 1.

Model parameters

In both model versions, a population consists of replicators with equal length $L = 10$, composed of A, U, G, C nucleobases. We consider randomly generated master sequence sets; however, we impose the constraint that the Hamming distance between any two master sequences must be at least equal to two, to avoid source-sink type dynamics.

We also generated sequences containing a predefined amount of G+C nucleobases ([Figure 3AB](#)), which would lead to increased probability of obtaining compositionally similar random sequences (i.e., generally smaller pairwise Hamming distances between masters, and between masters and their complementary sequences). In order to circumvent this distortion, the nucleobases were randomly shuffled among the loci until the Hamming distance > 3 condition was satisfied for each distinct master-master and master-complementary sequence pair and for the 10 master-complementary sequence pairs of the same type, thereby excluding palindromes.

Replication and mutation

To analyze the diversity-maintaining ability of the system, we generate a T -membered set of master sequences. The ϕ_i replicability of master sequence type i is

$$\phi_i = \phi_{\min} + (i - 1) \cdot \delta \quad (8)$$

where $i = 1, 2, \dots, T$. Note that the difference between the lowest and highest replicability is $(T - 1) \cdot \delta$, thus, δ can be interpreted as *replicability distance*. Mutations reduce the replicability of master sequences by a predefined error factor: $\phi = \varepsilon \cdot \phi_i$, where in case of one mutation, $\varepsilon = 0.2$, and with two mutations, $\varepsilon = 0.05$. A j sequence with more than two Hamming-distance to *any* i master sequence ($\Delta_{ij} > 2$) has a baseline replicability: $\phi_j = 0.05\phi_{\min}$. To determine the replicability of a given replicator, we compute its replicability against all master types and choose the highest value (for instance, if a replicator is a one-error copy of master type 5, a two-error copy of master type 3 and has more than two errors to all other master types, the replicability of the replicator is $0.2\phi_5$). This rule is implemented into the model to treat the (very rarely occurring) situation, when the Δ_{ij} Hamming-distance of a newly synthesized j strand is smaller than 3 to more than one i master sequences. For simplicity we assume that the rate constant of replication (see [Equation 1A](#) and [Equation 1B](#)) is equal to the replicability: $c_r = \phi$. Substitution mutations can occur during the replication with the probability $p_{5,-}$ per base. Because of the fixed sequence length, deletion and insertion are not permitted.

Dissociation, decay and association

Binding energy (E_b) of a duplex (in arbitrary units) is determined by the number of Watson-Crick pairs present in the double strand according to

$$E_b = n_{AU} + 2n_{GC}, \quad (9)$$

where n_{AU} and n_{GC} are the numbers of A-U and G-C pairs, respectively (frameshift is not allowed). The relatively high melting temperature of GC-rich double-stranded RNAs (e.g. [Freier et al., 1986](#)) is considered here by including a prefactor 2 in the H-bond strength of a G-C pair, which renders the binding energy of this pairing to be the double that of an A-U pair. Note that the maximum binding energy is consequently $2L$. If $n_{AU} + n_{GU} < 2$, association is not possible due to low duplex stability. As in our model, dissociation of the replicators is an enzyme-free process, the main driving force of the strand-separation mechanism is temperature. Consequently, the dissociation probability (p_{diss}) of a duplex with binding energy E_b follows the Boltzmann distribution:

$$p_{\text{diss}}(E_b) = \frac{\exp(-\kappa E_b)}{\sum_{E_b=2}^{2L} \exp(-\kappa E_b)}, \quad (10)$$

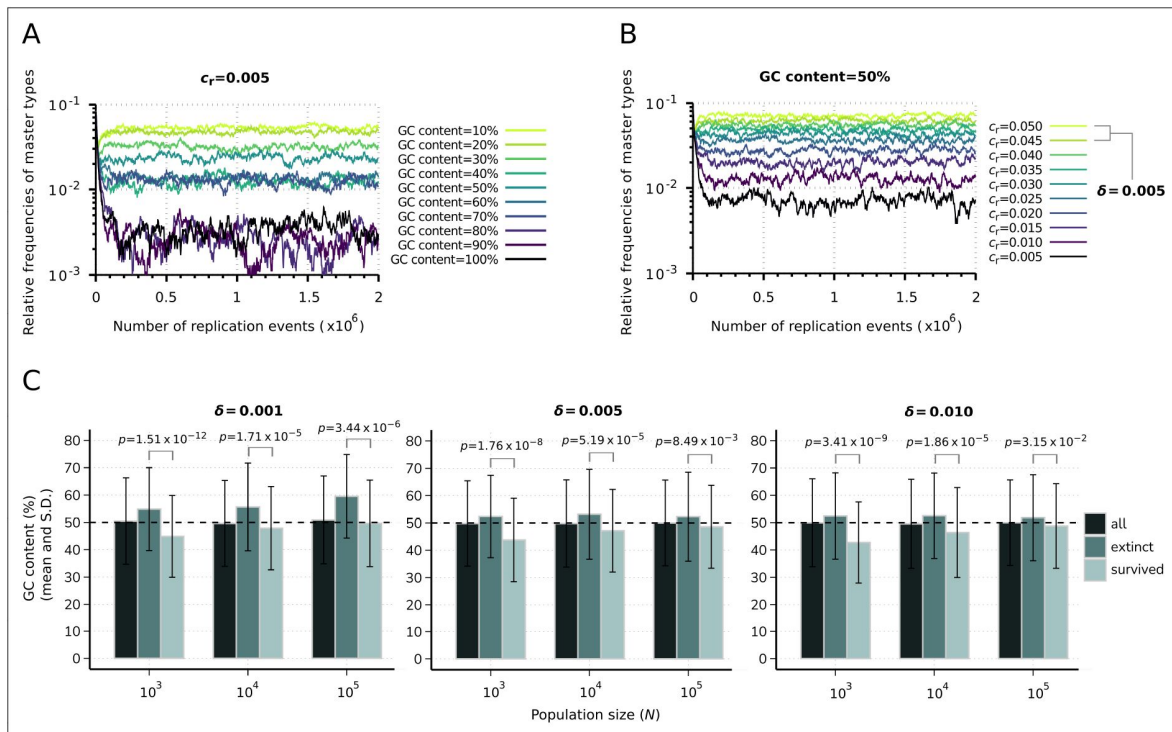


Figure 3.

Effect of sequence composition on replicator abundances in the constant population model. **(A)** Demonstration of the direct effect of sequence composition on replicator abundances. The master types (indicated by differently colored lines) are defined along a GC content gradient with identical replicabilities ($c_r = 0.005$). **(B)** Consequences of removal of the effect of sequence composition on replicator abundances. The master types (indicated by differently colored lines) are defined along a replicability gradient (replicability distance: $\delta = 0.005$) with identical GC contents (50%). Due to the relatively fast convergence, simulations were terminated after $2 \cdot 10^6$ replication events. Parameters that were used in both simulations: $N = 10^4$, $p_{\text{mut}} = 0.01$ (see [Table 1](#)). **(C)** Means and standard deviation of the relative GC content in the survived and extinct master types, corresponding to the simulation results shown in [Figure 2B](#). Columns with the darkest shade represent the average GC content of the randomly generated $n = 500$ master sequences (10 sequences for all 50 replicate runs) constituting the initial populations. Horizontal dashed lines indicate the expected (random) 50% initial GC content. The exact p values resulted from the two-sample Wilcoxon tests on the GC content of the survived-extinct subsets are indicated above the column-pairs.

where κ incorporates the (inverse) temperature and other possible environmental factors that can have potential effects on duplex separation (e.g., ionic concentration, pH, etc.), see e.g. [Le Vay & Mutschler \(2019\)](#). In order to speed up the convergence, we set the κ value so that the dissociation probability of a duplex with maximum binding energy is 0.01 (i.e., $p_{\text{diss}}(2L) = 0.01$). For sake of simplicity we assume that the rate constant of dissociation is equal to the dissociation probability: $c_d = p_{\text{diss}}$, see [Equation 3A](#) and [Equation 3B](#).

The degradation rate of single strands ($c_{\delta_{ss}}$) is constant, regardless of the sequence. The decay rate of duplex molecules is affected by the binding energy according to

$$c_{\delta_{ds}} = f \cdot c_{\delta_{ss}} \cdot 0.8^{E_b}, \quad (11)$$

where the duplex decay factor parameter f allows us to scale the duplex decay rates relative to that of the simplexes. By assuming $0 < f \leq 1$ and satisfying the condition of $E_b \geq 2$ for every duplex (which is ensured by the $n_{AU} + n_{GC} \geq 2$ condition), simplex molecules have higher decay rate than duplexes. This model assumption reflects the fact that the structure of double-stranded RNA provides a certain degree of stability against alkaline hydrolysis in solutions relative to that of single-stranded RNA ([Zhang et al., 2021](#)).

The association rate constant is sequence-independent: $c_a = 1$ for all pairs of single strands. Cross-hybridization between not fully complementary strands is allowed, the binding energy in this case is also computed according to [Equation 9](#).

Stochastic simulation algorithm

In order to take finite population sizes and stochastic effects into account, we consider an individual-based, stochastic approach of parabolic dynamics. Accordingly, the model is implemented as a Gillespie stochastic simulation algorithm ([Gillespie, 1976](#), [1977](#)) adapting the optimized direct method (ODM) formulation (see, for example, [Cao et al., 2004](#); for details). The algorithm randomly chooses a reaction to take place, based on a propensity function that is specified for every μ reaction channel as:

$$\alpha_{\mu}(t) = h_{\mu} c_{\mu}, \quad (12)$$

where $\mu = 1, \dots, M$, and M is the number of possible reaction channels, c_{μ} denotes the corresponding rate constant, h_{μ} is the number of the combinatorically distinct potential configurations of μ reaction-compatible molecules at time t . τ denotes the time when the next reaction occurs after t . This quantity is calculated as a function of the summed propensities:

$$\tau = \frac{1}{\sum_{\mu=1}^M \alpha_{\mu}(t)} \ln \frac{1}{r}, \quad (13)$$

where r is a uniformly distributed random number from the $[0, 1]$ interval. After one of the reactions is chosen and takes place, the frequencies of the corresponding chemical species are updated based on the reaction kinetic equations ([Equations 1–7](#)). Then the respective propensity functions are also updated and this procedure is repeated until reaching the termination condition.

In all simulations, initially each master type is present in equal copy numbers in simple-stranded form. We monitor the frequency of the master types, the replicability of all replicators including mutant copies of the masters, the number of surviving master types, the simplex-duplex asymmetry and the total number of replicators (in the chemostat system). The steady state characteristic of coexistence, i.e., whether the number and the identity of the surviving master

types are in dynamical equilibrium is assessed as follows: the relative frequency of each master sequence is tested for the condition of $\geq 10^{-3}$ (cut-off relative frequency for survival) in every 2,000th replication step in the interval between 10,000 replication steps before termination and actual termination (10^7 replication steps, unless stated otherwise). According to our numerical experiences, convergence to an apparent steady state is always reached until the beginning of this time interval, even though stochastic extinction events can occur at any time. If the above condition for the cut-off relative frequency for survival holds in more than one such sample, the master type in question is regarded as survived, otherwise it is considered to be extinct in dynamical equilibrium. All simulations were performed in C. Statistical tests included in this manuscript were performed in the R environment (R Core Team, 2023 [link](#)), using two-tailed hypothesis tests with alpha risk of 0.05.

Results

Constant population model

Coexistence of the replicators

First, we investigated the diversity-maintaining ability of the parabolic dynamics at different population sizes and replicability distances. **Figure 2A** [link](#) shows the results of individual runs with $T = 10$ master types in case of small ($N = 10^3$), medium ($N = 10^4$) and large ($N = 10^5$) population size and small ($\delta = 0.001$), medium ($\delta = 0.005$) and large ($\delta = 0.01$) replicability distances. In all cases, the smallest replicability of the master types is $\phi_1 = \phi_{\min} = 0.005$. Note that, the ratio of the smallest and largest master replicabilities $\left(\frac{\phi_{10}}{\phi_1}\right)$ in the cases of three different δ s are 2.8, 10 and 19,

respectively. These enormously large fitness ratios and fitness differences result in a highly competitive regime in Darwinian sense and also provide the opportunity to investigate the diversity-maintaining capacity of the parabolic regime. The other key parameter of the system is the population size. It is known that in the infinite population size limit, the sustainable diversity has no upper bound (Szathmáry & Gladkih, 1989 [link](#)), but we are interested in how the sustained diversity decreases with decreasing population size. As a general observation, we can state that with decreasing population size, the effect of demographic stochasticity increases and can drive some of the master types extinct. **Figure 2A** [link](#) shows that increasing the population size reduces the fluctuations in the frequencies and a higher population size together with smaller replicability distance increases the number of sustained master types. We measured the average and standard deviation of the number of maintained master types as the function of population size and replicability distance (**Figure 2B** [link](#)). In case of small population size, the parabolic regime can maintain 43.8%, 29.8% and 25.6% of the master types at an average, assuming small, medium and large replicability distance, respectively. The corresponding percentages of sustained master types in case of medium population size are: 77.6%, 56.2% and 48.2%; and in case of large population size: 87.2%, 62% and 59.8%. The average values are also in agreement with the expectations: both large population size and small replicability distance increase the sustainable diversity.

A set of survived master replicator types and thus the selectivity of a parabolically replicating system can be characterized by the sum of replicability rank indices of the coexisting master population. According to **Equation 8** [link](#), the master type with the highest replicability has rank 10, the second highest has rank 9, while the master type with the lowest replicability has rank 1. If T_{surv} different types of master replicators survive (coexist), the maximum of the sum of their ranks is $\rho_{\max} = \sum_{i=1}^{T_{\text{surv}}} 11 - i = \frac{T_{\text{surv}} \cdot (21 - T_{\text{surv}})}{2}$, ($1 \leq T_{\text{surv}} \leq T$). The expected value of the sum of ranks, when the surviving masters are distributed randomly with respect to their replicabilities is $\rho_{\text{rand}} = 5.5 \cdot T_{\text{surv}}$. According to our simulation results, the sum of replicability ranks of the coexisting types is typically lower than the theoretical maximum ($\rho_{\max}$), but higher than the

random composition (ρ_{rand}), indicating that the fitness of a replicator does not only depend on its replicability and therefore the dynamics does not *only* selects for relatively high replicabilities (**Figure 2C** [↗](#)).

Effect of sequence composition

We found that the key factor leading to the deviation from the theoretical maximums of the sum of replicability ranks of the surviving master subsets is the GC content (the sum of G and C nucleobases) of the sequences. The direct effect of GC content on the replicator dynamics and thus the selectivity of a parabolically replicating system with regard to sequence composition were investigated by observing the time evolution of $T = 10$ master sequences generated along a GC content gradient assuming equal replicabilities. For this investigation, we allocated the same number of G+C nucleobases (with 50-50% probability of G and C) to the master types as their indexes. Then A or U nucleobases (also with 50-50% probability) were allocated to the remaining loci and random shuffling of the nucleobases among the loci was carried out. This analysis showed that equilibrium master sequence frequencies follow an inverse relation with the GC content gradient, with increasing stochastic fluctuations in the frequencies at relatively large GC content values (**Figure 3A** [↗](#)).

To address the question of how the parabolic dynamics operates when the above-described effect of sequence composition is excluded, we examined master sequence abundances under the assumption of balanced GC content, along a replicability gradient and with cross-hybridization between different master types not being allowed. In this scenario, master sequences were generated in the following way: equal GC content of the sequences was ensured by allocating G or C (with 50-50% probability) nucleobases to the first half of the loci. Then, A or U nucleobases (also with 50-50% probability) were allocated to the remainder second half of the loci and random shuffling of the nucleobases among the loci was carried out. Another restricting condition considered during this investigation is that master-master and complementary-complementary sequence associations are not allowed, nor master-complementary hybridization of different types, thereby preventing the formation of cross-hybrid duplexes with different binding energies. This investigation showed that if the effects of sequence composition and cross-hybridization are excluded, (i) equilibrium master sequence frequencies are directly proportional to their replicabilities and exactly follow the replicability gradient, (ii) survival of each master sequence is ensured because of the fact that in this way each master inhibits (regulates) only itself and no other masters, i.e., the dynamics is “purely” parabolic (cf. Meszéna & Szathmáry, 2002 [↗](#)), (iii) despite a relatively low total population size ($N = 10^6$), stochastic fluctuations have negligible effects on the dynamics (**Figure 3B** [↗](#)). A direct investigation of whether the sequence composition of the master types is associated with their survival outcome was conducted using the data from the constant population model simulation results (**Figure 2** [↗](#)). In these data, the average GC content was measured to be lower in the surviving master subpopulations than in the extinct subpopulations (**Figure 3C** [↗](#)).

To determine whether this difference was statistically significant, nonparametric, two-sample Wilcoxon rank-sum tests (Hollander & Wolfe, 1999 [↗](#)) were performed on the GC content of the extinct-surviving master subsets. The GC content was significantly different between these two groups in all nine investigated parameter combinations of population size (N) and replicability distance (δ) at $p < 0.05$ level, indicating a selective advantage for a lower GC content in the constant population model context. The exact p values obtained from this analysis are shown in **Figure 3C** [↗](#).

Effect of mutations

We investigate the effect of the mutation rate on the sustainable diversity and the proportion of master types in the total replicator population. **Figure 4A** [↗](#) shows four realizations with different per bit mutation rates: $p_{\text{mut}} = 0.01, 0.05, 0.1, 0.15$. In accordance with the expectations, increasing

mutation rates reduce the equilibrium abundances of masters by widening the mutant cloud. **Figure 4B** shows this effect as average of 20 independent runs for each mutation rate. The corresponding average of percentages of sustained master types are: 65%, 61%, 43.5% and 16%, respectively. Summed relative steady-state master frequencies are a decreasing function of the mutation rate, see **Figure 4C**.

Lack of error threshold in case of parabolically replicating infinite populations

As a reference case, we analyzed the behavior of a simplified dynamics of parabolic replication (**Equation 15A** and **Equation 15B** in *Appendix 1*). For analytic tractability, we assume infinite population size and a single peak replication landscape in which a single genotype has high replication rate, all others have the same baseline value, the ratio of these two values (denoted by A) is the reproductive superiority. Furthermore, we consider the essential requisite that ensures parabolic dynamics, namely that the growth rate of the population is proportional to the square root of the actual size of the population.

Our analytical results show that in this simple model – except for $Q = 0$, where Q denotes the probability of the error-free replication of a master sequence – there is no error threshold: the master type will persist at any level of mutation rate (**Figure 5**, right panel). By contrast, in non-parabolic dynamics (i.e., in the Eigen-model, referred to as exponential dynamics, see, **Equation 14A** and **Equation 14B** in *Appendix 1*), the master type disappears from the system above a critical mutation rate $Q^* = \frac{1}{A}$ (**Figure 5**, left panel).

Introducing back mutations into the model (so far neglected as a rare process) fundamentally changes the dynamics: the error threshold also disappears from exponential the regime, see **Appendix 1—figure 2**. In the parabolic case, under a realistic parameter range – e.g., in the interval, where the probability of back mutation $p_b = [0; 0.1]$ – the smaller the probability of back mutations the larger the equilibrium concentration of the parabolic masters in comparison with that of the exponential masters (**Appendix 1—figure 3D**). In terms of the difference in the equilibrium concentrations between the parabolic and the exponential masters, the smaller the replication fidelity, the higher the relative parabolic master concentration for both mutation scenarios (**Appendix 1—figure 1** and **Appendix 1—figure 3A–C**).

Chemostat system

Diversity-maintaining ability

The chemostat model, by taking into account the components of replicator decay, inflow of nucleobases into the system and outflow of the molecules, shows a richer dynamical repertoire compared to the constant population model. The total population size and the diversity maintaining ability of the system are jointly affected by both the inflow rate of the nucleobases (c_{in}) and the decay rate of replicators in simplex ($c_{\delta_{ss}}$) and duplex ($c_{\delta_{ds}}$) form, cf. **Equation 11**. **Figure 6A** shows individual realizations of this dynamics with the parameter values that lead to comparable equilibrium population sizes to those investigated in the constant population model (**Figure 2**). Note that in the chemostat model, the total population size (N) changes in time, and therefore population sizes shown in **Figure 6A** are approximate measures. Consistent with the constant population model, the dynamical outcome approaches complete coexistence if large population sizes are combined with relatively similar replicabilities (**Figure 6A**, upper right panel) under the assumption of small duplex decay rates relative to that of the simplexes ($f = 0.01$). In line with the expectations, decreasing c_{in} reduces the average total population size ($\bar{N}$), and this simultaneously decreases the number of sustainable master types (**Figure 6AB**). Our results show that decreasing c_{in} by one order of magnitude, as expected, leads to approximately one

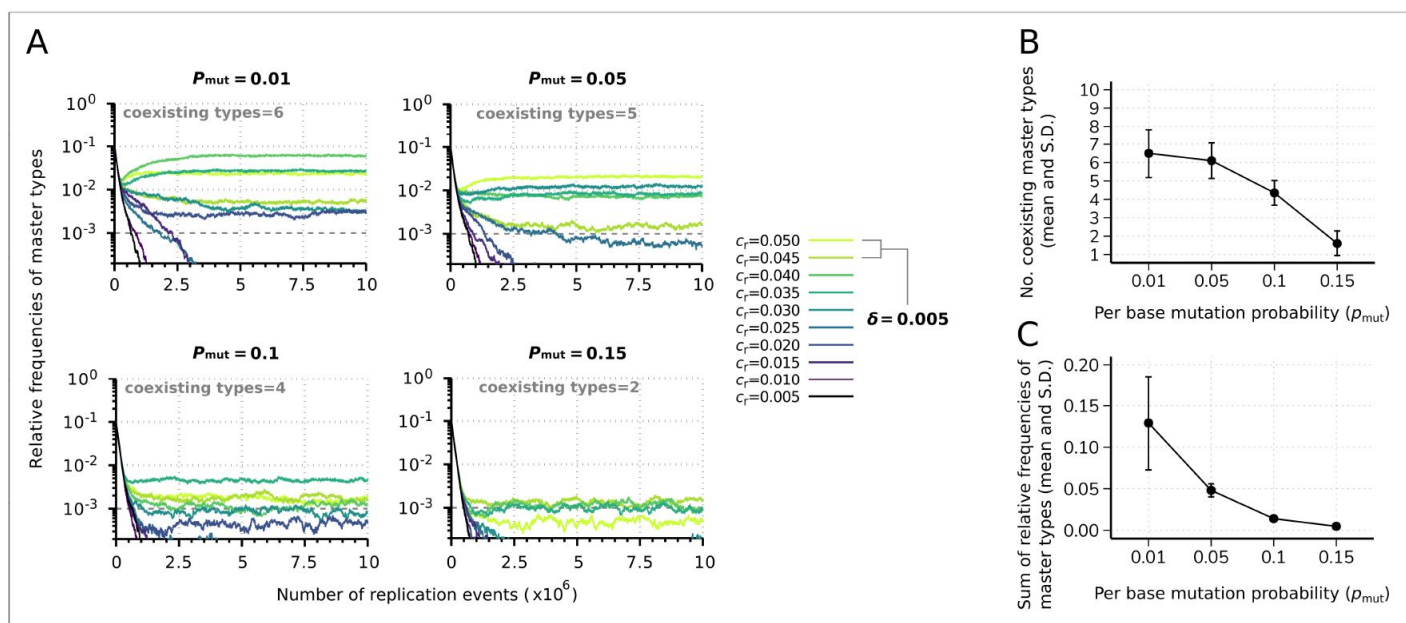


Figure 4.

Effect of the mutation rate on the replicator coexistence in the constant population model. **(A)** Time series of the relative frequencies of master types at different mutation rates (p_{mut}). Line colors denote master type replicabilities (values are indicated in the legend). The cut-off relative frequency for survival is indicated by horizontal dashed lines. **(B)** Mean and standard deviation of the coexisting master types as the function of p_{mut} calculated from 20 independent runs. **(C)** Mean and standard deviation of the sum of relative frequencies of master types in the total population as the function of p_{mut} corresponding to the simulation results shown in **(B)**. Except for p_{mut} , the default parameter set (see [Table 1](#)) was used.

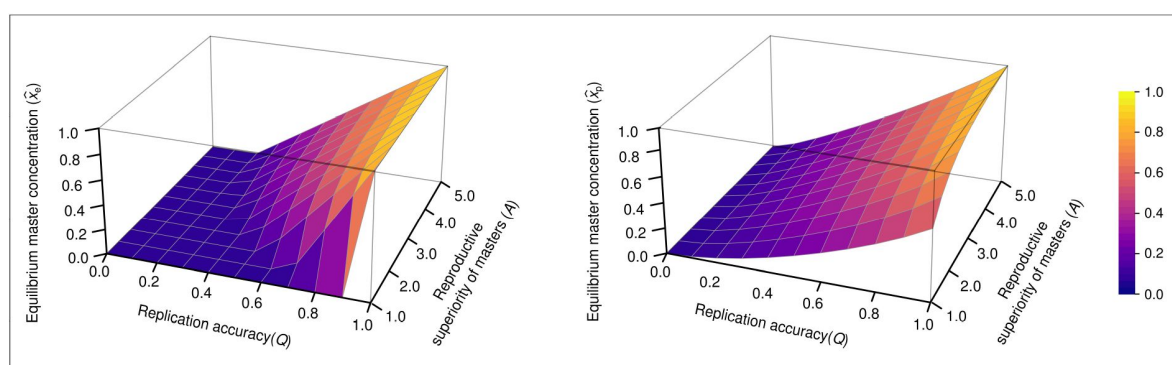


Figure 5.

The concentration of master type ($\hat{x}_e$ and $\hat{x}_p$) as the function of Q and A in the exponential model (left panel) and in the parabolic replication model (right panel). $\alpha = 1$ (where α stands for the reproductive value of the mutants) was used.

order of magnitude drop in the average total population sizes (**Figure 6B**). Relatively large replicability distance parameter (δ) values, however, while narrowing down the range of the coexistence, increase the average total population sizes at the same time (**Figure 6B**) due to the fact that they imply larger mean master replicabilities.

For further analysis, we introduce the simplex-duplex asymmetry measure: $\chi = \frac{N_{\text{simplex}} - N_{\text{duplex}}}{N}$, whose negative values indicate duplex, positive values indicate simplex dominance, -1 (+1) value means that the whole population is in duplex (simplex) state. Along an increasing gradient of the duplex decay factor (f), the average total population sizes are again lowered (**Figure 6B**). The mechanistic explanation for this population-shrinking effect at relatively large values of f is well demonstrated by the population compositions, showing that the overwhelming majority of the replicators tend to be in double-stranded state independent from δ (**Figure 6C**). Note, however, the considerable shift toward the single-stranded state along a decreasing resource influx (c_{in}) and average total population size gradient (**Figure 6C**).

Shift in selectivity

In case of a strongly resource-limited regime, the diversity-maintaining potential of the dynamics can ultimately drop to a level at which the system can maintain only one master type (**Figure 6B**). Our results show that such a single-species equilibrium generally occurs under the assumption of the smallest resource inflow rate ($c_{\text{in}} = 0.01$), combined with large replicability distance ($\delta = 0.01$) and duplex decay factor ($f = 1$). In the chemostat model binding energy of a duplex (and thus its sequence composition) is also taken into account in the decay probability distribution (c.f. **Equation 11**). Therefore, a higher GC content results in a lower decay rate, compensating to some extent the otherwise disadvantageous effect of a relatively high GC content (**Figure 3A** and **Figure 3C**) and the consequential low dissociation probability (only dissociated, single-stranded sequences are available as templates for complementary strand synthesis). This sequence composition effect-mediated trade-off between lowered decay rate in duplex state and templating affinity in simplex state, together with type-specific replicabilities, determines the identity of the surviving master types and thus the selectivity of the system in single species equilibria.

Discussion

In this modelling study we focused on the coexistence of polymer variants involved in a template-directed, non-enzymatic replication mechanism. Prior to the emergence of a helicase ribozyme that could have facilitated strand separation in RNA duplexes (or a similar, relatively effective strand-separation mechanism, see, e.g., Szostak et al., 2001; Müller, 2006), such an enzyme-free replication mode was likely to result in parabolic growth profiles (von Kiedrowski et al., 1991). Therefore, parabolic population growth is of conceptual relevance to the origin and the subsequent evolution of heritable chemical information. Within this framework, we demonstrated that – depending on the total replicator population size, GC content, the extent of replication rate differences and copying error as well as resource influx rates – parabolic dynamics can sustain a large variety of sequence modules.

A small total population size (10^3) combined with large replication rate differences ($\delta = 0.01$), on average, can only maintain 26% of all species (**Figure 2B**). These results corroborate Davis's (2000) analytical findings that steady-state replicator variant frequencies in a finite, sublinear system are truncated in such a way that variants with insufficient fitness are eliminated. In contrast, large population sizes and smaller between-species fitness differences are expected to approach the deterministic, unlimited coexistence limit (Szathmáry & Gladkih, 1989; Varga & Szathmáry, 1997). This is confirmed here by considering the combination of a relatively large

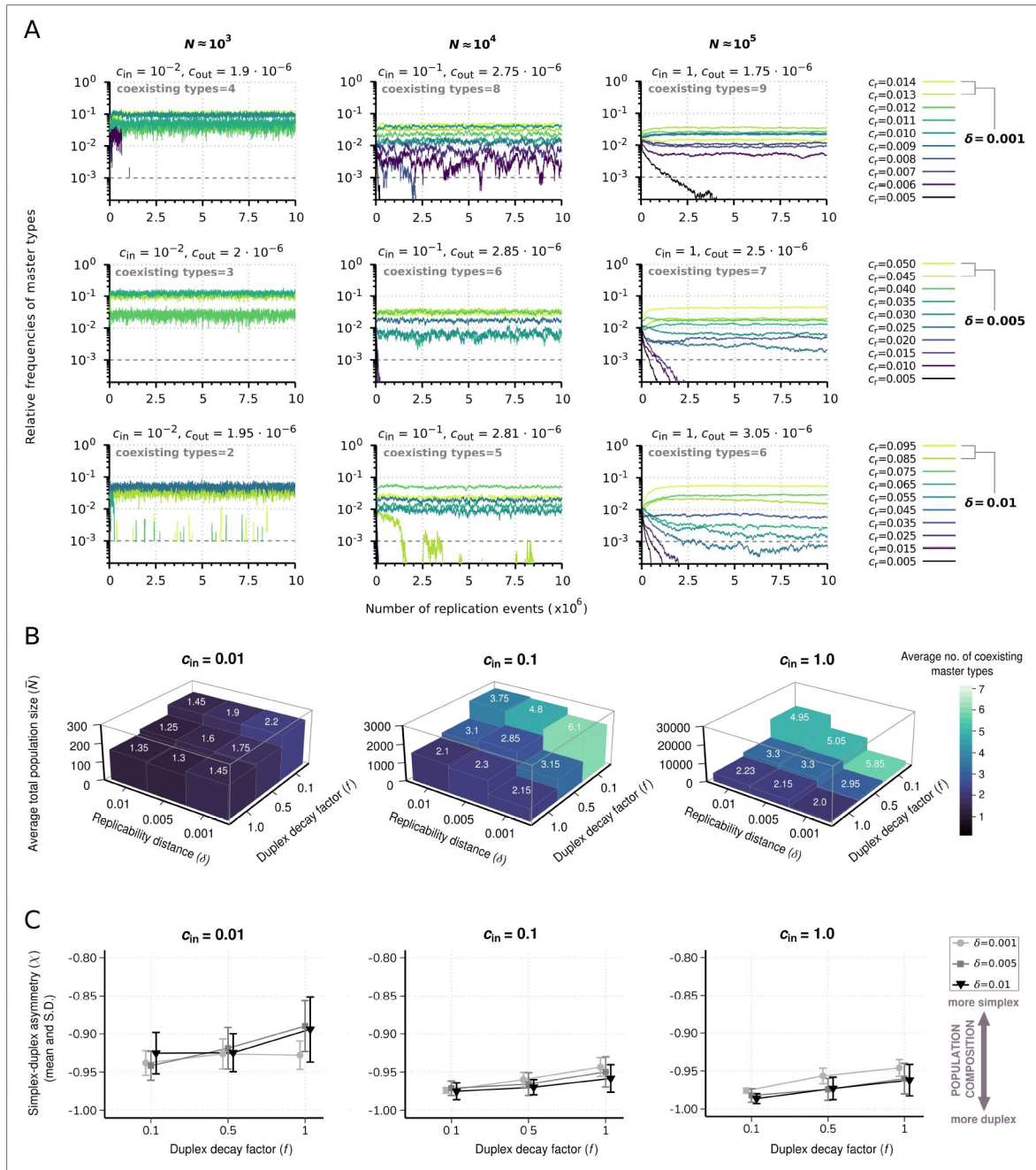


Figure 6.

Effect of resource competition on replicator coexistence in the chemostat system model. **(A)** Time series of the relative frequencies of master types. Panels corresponding to different columns demonstrate different population sizes (N). Panels corresponding to different rows demonstrate different replicability distances (δ). Line colors indicate master types with distinct replicabilities (as explained in the legends). The cut-off relative frequency for survival is indicated by horizontal dashed lines. **(B)** Averages for the total population size ($\bar{N}$) and the number of coexisting master types (color and numbers on columns) as functions of δ and duplex decay factor (f) at different resource inflow rates (c_{in}). Averages were obtained from 20 individual runs for every parameter combination. **(C)** Mean and standard deviation of simplex-duplex asymmetry ($\chi = \frac{N_{\text{simplex}} - N_{\text{duplex}}}{N}$) as the function of δ and f at different c_{in} values, corresponding to the simulation results shown in **(B)**. The default parameter set (see [Table 1](#)) was used, unless otherwise indicated.

total population size (10^5) and small equidistant difference in the replication rates ($\delta = 0.001$), which leads to the coexistence of almost 90% of the variants, on average (**Figure 2B**). The replicability (fitness) distance dependency of the number (or the proportion) of the coexisting variants in dynamical equilibrium (**Figure 2B**) is viewed as the manifestation of fitness-driven (Darwinian) selection during competitive replication at sublinear propagation rates (Davis, 1996a, 1996b, 2000). Analytical and numerical results show that the threshold fitness for coexistence depends on the fitness sum in such a system (Davis, 2000). Consistent with this, elevation in the sum of species fitness – that can be expressed in the present model in terms of the increasing sum of replication rates of master species with increasing replicability distance ($\delta = 0.001$: $\sum_{i=1}^T c_{r(i)} = 0.095$; $\delta = 0.005$: $\sum_{i=1}^T c_{r(i)} = 0.275$; $\delta = 0.01$: $\sum_{i=1}^T c_{r(i)} = 0.5$) – raises the fitness threshold for survival and thus the number of subthreshold species that are eliminated (**Figure 2AB**) due to the fact that master species with higher fitness are present in higher copy numbers (**Figure 2A**).

With regard to demographic stochasticity, we considered order of magnitude-scaled distances between the investigated population sizes, while a constant cut-off relative frequency for survival (10^{-3}) was applied, implying that the minimum copy number to avoid exclusion also increased by an order of magnitude with increasing population size (**Figure 2A**). The rationale behind this assumption is that in this way the average number of the maintained variants is expected to change linearly as a function of population size, if fluctuations associated in population size shifts affect the diversity-maintaining potential linearly. In contrast to linear dependence, however, we found that the slope of this function is steeper between the smallest and the medium population sizes than between the medium and the largest one and that this non-linearity is becoming more pronounced with decreasing fitness differences (i.e., with smaller δ , **Figure 2B**). Accelerating species extinction at the smallest population size, resulting from increased variability in relative species frequencies over time (i.e., increased amplitude of stochastic fluctuations), is consistent with prior predictions and analytical findings that suggested the existence of a critical total replicator population size above which parabolic coexistence is guaranteed (Epstein, 1979; Scheuring & Szathmáry, 2001). These remarks emphasized that effective diversity maintenance by parabolic replication should have required sufficiently large population sizes during early molecular evolution, which, according to our results, must have been $>10^3$.

Moreover, we found evidence that under sublinear dynamics, GC content of the replicator sequences has also a significant effect on the steady-state master frequencies (**Figure 3A**) and, thus, on the identity of the surviving master types (**Figure 2C** and **Figure 3C**). Consequently, this feature also plays a decisive role in determining the number of the surviving variants in equilibrium in a manner that a balanced GC content favors a larger number of coexisting variants (**Figure 3B**), whereas, for example, an extremely low GC content of a few master sequences is expected to lead the survival of this narrow subpopulation of the sequences with low GC content. This finding can be interpreted in the light of the lifetime reproductive ratio of the replicators. Reproductive ratio in this case can be defined as the expected number of copies (“offspring”) produced by a single strand or, in case of complementary template replication, the number of complementary strands synthesized from a given sequence during its lifespan (Mészéna & Szathmáry, 2002). According to this interpretation, a relatively high dissociation probability and the consequential higher propensity of being in single-stranded form provides an advantage to the sequences with relatively low GC content in terms of their replication affinity, that is, the expected number of offspring in case of such variants will be relatively high. This model characteristic is also in accordance with the results of a computational study on RNA folding from randomly generated sequences, showing that abundant and topologically simple fold structures tend to arise from sequences depleted in G, thereby suggesting that the first catalytic RNA sequences could have been characterized by a somewhat reduced GC content (Briones et al., 2009).

We demonstrated that during competitive replication of parabolically growing replicator variants, monomer resource inflow rate (c_{in}) can act as a control parameter of selectivity. As such, by changing the values of this parameter, the dynamics can switch from a non-Darwinian steady state that ensures coexistence towards a Darwinian state, where one competitively superior species prevails, other variants are competitively excluded. In an analogous replicator competition model, Szilágyi and co-workers applied Gause's principle of competitive exclusion (Gause, 1934) to replicator populations and showed both analytically and numerically that if resources (nucleotides) affect the growth linearly, then the number of coexisting replicator species cannot be more than that of the nucleotide building block types (note that plus and minus copies of the same replicator count as one species: Szilágyi et al., 2013). By contrast, the present study provides evidence that within the context of sublinear propagation, at high monomer resource concentrations that ensure large population sizes (Figure 6A), the average number of the maintained replicator types can be higher than the number of resource types (4) (Figure 6B). However, with small monomer inflow rate, sublinear population growth also becomes resource limited concomitant with a considerable drop in the average total population size (Figure 6B), leading to the survival of the fittest variant.

We also investigated the consequences of errors in the replication mechanism in the context of the theoretical error threshold of parabolic dynamics. Our analytical results show that a master sequence vanishes from the system, only if the replication accuracy is zero (Figure 5, right panel). Thus, in this deterministic scenario, the error threshold problem is seemingly circumvented, because persistent maintenance of essential genetic information by non-enzymatic, template-directed synthesis of short RNA-like information carrying sequence modules is not an irresolvable task for a putative rudimentary replication mechanism, even if it is highly inaccurate. However, when finite population sizes and stochastic effects are taken into account, at the largest investigated per-base mutation rate ($p_{mut} = 0.15$), the summed relative steady-state master frequencies approach zero (Figure 4C) with accelerating species extinction (Figure 4B), indicating that this value is close to the system's empirical error threshold. We note that this value is above the one that can be anticipated if one applies the inverse genome length rule of thumb for the error threshold (Eigen, 1971; Joyce, 2002) to our system, which gives $1/L = 0.1$ error rate of replication per nucleotide (but note that applying this approach to our system is a serious oversimplification). Moreover, the applied constant cut-off relative master frequency for survival (10^{-3}) renders our empirical error threshold approximation rather conservative. Former studies suggest that a 0.1 per-base error rate, at which we measured a decent diversity-maintaining potential (more than 40% of all variants), is a reasonable assumption for non-enzymatic RNA copying. Although the average error rate in this process was estimated to be ~ 0.17 (Leu et al., 2011), using optimized nucleotide ratios could potentially lower this measure below ~ 0.1 with a possibility for a further reduction to ~ 0.05 on GC-rich templates (Szostak, 2012). However, all of the average error rates reported by these studies pertain to non-enzymatic polymerization reactions involving primers. Triplet substrates, albeit in an enzymatic context (i.e., with in vitro evolved ribozymes), have proven to be able to successfully bypass the requirement of primers (Attwater et al., 2018). A similarly primer-free and therefore – under prebiotic conditions – more realistic replication mode has been proposed for non-enzymatic RNA replication in which complementary strand synthesis takes place by oligonucleotide ligation together with gap filling reactions, making use of monomers and shorter oligomers (Szostak, 2012). Although the fidelity of this RNA replication mechanism is largely unknown, a pilot experimental study (James & Ellington, 1997), which was performed with constant-sequence DNA templates, suggests a surprisingly high accuracy of complementary strand synthesis during this process.

Ideas and Speculation

As we have demonstrated, template-directed, non-enzymatic replication of oligomer modules leading to parabolic growth profiles is an effective genetic information maintaining mechanism which may thus have constituted an indispensable bridge from the very first abiotic synthetic pathways for RNA to a ribozyme-dominated stage of the RNA world. The demonstrated diversity-

maintaining mechanism of finite parabolic populations can be used as a plug-in model to investigate the coevolution of naked and encapsulated molecular replicators (e.g., Babajanyan et al., 2023 [DOI](#)). It is, however, noteworthy that besides the “information first” view of the RNA world hypothesis, there are other concepts as to how life arose; for example, the “metabolism first” or autotrophic origins view (Wächtershäuser, 1988 [DOI](#); Joyce, 2002 [DOI](#); Martin et al., 2008 [DOI](#)). Moreover, this view does not exclude the possibility of the existence of a pre-RNA world, in which genetic information resided in polymer structures similar to RNA, such as peptide- or tetrose nucleic acids (Schöning et al., 2000 [DOI](#); Bada, 2004 [DOI](#); Orgel, 2004 [DOI](#)). Considering that we applied a rather broad approach to replicator representation we believe that the results of the present model are general enough for being naturally applicable to pre-RNA systems as well. Although short oligonucleotides like those we have investigated in the present study are not characteristically prone to intramolecular 2D folding and therefore typically bear with limited catalytic activities, miniribozymes with surprisingly short sequence lengths can still catalyze a wide variety of chemical reactions (Vlassov et al., 2005 [DOI](#)). Therefore, catalytic aid and/or metabolic cooperation among independently replicating sequences (i.e., *cooperative coexistence*, see e.g., Epstein, 1979 [DOI](#); Mizuuchi & Ichihashi, 2018 [DOI](#)) can considerably broaden the parameter space in which we reported *competitive coexistence* of different replicator variants.

In this study we considered kinetic constants that correspond to the $p \approx 0.5$ growth order. Further investigations into the $0.5 < p < 1$ interval (Issac & Chmielewski, 2002 [DOI](#); Li & Chmielewski, 2003 [DOI](#); Lincoln & Joyce, 2009 [DOI](#)) are expected to show that an increase in the kinetic growth order decreases the scope of competitive coexistence and drives the system more toward the Darwinian regime. Such studies may help gain a deeper insight into the question whether the emergence of an effective strand separation mechanism, or other conditions implying kinetic constants that shift the dynamics toward exponential amplification, allowed for a still sub-exponential diversity maintaining mechanism during later phases of chemical evolution and thus a gradual evolutionary transition from non-enzymatic replication to ribozyme-aided amplification.

The two key components of evolutionary algorithms are exploration and exploitation (Bäck et al., 2000 [DOI](#)). If diversity decreases too fast under selection, this can lead to premature convergence, resulting in a population getting stuck at a local adaptive peak. In order to avoid premature convergence, various diversity-maintaining mechanisms have been introduced. We call attention to the fact that parabolic growth is an organic form of diversity maintenance. Referring to the chemostat model, we propose that a periodically changing environment with alternating resource abundance and shortage can drive, and in the distant past could have driven, an efficient evolutionary exploration-exploitation algorithm.

Acknowledgements

This work was supported through National Research, Development and Innovation Office Élvonal KKP129848 and the Templeton World Charity Foundation (“Learning in evolution, evolution in learning” award - TWCF0268). AS received support from the Hungarian Academy of Sciences through Bolyai János Research Fellowship program. MP received support from the ELTE Eötvös Loránd University through Hungarian state PhD scholarship. We thank Géza Meszéna for helpful discussions, as well as János Podani for comments on the manuscript. We would like to thank Balázs Könnyű for his advice on the statistical analysis and Dániel Vörös for his help with the coding.

Additional Information

Funder	Grant reference number	Author
National Research, Development and Innovation Office	Élvonal KKP129848	Mátyás Paczkó Eörs Szathmáry András Szilágyi
Templeton World Charity Foundation	TWCF0268	Mátyás Paczkó Eörs Szathmáry András Szilágyi
Hungarian Academy of Sciences	Bolyai János Research Fellowship	András Szilágyi
ELTE Eötvös Loránd University	Hungarian state PhD scholarship	Mátyás Paczkó

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Competing interests

The authors declare that they have no competing interests.

Author contributions

Mátyás Paczkó, Conceptualization, Software, Visualization, Writing – original draft, Writing – editing; Eörs Szathmáry, Conceptualization, Funding acquisition, Formal analysis, Methodology, Writing – original draft, Writing – review; András Szilágyi, Conceptualization, Funding acquisition, Supervision, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review

Data availability

The current manuscript is a computational study, so no data have been generated for this manuscript. Modelling code is publicly available on GitHub under the GNU General Public License v3.0 license: https://github.com/paczkomatyas/Parabolic_Replicator_2023 

Appendix 1

Simple Maynard-Smith-type error threshold for parabolic replicators

No back mutation

The dynamics of the master type (denoted by x) and the mutant types (y) for the exponential regime in the original Maynard-Smith model (Maynard Smith & Szathmáry, 1995 [DOI](#); pp. 44-49) are as follows:

$$\dot{x} = AQx - x(Ax + ay) \quad (14A)$$

$$\dot{y} = ay + A(1 - Q)x - y(Ax + ay), \quad (14B)$$

while its parabolic counterpart has the following form:

$$\dot{x} = AQ\sqrt{x} - x(A\sqrt{x} + a\sqrt{y}) \quad (15A)$$

$$\dot{y} = a\sqrt{y} + A(1 - Q)\sqrt{x} - y(A\sqrt{x} + a\sqrt{y}), \quad (15B)$$

where A and a the master and mutant reproductive values, respectively, if $0 \leq Q \leq 1$ is the probability of the error-free replication of a master sequence, the last terms in the equations guarantees the constant unit concentration ($x + y = 1$).

The non-trivial fixed points of [Equation 14A](#) and [Equation 14B](#) are:

$$(\hat{x}_0, \hat{y}_0) = (0, 1) \quad (16A)$$

$$(\hat{x}_e, \hat{y}_e) = \left(\frac{AQ - a}{A - a}, \frac{A(1 - Q)}{A - a} \right). \quad (16B)$$

If $Q > \frac{a}{A}$ the $(\hat{x}_e, \hat{y}_e)$ fixed point is asymptotically stable, the master type can coexist with its mutants, $Q^* = \frac{a}{A}$ is the error threshold where this fixed point loses stability and when $Q < \frac{a}{A}$ the $(\hat{x}_0, \hat{y}_0)$ becomes the stable fixed point, the master type dies out.

In contrast to the exponential system, parabolic dynamics does not exhibit a sharp change in the concentration of the master as the function of the replication fidelity Q . The interior fixed point of the system (which is asymptotically stable in the $0 < Q < 1$ region):

$$\hat{x}_p = \frac{a^2 + 2A^2Q - \sqrt{a^4 + 4a^2A^2Q(1 - Q)}}{2(a^2 + A^2)} \quad (17A)$$

$$\hat{y}_p = \frac{a^2 + 2A^2 - 2A^2Q + \sqrt{a^4 + 4a^2A^2Q(1 - Q)}}{2(a^2 + A^2)}. \quad (17B)$$

It is easy to prove that x_{0R} is always positive in $0 < Q < 1$, since the expression under the root sign can be estimated as follows:

$$\sqrt{a^4 + 4a^2A^2Q(1-Q)} = \sqrt{[a^2 + 2A^2Q(1-Q)]^2 - 4A^4Q^2(1-Q)^2} < a^2 + 2A^2Q.$$

Note that the master type vanishes if and only if $Q = 0$ ([Figure 5](#), right panel), and mutants cannot vanish because of parabolic competition. The master type can be lost only through stochastics at sufficiently low equilibrium $\hat{x}_p$ values, but not otherwise.

When $Q^{**} = \frac{a+A}{2A}$, both regimes have the same master concentrations ($\hat{x}_p = \hat{x}_e$), below this value the parabolic, above this value the exponential system has more masters. As $Q^{**} > Q^*$ (whenever $A > a$), the length of the replication accuracy interval where the parabolic regime maintains a higher equilibrium master concentration ($Q^* < Q < Q^{**}$):

$$Q^{**} - Q^* = \frac{A - a}{2A}. \quad (18)$$

This means that the “parabolic higher” Q range increases with the reproductive superiority of masters (A). [Appendix 1—figure 1](#) shows the different outcomes as the function of the A and Q , the continuous red and dashed black curves correspond to $Q^*(A)$ and $Q^{**}(A)$, respectively. The Q range at a given A is the vertical distance between the two curves.

With back mutation

The previous model can be extended by back mutation, B is the probability that a replication of a mutant results in a master. The corresponding dynamics can be read as

$$\dot{x} = AQx + aBy - x(Ax + ay) \quad (19A)$$

$$\dot{y} = A(1-Q)x + a(1-B)y - y(Ax + ay), \quad (19B)$$

while its parabolic counterpart has the following form:

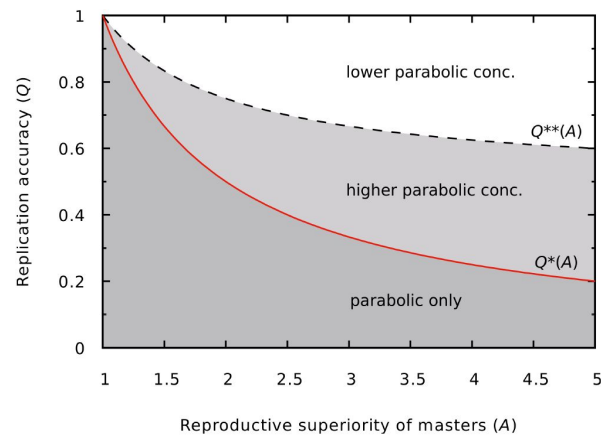
$$\dot{x} = AQ\sqrt{x} + aB\sqrt{y} - x(A\sqrt{x} + a\sqrt{y}) \quad (20A)$$

$$\dot{y} = A(1-Q)\sqrt{x} + a(1-B)\sqrt{y} - y(A\sqrt{x} + a\sqrt{y}). \quad (20B)$$

The non-trivial equilibrium master concentration of the exponential dynamics of [Equation 19A](#) and [Equation 19B](#) is

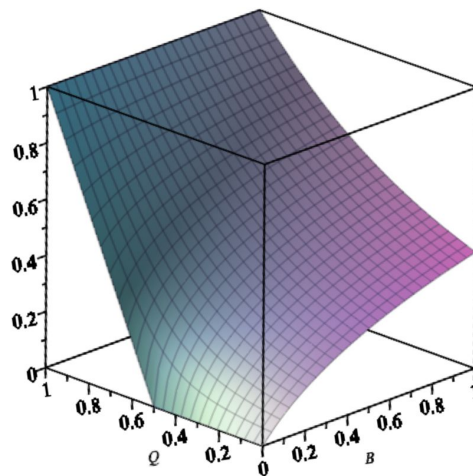
$$\hat{x}_{eB} = \frac{AQ - a(B+1) + \sqrt{[AQ - a(B+1)]^2 + 4a(A-a)B}}{2(A-a)}. \quad (21)$$

Due to back mutation, the error threshold disappears from the exponential system (as a simple consequence of the Perron–Forbenius theorem), if $B > 0$ the master concentration is always positive. [Appendix 1—figure 2](#) shows $\hat{x}_{eB}$ as the function of Q and B .



Appendix 1—figure 1.

Different outcomes of the dynamics as a function of Q and A . In the darker grey region (low copying fidelity, high reproductive superiority) the master survives only in the parabolic regime. In the lighter grey region (medium copying fidelity range) the master has a higher equilibrium concentration in the parabolic regime than in the exponential regime. The continuous red line is the error threshold for the exponential system (Q^*), the dashed black line divides the higher and lower parabolic concentration regions of the parameter space (Q^{**}). $a = 1$ was used.



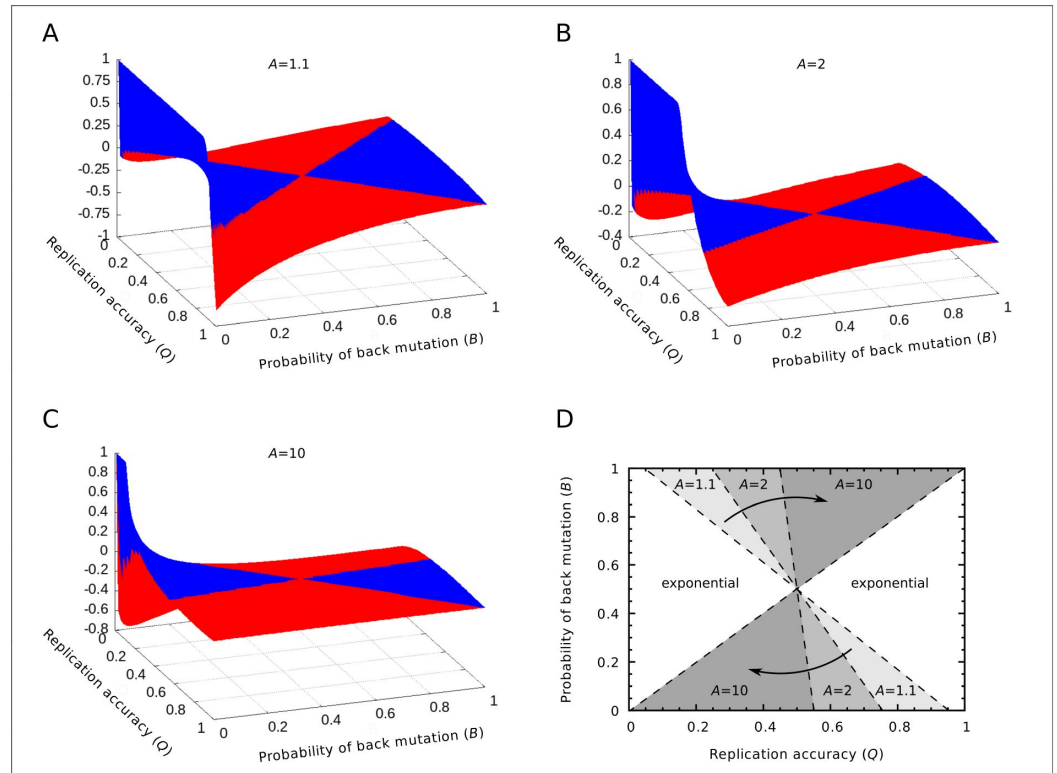
Appendix 1—figure 2.

Equilibrium concentration of the master in the exponential system with back mutation ($\hat{x}_{eB}$) as a function of Q and B . Parameters are: $A = 2$, $a = 1$.

As the analytical tractability of the equilibrium solution of [Equation 20A](#) and [Equation 20B](#) is limited, we carried out numerical simulations to analyze the concentrations of masters in both regimes ([Appendix 1—figure 3](#)). The equation of the delimiter line can be computed analytically by solving the $\hat{x}_{eB} = \hat{x}_{pB}$ equation for B with computer algebra ($\hat{x}_{pB}$ is the relevant fixed point of [Equation 20A](#) and [Equation 20B](#)). The result is:

$$B_1(Q) = Q \quad (22A)$$

$$B_2(Q) = -AQ + \frac{1}{2}(A + 1) \quad (22B)$$



Appendix 1—figure 3.

(A–C) The relative differences between the equilibrium concentrations of masters in the two regimes: $\frac{\hat{x}_p - \hat{x}_e}{\hat{x}_p}$ as the function of the replication accuracy (Q) and the probability of back mutation (B) at $A = 1.1, 2, 10$, respectively. Blue color indicates higher, red indicates lower parabolic concentrations. (D) The area of the parameter space where the parabolic regime has higher equilibrium concentration. The area shrinks as A increases. If $A \approx a$, the parabolic case has a higher concentration in the half of the parameter space, while if $A \gg a$, this area is reduced to the quarter of the space ($a = 1$ was used).

References

Attwater J., Raguram A., Morgunov A. S., Gianni E., Holliger P. (2018) **Ribozyme-catalysed RNA synthesis using triplet building blocks** *eLife* 7 <https://doi.org/10.7554/eLife.35255>

- Babajanyan S. G., Wolf Y. I., Khachatryan A., Allahverdyan A., Lopez-Garcia P., Koonin E. V. (2023) **Coevolution of reproducers and replicators at the origin of life and the conditions for the origin of genomes** *Proceedings of the National Academy of Sciences* **120** <https://doi.org/10.1073/pnas.2301522120>
- Bäck T., Fogel D. B., Michalewicz Z. (2000) **Evolutionary computation**
- Bada J. L. (2004) **How life began on Earth: A status report** *Earth and Planetary Science Letters* **226**:1–15 <https://doi.org/10.1016/j.epsl.2004.07.036>
- Benner S. A., Ellington A. D., Tauer A. (1989) **Modern metabolism as a palimpsest of the RNA world** *Proceedings of the National Academy of Sciences* **86**:7054–7058 <https://doi.org/10.1073/pnas.86.18.7054>
- Bernhardt H. S. (2012) **The RNA world hypothesis: The worst theory of the early evolution of life (except for all the others)** *Biology Direct* **7** <https://doi.org/10.1186/1745-6150-7-23>
- Briones C., Stich M., Manrubia S. C. (2009) **The dawn of the RNA World: Toward functional complexity through ligation of random RNA oligomers** *RNA* **15**:743–749 <https://doi.org/10.1261/rna.1488609>
- Cao Y., Li H., Petzold L. (2004) **Efficient formulation of the stochastic simulation algorithm for chemically reacting systems** *The Journal of Chemical Physics* **121**:4059–4067 <https://doi.org/10.1063/1.1778376>
- Crick F. H. C. (1968) **The origin of the genetic code** *Journal of Molecular Biology* **38**:367–379 [https://doi.org/10.1016/0022-2836\(68\)90392-6](https://doi.org/10.1016/0022-2836(68)90392-6)
- (1996) **A theory of evolution that includes prebiotic self-organization and episodic species formation** *Bulletin of Mathematical Biology* **58**:65–97 <https://doi.org/10.1007/BF02458282>
- Davis (1996) **Darwinian evolution by self-propagating species under fractional order kinetics** *Bulletin of Mathematical Biology* **58**:99–101 <https://doi.org/10.1007/BF02458283>
- (2000) **Darwinian Aspects of Molecular Evolution at Sublinear Propagation Rates** *Bulletin of Mathematical Biology* **62**:675–694 <https://doi.org/10.1006/bulm.2000.0173>
- Eigen M. (1971) **Selforganization of matter and the evolution of biological macromolecules** *Die Naturwissenschaften* **58**:465–523 <https://doi.org/10.1007/BF00623322>
- Epstein I. R. (1979) **Competitive coexistence of self-reproducing macromolecules** *Journal of Theoretical Biology* **78**:271–298 [https://doi.org/10.1016/0022-5193\(79\)90269-8](https://doi.org/10.1016/0022-5193(79)90269-8)
- Freier S. M., Kierzek R., Jaeger J. A., Sugimoto N., Caruthers M. H., Neilson T., Turner D. H. (1986) **Improved free-energy parameters for predictions of RNA duplex stability** *Proceedings of the National Academy of Sciences* **83**:9373–9377 <https://doi.org/10.1073/pnas.83.24.9373>
- Gause G. F. (1934) **The struggle for existence** <https://doi.org/10.5962/bhl.title.4489>
- Gilbert W. (1986) **Origin of life: The RNA world** *Nature* **319**:618–618 <https://doi.org/10.1038/319618a0>

- Gillespie D. T. (1976) **A general method for numerically simulating the stochastic time evolution of coupled chemical reactions** *Journal of Computational Physics* **22**:403–434 [https://doi.org/10.1016/0021-9991\(76\)90041-3](https://doi.org/10.1016/0021-9991(76)90041-3)
- Gillespie D. T. (1977) **Exact stochastic simulation of coupled chemical reactions** *The Journal of Physical Chemistry* **81**:2340–2361 <https://doi.org/10.1021/j100540a008>
- Hollander M., Wolfe D. A. (1999) **Nonparametric statistical methods**
- Issac R., Chmielewski J. (2002) **Approaching Exponential Growth with a Self-Replicating Peptide** *Journal of the American Chemical Society* **124**:6808–6809 <https://doi.org/10.1021/ja026024i>
- James K. D., Ellington A. D. (1997) **Surprising fidelity of template-directed chemical ligation of oligonucleotides** *Chemistry & Biology* **4**:595–605 [https://doi.org/10.1016/S1074-5521\(97\)90245-3](https://doi.org/10.1016/S1074-5521(97)90245-3)
- Jiménez J. I., Xulvi-Brunet R., Campbell G. W., Turk-MacLeod R., Chen I. A. (2013) **Comprehensive experimental fitness landscape and evolutionary network for small RNA** *Proceedings of the National Academy of Sciences* **110**:14984–14989 <https://doi.org/10.1073/pnas.1307604110>
- Joyce G. F. (2002) **The antiquity of RNA-based evolution** *Nature* **418**:214–221 <https://doi.org/10.1038/418214a>
- Le Vay K., Mutschler H. (2019) **The difficult case of an RNA-only origin of life** *Emerging Topics in Life Sciences* **3**:469–475 <https://doi.org/10.1042/ETLS20190024>
- Lee B. D. *et al.* (2023) **Mining metatranscriptomes reveals a vast world of viroid-like circular RNAs** *Cell* **186**:646–661 <https://doi.org/10.1016/j.cell.2022.12.039>
- Leu K., Kervio E., Obermayer B., Turk-MacLeod R. M., Yuan C., Luevano J.-M., Chen E., Gerland U., Richert C., Chen I. A. (2013) **Cascade of Reduced Speed and Accuracy after Errors in Enzyme-Free Copying of Nucleic Acid Sequences** *Journal of the American Chemical Society* **135**:354–366 <https://doi.org/10.1021/ja3095558>
- Leu K., Obermayer B., Rajamani S., Gerland U., Chen I. A. (2011) **The prebiotic evolutionary advantage of transferring genetic information from RNA to DNA** *Nucleic Acids Research* **39**:8135–8147 <https://doi.org/10.1093/nar/gkr525>
- Li X., Chmielewski J. (2003) **Peptide Self-Replication Enhanced by a Proline Kink** *Journal of the American Chemical Society* **125**:11820–11821 <https://doi.org/10.1021/ja036569s>
- Lifson S., Lifson H. (1999) **A Model of Prebiotic Replication: Survival of the Fittest versus Extinction of the Unfittest** *Journal of Theoretical Biology* **199**:425–433 <https://doi.org/10.1006/jtbi.1999.0969>
- Lincoln T. A., Joyce G. F. (2009) **Self-Sustained Replication of an RNA Enzyme** *Science* **323**:1229–1232 <https://doi.org/10.1126/science.1167856>
- Manrubia S. C., Briones C. (2007) **Modular evolution and increase of functional complexity in replicating RNA molecules** *RNA* **13**:97–107 <https://doi.org/10.1261/rna.203006>

- Martin W., Baross J., Kelley D., Russell M. J. (2008) **Hydrothermal vents and the origin of life** *Nature Reviews Microbiology* **6**:805–814 <https://doi.org/10.1038/nrmicro1991>
- Maynard Smith J., Szathmáry E. (1995) **The major transitions in evolution**
- Meszéna G., Szathmáry E. (2002) **Adaptive Dynamics of Parabolic Replicators** *Selection* **2**:147–159 <https://doi.org/10.1556/Select.2.2001.1-2.11>
- Mizuuchi R., Ichihashi N. (2018) **Sustainable replication and coevolution of cooperative RNAs in an artificial cell-like system** *Nature Ecology & Evolution* **2**:1654–1660 <https://doi.org/10.1038/s41559-018-0650-z>
- Moran P. A. P. (1958) **Random processes in genetics** *Mathematical Proceedings of the Cambridge Philosophical Society* **54**:60–71 <https://doi.org/10.1017/S0305004100033193>
- Müller U. F. (2006) **Re-creating an RNA world** *Cellular and Molecular Life Sciences* **63**:1278–1293 <https://doi.org/10.1007/s00018-006-6047-1>
- Orgel L. E. (1968) **Evolution of the genetic apparatus** *Journal of Molecular Biology* **38**:381–393 [https://doi.org/10.1016/0022-2836\(68\)90393-8](https://doi.org/10.1016/0022-2836(68)90393-8)
- Orgel L. E. (1992) **Molecular replication** *Nature* **358**:203–209 <https://doi.org/10.1038/358203a0>
- Orgel L. E. (2004) **Prebiotic Chemistry and the Origin of the RNA World** *Critical Reviews in Biochemistry and Molecular Biology* **39**:99–123 <https://doi.org/10.1080/10409230490460765>
- Patzke V., von Kiedrowski G. (2007) **Self replicating systems** *Arkivoc* **2007**:293–310 <https://doi.org/10.3998/ark.5550190.0008.522>
- Prywes N., Blain J. C., Del Frate F., Szostak J. W. (2016) **Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides** *eLife* **5** <https://doi.org/10.7554/eLife.17756>
- R Core Team (2023) **R: A Language and Environment for Statistical Computing**
- Rosenberger J. H., Göppel T., Kudella P. W., Braun D., Gerland U., Altaner B. (2021) **Self-Assembly of Informational Polymers by Templated Ligation** *Physical Review X* **11** <https://doi.org/10.1103/PhysRevX.11.031055>
- Ruiz-Mirazo K., Briones C., de la Escosura A. (2017) **Chemical roots of biological evolution: The origins of life as a process of development of autonomous functional systems** *Open Biology* **7** <https://doi.org/10.1098/rsob.170050>
- Scheuring I., Szathmáry E. (2001) **Survival of Replicators with Parabolic Growth Tendency and Exponential Decay** *Journal of Theoretical Biology* **212**:99–105 <https://doi.org/10.1006/jtbi.2001.2360>
- Schöning K.-U., Scholz P., Guntha S., Wu X., Krishnamurthy R., Eschenmoser A. (2000) **Chemical Etiology of Nucleic Acid Structure: The α -Threofuranosyl-(3' $\rightarrow$ 2') Oligonucleotide System** *Science* **290**:1347–1351 <https://doi.org/10.1126/science.290.5495.1347>
- Schuster P. (2011) **Mathematical modeling of evolution** *Solved and open problems. Theory in Biosciences* **130**:71–89 <https://doi.org/10.1007/s12064-010-0110-z>

- Sievers D., von Kiedrowski G. (1998) **Self-Replication of Hexadeoxynucleotide Analogues: Autocatalysis versus Cross-Catalysis** *Chemistry - A European Journal* **4**:629–641 [https://doi.org/10.1002/\(SICI\)1521-3765\(19980416\)4:4<629::AID-CHEM629>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1521-3765(19980416)4:4<629::AID-CHEM629>3.0.CO;2-0)
- Strazewski P. (2019) **The Beginning of Systems Chemistry** *Life* **9** <https://doi.org/10.3390/life9010011>
- Szathmáry E. (1991) **Simple growth laws and selection consequences** *Trends in Ecology & Evolution* **6**:366–370 [https://doi.org/10.1016/0169-5347\(91\)90228-P](https://doi.org/10.1016/0169-5347(91)90228-P)
- Szathmáry E., Gladkih I. (1989) **Sub-exponential growth and coexistence of non-enzymatically replicating templates** *Journal of Theoretical Biology* **138**:55–58 [https://doi.org/10.1016/S0022-5193\(89\)80177-8](https://doi.org/10.1016/S0022-5193(89)80177-8)
- Szilágyi A., Zachar I., Scheuring I., Kun Á., Könnnyű B., Czárán T. (2017) **Ecology and Evolution in the RNA World Dynamics and Stability of Prebiotic Replicator Systems** *Life* **7** <https://doi.org/10.3390/life7040048>
- Szilágyi A., Zachar I., Szathmáry E. (2013) **Gause's Principle and the Effect of Resource Partitioning on the Dynamical Coexistence of Replicating Templates** *PLoS Computational Biology* **9** <https://doi.org/10.1371/journal.pcbi.1003193>
- Szostak J. W. (2011) **An optimal degree of physical and chemical heterogeneity for the origin of life?** *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**:2894–2901 <https://doi.org/10.1098/rstb.2011.0140>
- Szostak J. W. (2012) **The eightfold path to non-enzymatic RNA replication** *Journal of Systems Chemistry* **3** <https://doi.org/10.1186/1759-2208-3-2>
- Szostak J. W., Bartel D. P., Luisi P. L. (2001) **Synthesizing life** *Nature* **409**:387–390 <https://doi.org/10.1038/35053176>
- Tjivikua T., Ballester P., Rebek J. (1990) **Self-replicating system** *Journal of the American Chemical Society* **112**:1249–1250 <https://doi.org/10.1021/ja00159a057>
- Varga Z., Szathmáry E. (1997) **An extremum principle for parabolic competition** *Bulletin of Mathematical Biology* **59**:1145–1154 <https://doi.org/10.1007/BF02460105>
- Vlassov A. V., Kazakov S. A., Johnston B. H., Landweber L. F. (2005) **The RNA World on Ice: A New Scenario for the Emergence of RNA Information** *Journal of Molecular Evolution* **61**:264–273 <https://doi.org/10.1007/s00239-004-0362-7>
- von Kiedrowski G. (1986) **A Self-Replicating Hexadeoxynucleotide** *Angewandte Chemie International Edition in English* **25**:932–935 <https://doi.org/10.1002/anie.198609322>
- von Kiedrowski G., Szathmáry E. (2001) **Selection versus Coexistence of Parabolic Replicators Spreading on Surfaces** *Selection* **1**:173–180 <https://doi.org/10.1556/Select.1.2000.1-3.17>
- von Kiedrowski G., Wlotzka B., Helbing J., Matzen M., Jordan S. (1991) **Parabolic Growth of a Self-Replicating Hexadeoxynucleotide Bearing a 3'-5'-Phosphoamidate Linkage** *Angewandte Chemie International Edition in English* **30**:423–426 <https://doi.org/10.1002/anie.199104231>

Wächtershäuser G. (1988) **Before enzymes and templates: Theory of surface metabolism** *Microbiological Reviews* **52**:452–484 <https://doi.org/10.1128/mr.52.4.452-484.1988>

Woese C. R. (1965) **On the evolution of the genetic code** *Proceedings of the National Academy of Sciences* **54**:1546–1552 <https://doi.org/10.1073/pnas.54.6.1546>

Zhang K., Hodge J., Chatterjee A., Moon T. S., Parker K. M. (2021) **Duplex Structure of Double-Stranded RNA Provides Stability against Hydrolysis Relative to Single-Stranded RNA** *Environmental Science & Technology* **55**:8045–8053 <https://doi.org/10.1021/acs.est.1c01255>

Zielinski W. S., Orgel L. E. (1987) **Autocatalytic synthesis of a tetranucleotide analogue** *Nature* **327**:346–347 <https://doi.org/10.1038/327346a0>

Article and author information

Mátyás Paczkó

HUN-REN Centre for Ecological Research, Institute of Evolution, Budapest, Hungary, Doctoral School of Biology, Institute of Biology, ELTE Eötvös Loránd University, Budapest, Hungary
ORCID iD: [0000-0002-0162-7927](https://orcid.org/0000-0002-0162-7927)

Eörs Szathmáry

HUN-REN Centre for Ecological Research, Institute of Evolution, Budapest, Hungary, Center for the Conceptual Foundations of Science, Parmenides Foundation, Pöcking, Germany, Department of Plant Systematics, Ecology and Theoretical Biology, Eötvös Loránd University, Budapest, Hungary
For correspondence: szathmary.eors@ecolres.hu
ORCID iD: [0000-0001-5227-2997](https://orcid.org/0000-0001-5227-2997)

András Szilágyi

HUN-REN Centre for Ecological Research, Institute of Evolution, Budapest, Hungary
ORCID iD: [0000-0002-6894-4652](https://orcid.org/0000-0002-6894-4652)

Copyright

© 2024, Paczkó et al.

This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editors

Reviewing Editor

Anne-Florence Bitbol

Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland

Senior Editor

Aleksandra Walczak

École Normale Supérieure - PSL, Paris, France

Reviewer #1 (Public Review):

Summary:

Szathmary and colleagues explore the parabolic growth regime of replicator evolution. Parabolic growth occurs when nucleic acid strain separation is the rate limiting step of the replication process which would have been the case for non-enzymatic replication of short oligonucleotide that could precede the emergence of ribozyme polymerases and helicases. The key result is that parabolic replication is conducive to the maintenance of genetic diversity, that is, coexistence of numerous master sequences (the Gause principle does not apply). Another important finding is that there is no error threshold for parabolic replication except for the extreme case of zero fidelity.

Strengths:

I find both the analytic and the numerical results to be quite convincing and well described. The results of this work are potentially important because they reveal aspects of a realistic evolutionary scenario for the origin of replicators.

Weaknesses:

There are no obvious technical weaknesses. It can be argued that the results represent an incremental advance because many aspects of parabolic replication have been explored previously (the relevant publications are properly cited). Obviously, the work is purely theoretical, experimental study of parabolic replication is due. In the opinion of this reviewer, though, these are understandable limitations that do not actually detract from the value of this work.

<https://doi.org/10.7554/eLife.93208.2.sa1>

Reviewer #2 (Public Review):

Summary:

A dominant hypothesis concerning the origin of life is that, before the appearance of the first enzymes, RNA replicated non-enzymatically by templating. However, this replication was probably not very efficient, due to the propensity of single strands to bind to each other, thus inhibiting template replication. This phenomenon, known as product inhibition, has been shown to lead to parabolic growth instead of exponential growth. Previous works have shown that this situation limits competition between alternative replicators and therefore promotes RNA population diversity. The present work examines this scenario in an agent-based model of RNA replication, taking into account finite population size, mutations and differences in GC content. The main results are (1) confirmation that parabolic growth promotes diversity, but that when the population size is small enough, sequences least efficient at replicating may nevertheless go extinct; (2) the observation that fitness is not only controlled by the replicability of sequences, but also by their GC content ; (3) the observation that parabolic growth attenuates the impact of mutations and, in particular, that the error threshold to which exponentially growing sequences are subject can be exceeded, enabling sequence identity to be maintained at higher mutation rates.

Strengths:

The analyses are sound and the observations intriguing. Indeed, while it has been noted previously that parabolic growth promotes coexistence, this is the first work to show that it can also mitigate the error threshold catastrophe, which is often presented as a major obstacle to our understanding of the origin of life.

Weaknesses:

A general weakness, which can however be seen as inherent in an agent-based model that aims to be more realistic than earlier, more phenomenological models, is the proliferation of parameters. The choice and values of these parameters are generally justified and, in many cases, several values are tested to assess the robustness of the results, but it can be difficult for the reader to identify the modeling choices that are truly critical from those that are less so.

<https://doi.org/10.7554/eLife.93208.2.sa0>

Author Response

The following is the authors' response to the original reviews.

General remarks for the Editor and the Reviewers

We would like to thank the Editor and the Reviewers for their feedback. Below we address their comments and present our point-by-point responses as well as the related changes in the manuscript.

In addition to these changes, in a few cases we have found it necessary to move some texts and provide some additional explanations within the manuscript. We emphasize that these amendments have been made for only technical reasons, and do not alter the results and conclusions of the paper, but may help to render the text more coherent and understandable to readers with little knowledge of the subject.

These minor corrections are:

- We extended the Introduction section by a sentence (lines 40-42) that is intended to fit the proposed template directed, non-enzymatic replication mechanism into a more general prebiotic evolutionary context, thus emphasizing its biological relevance. This sentence includes an additional reference (Rosenberger et al., 2021).
- Two very methodologically oriented and repeated descriptions of random sequence generation have been moved to the Methods section (lines 178-185) from the Results section (lines 336-339 and lines 351-354).
- We complemented the Data availability statement with licensing information (lines 684-685).
- Further minor changes (also indicated by red texts) have been implemented to remedy logical and grammatical glitches.

Public Reviews:

Reviewer #1 (Public Review):

Summary:

Szathmáry and colleagues explore the parabolic growth regime of replicator evolution. Parabolic growth occurs when nucleic acid strain separation is the rate-limiting step of the replication process which would have been the case for non-enzymatic replication of short oligonucleotide that could precede the emergence of ribozyme polymerases and helicases. The key result is that parabolic replication is conducive to the maintenance of genetic diversity, that is, the coexistence of numerous master sequences (the Gause

principle does not apply). Another important finding is that there is no error threshold for parabolic replication except for the extreme case of zero fidelity.

Strengths:

I find both the analytic and the numerical results to be quite convincing and well-described. The results of this work are potentially important because they reveal aspects of a realistic evolutionary scenario for the origin of replicators.

Weaknesses:

There are no obvious technical weaknesses. It can be argued that the results represent an incremental advance because many aspects of parabolic replication have been explored previously (the relevant publications are properly cited). Obviously, the work is purely theoretical, experimental study of parabolic replication is due. In the opinion of this reviewer, though, these are understandable limitations that do not actually detract from the value of this work.

We are grateful that this Reviewer appreciates our work. We completely agree that the ultimate validation must come from experiments. It is important to stress that in this field theory often preceded experimental work by decades, and the former often guided the latter. We hope that for the topic of the present paper experiments will follow considerably faster.

Reviewer #2 (Public Review):

Summary:

A dominant hypothesis concerning the origin of life is that, before the appearance of the first enzymes, RNA replicated non-enzymatically by templating. However, this replication was probably not very efficient, due to the propensity of single strands to bind to each other, thus inhibiting template replication. This phenomenon, known as product inhibition, has been shown to lead to parabolic growth instead of exponential growth. Previous works have shown that this situation limits competition between alternative replicators and therefore promotes RNA population diversity. The present work examines this scenario in a model of RNA replication, taking into account finite population size, mutations, and differences in GC content. The main results are (1) confirmation that parabolic growth promotes diversity, but that when the population size is small enough, sequences least efficient at replicating may nevertheless go extinct; (2) the observation that fitness is not only controlled by the replicability of sequences, but also by their GC content; (3) the observation that parabolic growth attenuates the impact of mutations and, in particular, that the error threshold to which exponentially growing sequences are subject can be exceeded, enabling sequence identity to be maintained at higher mutation rates.

Strengths:

The analyses are sound and the observations are intriguing. Indeed, it has been noted previously that parabolic growth promotes coexistence, its role in mitigating the error threshold catastrophe - which is often presented as a major obstacle to our understanding of the origin of life - had not been examined before.

Weaknesses:

Although all the conclusions are interesting, most are not very surprising for people familiar with the literature. As the authors point out, parabolic growth is well known to promote diversity (Szathmari & Gladkih 89) and it has also been noted previously that a form of Darwinian selection can be found at small population sizes (Davis 2000).

Given that under parabolic growth, no sequence is ever excluded for infinite populations, it is also not surprising to find that mutations have a less dramatic exclusionary impact.

In the two articles cited (Szathmary-Gladkih 1989 and Davis 2000) the subexponentiality of the system was implemented in a mechanistic way, by introducing the exponent $0 < p < 1$. Although the behaviour of these models is more or less consistent with experimental findings (von Kiedrowski, 1986; Zielinski and Orgel, 1987), the divergence of per capita growth rates ($\dot{x}/x$) at very low concentrations—which guarantees the ability to maintain unlimited diversity in the case of infinite population sizes—makes this formal approach partly unrealistic.

To avoid the possible artefacts of this mechanistic approach, and as there are no previous studies analysing the diversity maintaining ability of finite populations of parabolic replicators in an individual-based model context, we implemented a simplified template replication mechanism leading to parabolic growth and analysed the dynamics in an individual-based stochastic model context. The key point of our investigation is that considerable diversity can be maintained in the system even when the population size is quite small.

Regarding the Reviewer's comment on selection: Darwinian selection can only occur in a simple subexponential dynamics if the ratio of replicabilities diverges, cf. Eq. (8) and the preceding paragraph in Davis, 2000.

Our results also show (Figs. 4B and 4C) that high mutation rates and the error threshold problem can still be considered as a major limiting factor for parabolically replicating systems in terms of their diversity-maintaining ability. In the light of the above, potential mechanisms to relax the error threshold in such systems, one of which is demonstrated in the present study, seem to be important steps to account for the sequence diversification and increase in molecular complexity during the early evolution of RNA replicators.

A general weakness is the presentation of models and parameters, whose choices often appear arbitrary. Modeling choices that would deserve to be further discussed include the association of the monomers with the strands and the ensuing polymerization, which are combined into a single association/polymerization reaction (see also below), or the choice to restrict to oligomers of length $L = 10$. Other models, similar to the one employed here, have been proposed that do not make these assumptions, e.g. Rosenberger et al. Self-Assembly of Informational Polymers by Templated Ligation, PRX 2021. To understand how such assumptions affect the results, it would be helpful to present the model from the perspective of existing models.

The assumption of one-step polymerization reactions that we used here is a common technique for modelling template replication of sequence-represented replicators [see, e.g., Fontana and Schuster, 1998 (10.1126/science.280.5368.1451), Könnyű et al., 2008 (10.1186/1471-2148-8267), Vig-Milkovics et al, 2019 (10.1016/j.jtbi.2018.11.020) or Szilágyi et al., 2020 (10.1371/journal.pgen.1009155)]. This is because assuming base-to-base polymerisation of the copy would lead to a very large number of different types of intermediates, which a Gillespie-type stochastic simulation algorithm could not handle in reasonable computation times, even if the sequences were relatively short. For comparison, in our model, where polymerization is one-step, the characteristic time of a simulation for $L = 10$, $N = 105$ and $\delta = 0.01$ was 552 hours.

Note that in Rosenberg et al. (PRX 2021), in contrast to a pioneering work [Fernando et al, 2007 (10.1007/s00239-006-0218-4)], sequences of replicators are not represented, which makes this approach completely inapplicable to our case, in which sequence defines the fitness. In sum, we suggest that this valid criticism points to possible future work.

The values of the (many) parameters, often very specific, also very often lack justifications. For example, why is the "predefined error factor" $\varepsilon = 0.2$ and not lower or higher? How would that affect the results?

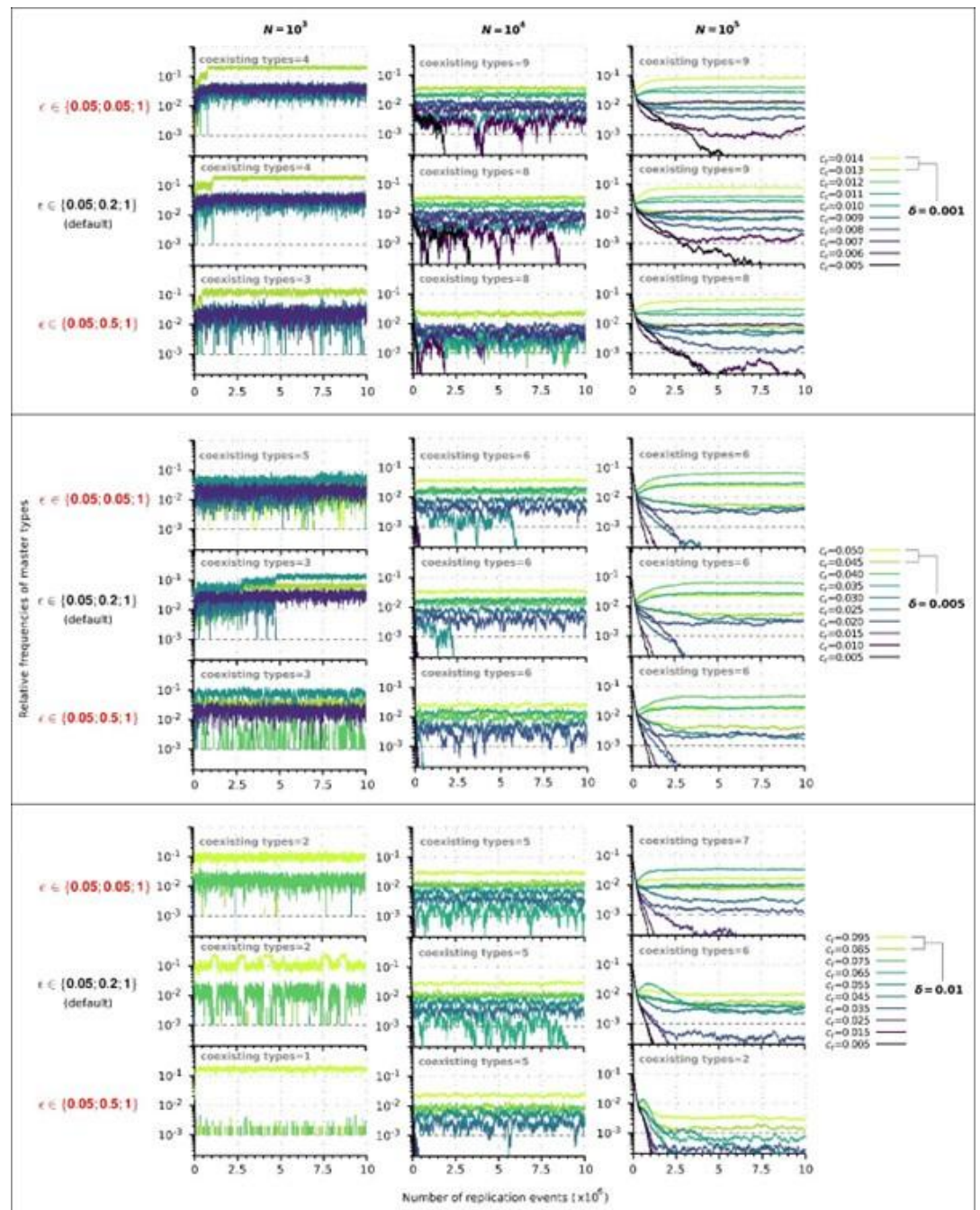
A general remark. For the more important parameters $(N, \delta, p_{mut}, f, c_{in}, c_{\delta ds}, c_{out}, c_r, c_d)$,

several values were used to test the behaviour of the model (see Table 1), but due to the considerable number of parameters, it is impossible to examine all possible combinations. $c^+ = 1$ fixes the timescale, L is set to 10 to obtain reasonable running times (see above).

ε characterizes how replicability decreases as the number of mutations increases. In the manuscript we used the following default vector: $\varepsilon = (0.05, 0.2, 1)$ in which the third element corresponds to the mutation-free sequence, so it must to be 1. The first element determines the baseline replicability (see Methods), which we preferred not to change because it would fundamentally alter the ratio of replication propensities to association and dissociation propensities (as the substantial amount of complementary sequences of the master sequences are of baseline replicability) and thus would alter the reaction kinetics to an extent that it is not comparable with the original results. Therefore, only the second element can be adjusted. Accordingly, we have analysed the behaviour of the model in the cases of a steeper and a more gradual loss of replicability using the following two vectors, respectively: $\varepsilon_s = (0.05, 0.05, 1)$ and $\varepsilon_g = (0.05, 0.5, 1)$. The choice of ε_s is chemically more plausible, since for very short oligomers the loss of chemical activity and replicability as a function of the number of mutations can be very sharp. We performed a series of simulations with all possible combinations of $\delta = 0.001, 0.005, 0.1$ and $N = 103, 104, 105$ for ε_s and ε_g in the constant population and chemostat model context (36 different runs). For other parameters, we took the default values, see Table 1. These values also correspond to the parameters we used in Figures 2 and 6. The results show that the steeper loss of replicability (ε_s) slightly increases the diversity maintaining ability of the system, whereas the more gradual loss of replicability (ε_g) moderately decreases the diversity-maintaining ability of the system, and that these shifts are more pronounced in the constant population size model (Author response image 1) than in the chemostat model (Author response image 2). Altogether, these results confirm that the qualitative outcome of the model is robust in a wide range of loss of replicability (ε vector) values.

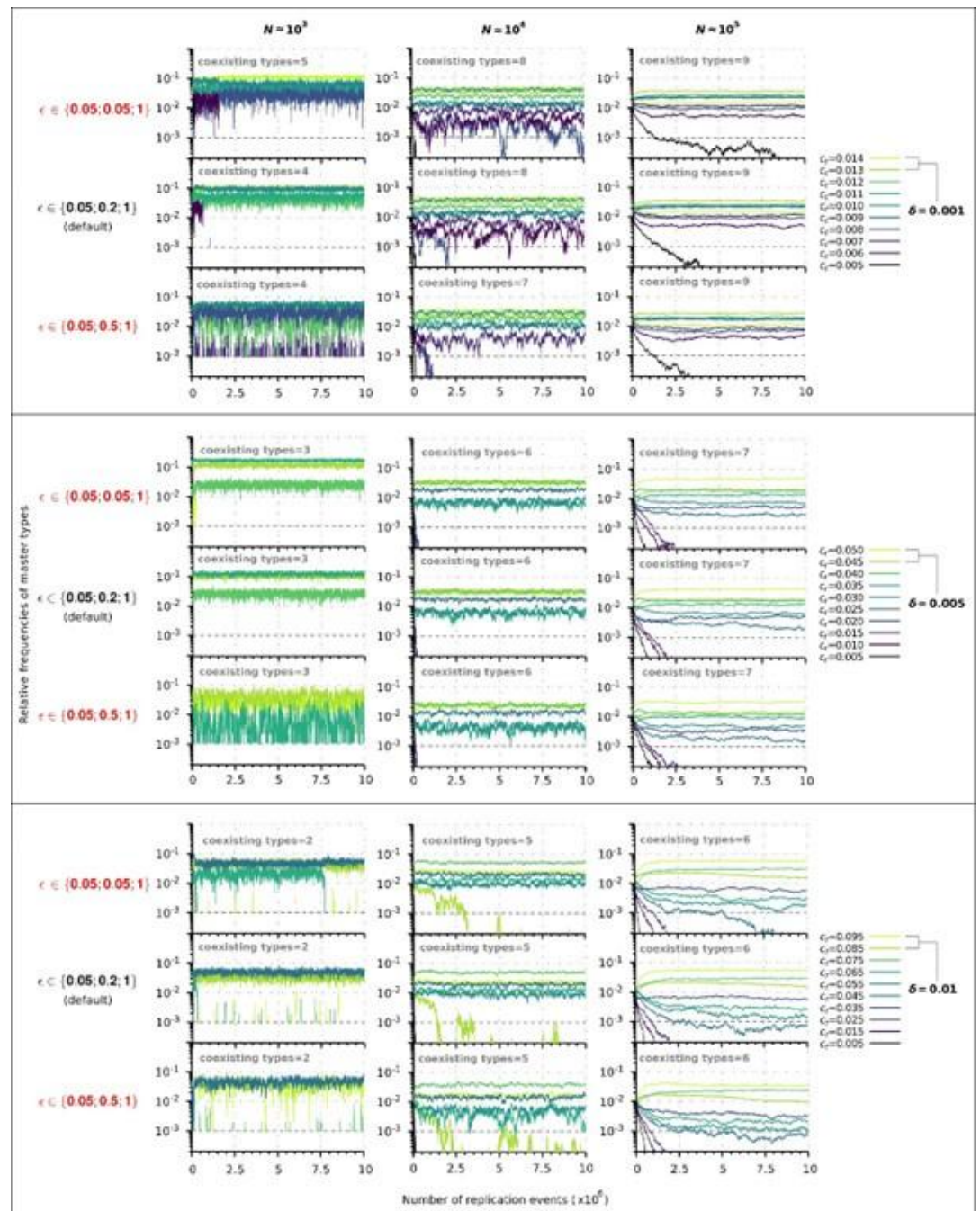
Author response image 1.

Replicator coexistence in the constant population model with different loss of replicability (ε vector) values. Within a given combination of δ and N parameter values, the upper panel corresponds to the steeper loss of replicability (ε_s), the middle panel to the default ε vector (Figure 2A), and the bottom panel to the more gradual loss of replicability vector (ε_g). Within each $\delta; N$ parameter combination, the same master sequence set was used with the three different ε vectors for comparability.



Author response image 2.

Replicator coexistence in the chemostat model with different loss of replicability (ϵ vector) values. Within a given combination of δ and N parameter values, the upper panel corresponds to the steeper loss of replicability (ϵ !), the middle panel to the default ϵ vector (Figure 6A), and the bottom panel to the more gradual loss of replicability vector (ϵ !). Within each δ ; N parameter combination, the same master sequence set was used with the three different ϵ vectors for comparability.



Similarly, in equation (11), where does the factor 0.8 come from?

This factor scales the decay rate of duplex sequences (c'') as the function of the binding energy

(Eb). The value of 0.8 is an arbitrary choice, the value should be in the interval (0,1) and is only relevant in the chemostat model. It is expected to have a similar effect on the dynamics as the duplex decay factor parameter f , which we have investigated in a wide range of different values (cf. Table 1, Fig. 6), although f is independent of the binding energy (E): increasing/decreasing the 0.8 factor is expected to decrease/increase the average total population size. We have investigated the diversity maintaining ability of the system at smaller (0.6) and larger (0.9) parameter values at different population sizes ($N \approx 103$, 104 and

105) and at different replicability distances ($\delta = 0.001, 0.005$ and 0.01) as shown in Fig. 6. We have found that the number of coexisting master types changes very little in response to changes in this factor. Only two shifts could be detected (underlined): factor 0.9 combined with $N \approx 104$ and $\delta = 0.001$ caused the number of surviving master types to decrease by one, while factor 0.9 combined with $N \approx 103$ and $\delta = 0.01$ caused the number of surviving master types to increase by one (Author response table 1). Factor 0.6 produced the same number of surviving types as the default (Author response table 1). In summary, the model shows marked robustness to changes in the values of this parameter.

Author response table 1.

Number of coexisting master types in the chemostat model with different binding energy dependent duplex decay rates. Within each δ ; N parameter combination, the same master sequence set was used with the three different factor values: 0.6, 0.8 (the original) and 0.9 for comparability.

	$N \approx 10^3$	$N \approx 10^4$	$N \approx 10^5$	
$c_{eq} = f \cdot c_{min} \cdot 0.6^{\delta_{ss}}$	4	8	9	$\delta = 0.001$
$c_{eq} = f \cdot c_{min} \cdot 0.8^{\delta_{ss}}$ (default)	4	8	9	
$c_{eq} = f \cdot c_{min} \cdot 0.9^{\delta_{ss}}$	4	<u>7</u>	9	
$c_{eq} = f \cdot c_{min} \cdot 0.6^{\delta_{ss}}$	3	6	7	$\delta = 0.005$
$c_{eq} = f \cdot c_{min} \cdot 0.8^{\delta_{ss}}$ (default)	3	6	7	
$c_{eq} = f \cdot c_{min} \cdot 0.9^{\delta_{ss}}$	3	6	7	
$c_{eq} = f \cdot c_{min} \cdot 0.6^{\delta_{ss}}$	2	5	6	$\delta = 0.01$
$c_{eq} = f \cdot c_{\delta_{ss}} \cdot 0.8^{E_b}$ (default)	2	5	6	
$c_{eq} = f \cdot c_{min} \cdot 0.9^{\delta_{ss}}$	<u>3</u>	5	6	

Why is the kinetic constant for duplex decay reaction $1.15e10-8$?

Note that this value is the minimum of the duplex decay rate, Table 1 correctly shows the interval of this kinetic constant as: $[1.15 \cdot 10^{-8}, 6.4 \cdot 10^{-5}]$. Both values are derived from the basic parameters of the system ($f, c_{\delta_{ss}}$ and E_b) and can be computed according to Eq. (11). The

minimum: $c_{\delta_{ss}} = f \cdot c_{\delta_{ss}} \cdot 0.8^{E_b} = 0.01 \cdot 10^{-4} \cdot 0.8^{-20} = 1.15 \cdot 10^{-8}$ as the parameter set

corresponding to this value is: $f = 0.01, c_{\delta_{ss}} = 10^{-4}$ and $E_b = 20$. The maximum:

$c_{\delta_{ss}} = f \cdot c_{\delta_{ss}} \cdot 0.8^{E_b} = 1 \cdot 10^{-4} \cdot 0.8^2 = 6.4 \cdot 10^{-5}$ with $f = 1, c_{\delta_{ss}} = 10^{-4}$ and $E_b = 2$.

Are those values related to experiments, or are they chosen because specific behaviors can happen only then?

See above.

The choice of the model and parameters potentially impact the two main results, the attenuation of the error threshold and the role of GC content:

Regarding the error threshold, it is also noted (lines 379-385) that it disappears when back mutations are taken into account. This suggests that overcoming the error threshold might not be as difficult as suggested, and can be achieved in several ways, which calls into question the importance of the particular role of parabolic growth. Besides, when the concentration of replicators is low, product inhibition may be negligible, such that a "parabolic replicator" is effectively growing exponentially and an error catastrophe may occur. Do the authors think that this consideration could affect their conclusion? Can simulations be performed?

The assumption of back mutation only provides a theoretical solution to the error threshold problem: back mutation guarantees a positive (non-zero) concentration of a master type, but, since the probability of back mutation is generally very low, this equilibrium concentration may be extremely low, or negligible for typical system sizes. Consequently, back mutation alone does not solve the problem of the error catastrophe: in our system back mutation is present (the probability that a sequence with k errors mutates back to a master sequence is $\mu k(1-\mu)^{L-k}$), and the diversity-maintaining ability is limited. The effect of back mutation decreases exponentially with increasing sequence length.

Regarding the role of the GC content, GC-rich oligomers are found to perform the worst but no rationale is provided.

For GC-rich oligonucleotides the dissociation probability of a template-copy complex is relatively low (cf. Eqs. (9, 10)), thus they have a relatively low number of offspring, cf. lines 557-561: "a relatively high dissociation probability and the consequential higher propensity of being in a simple stranded form provides an advantage for sequences with relatively low GC content in terms of their replication affinity, that is, the expected number of offspring in case of such variants will be relatively high." Note that the simulation results shown in Fig. 3A, demonstrate the realization of this effect with prepared sequences (along a GC content gradient).

One may assume that it happens because GC-rich sequences are comparatively longer to release the product. However, it is also conceivable that higher GC content may help in the polymerization of the monomers as the monomers attach longer on the template (as described in Eq. (9)). This is an instance where the choice to pull into a single step the association and polymerization reactions are pulled into a single step independent of GC content may be critical.

It would be important to show that the result arises from the actual physics and not from this modeling choice.

Some more specific points that would deserve to be addressed:

- *Line 53: it is said that p "reflects how easily the template-reaction product complex dissociates". This statement is not correct. A reaction order $p < 1$ reflects product inhibition, the propensity of templates to bind to each other, not slow product release. Product release can be limiting, yet a reaction order of 1 can be achieved if substrate concentrations are sufficiently high relative to oligomer concentrations (von Kiedrowski et al., 1991).*

We think the key reference is Von Kiedrowski (1993) in this case. Other things being equal, his Table 1 on p. 134 shows that a sufficient increase in K_4 , i.e., the stability of the duplex

(template and copy) (association rate divided by dissociation rate) throws the system into the parabolic regime. This is what we had in mind. In order to clarify this, we modified the quoted sentence thus: “In this kinetics, the growth order is equal or close to 0.5 (i.e., the dynamics is sub-exponential) because increased stability of the template-copy complex (rate of association divided by dissociation) promotes parabolic growth (von Kiedrowski et al., 1991; von Kiedrowski & Szathmáry, 2001).”

- *Population size is a key parameter, and a comparison is made between small (10^3) and large (10^5) populations, but without explaining what determines the scale (small/large relative to what?).*

The “small” value (103) corresponds to the smallest meaningful population size, significantly smaller population sizes (e.g. 102) cannot maintain the 10 master types (or any subset of them) and are chemically unrealistic. The “large value” (105) is the largest population size for which simulation times are still acceptable, in the case of 106 the runtimes are in the order of months.

- *In the same vein, we might expect size not to be the only important parameter, but also concentration.*

With constant volume population size and concentration are strictly coupled.

- *Lines 543-546: if understanding correctly, the quantitative result is that the error threshold rises from 0.1 in the exponential case to 0.196 in the parabolic. Are the authors suggesting that a factor of 2 is a significant difference?*

In this paragraph we compared the empirical error threshold of our system (which is close to $p_{\text{mut}} = 0.15$) with the error threshold of the well-known single peak fitness landscape (which

can be approximated by $\frac{1}{L} = 0.1, (L = 10)$ as a reference case. To make the message even clearer

we have extended the last sentence (lines 596-597) as follows: “but note that applying this approach to our system is a serious oversimplification”. The 0.196 is simply the probability of error-free replication of a sequence when $p_{\text{mut}} = 0.15$: $Q = (1 - p_{\text{mut}})^L = 0.196$, but we have

removed this sentence (“corresponding to the $Q = (1 - 0.15)^{10} = 0.196$ replication accuracy of a master sequence”) from the manuscript as it seems to be confusing.

- *Figure 3C: this figure shows no statistically significant effect?*

Thank you for pointing out this. We statistically tested the hypothesis that the GC content between the survived and the extinct master subsets are different. This analysis revealed that the differences between these two groups are statistically significant, which we now included in the manuscript at lines 380-390: “A direct investigation of whether the sequence composition of the master types is associated with their survival outcome was conducted using the data from the constant population model simulation results (Figure 2). In these data, the average GC content was measured to be lower in the surviving master subpopulations than in the extinct subpopulations (Figure 3C). To determine whether this difference was statistically significant, nonparametric, two-sample Wilcoxon rank-sum tests (Hollander & Wolfe, 1999) were performed on the GC content of the extinct-surviving master subsets. The GC content was significantly different between these two groups in all nine

investigated parameter combinations of population size (N) and replicability distance (δ) at $p < 0.05$ level, indicating a selective advantage for a lower GC content in the constant population model context. The exact p values obtained from this analysis are shown in Figure 3C.”

- line 542: "phase transition-like species extension (Figure 4B)": such a clear threshold is not apparent.

Thank you for pointing out the incorrect phrasing. As there is no clear threshold in the number of coexisting types as a function of the mutation rate, we removed the “phase transition-like” expression: “However, when finite population sizes and stochastic effects are taken into account, at the largest investigated per-base mutation rate ($\rho_{\text{mut}} = 0.15$), the summed relative steady-state master frequencies approach zero (Figure 4C) with accelerating species extinction (Figure 4B), indicating that this value is close to the system’s empirical error threshold.” (lines 589-594).

Recommendations for the authors:

Reviewer #1 (Recommendations For The Authors):

On the whole, the work is well done and presented, there are no major recommendations. It seems a good idea to cite and briefly discuss this recent paper: <https://pubmed.ncbi.nlm.nih.gov/36996101/> which develops a symbiotic scenario of the coevolution of primordial replicators and reproducers that appears to be fully compatible with the results of the current work.

Thank you for bringing this article to our attention. We have inserted the following sentence at lines 621-624: “The demonstrated diversity-maintaining mechanism of finite parabolic populations can be used as a plug-in model to investigate the coevolution of naked and encapsulated molecular replicators (e.g., Babajanyan et al., 2023).”

The manuscript is well written, but there are some minor glitches that merit attention. For example:

l. 5 "carriers presents a problem, because product formation and mutual hybridization" - "mutual" is superfluous here, delete

l. 13 "amplification. In addition, sequence effects (GC content) and the strength of resource" - hardly "effects" - should be 'features' or 'properties'

l. 41 "If enzyme-free replication of oligomer modules with a high degree of sequence" - "modules" here is only confusing - simply, "oligomers"

l. 44 "under ecological competition conditions with which distinct replicator types with different" - delete "with" etc, there are many such minor glitches that are best corrected.

Thank you for pointing out, we have corrected! Other drafting errors, glitches, superfluous sentences have also been corrected.

None

Reviewer #2 (Recommendations For The Authors):

None

Editor (Recommendations For The Authors):

In the manuscript, it appears that coexistence is assessed at a given point in time, while figures seem to show that it remains time-dependent. It would be great if the authors could clarify this and/or discuss this.

We appreciate you bringing this to our attention, as we have indeed missed to elaborate on this important point. The steady state characteristic of the coexistence is assessed in our model in the following way: the relative frequency of each master sequence is tested for the condition of ≥ 100 - (cut-off relative frequency for survival) in every 2,000th replication step in the interval between 10,000 replication steps before termination and actual termination (10= replication steps). If the above condition is true more than once, we consider the master type in question as survived (we have included this explanation in the Methods section: lines 258-268). Although this relatively narrow time interval can still be regarded as a snapshot of the state of the system, according to our numerical experiences, the resulting measure is a reliable quantitative indicator of the apparent stability of species coexistence in the parabolic dynamics.