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

Authors: Mátyás Paczkó[1,2], Eörs Szathmáry[1,3,4]*, András Szilágyi[1]

Affiliations:

[1] HUN-REN Centre for Ecological Research, Institute of Evolution, Budapest, Hungary

[2] Doctoral School of Biology, Institute of Biology, ELTE Eötvös Loránd University, Budapest, Hungary

[3] Center for the Conceptual Foundations of Science, Parmenides Foundation, Pöcking, Germany

[4] Department of Plant Systematics, Ecology and Theoretical Biology, Eötvös Loránd University, Budapest, Hungary

* Correspondence: szathmary.eors@ecolres.hu

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 mutual 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 found that large population sizes and small differences in the replication rates favor the coexistence of the vast majority of replicator species ("genes"), while the error-threshold problem is alleviated relative to exponential amplification. In addition, sequence effects (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.

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 those concerning 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 remarkable concept among these suggests that prior to template replication of complex polymers catalyzed by an RNA polymerase ribozyme (i.e., replicase), genetic information and catalytic functions were distributed among short sequence modules that could be occasionally 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, although they may have already showed 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). Consistent with this, 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 oligomer modules with a high degree of sequence variability were attainable from prebiotic chemistry, then it becomes a vital issue how a critical level of diversity of the associated genetic information (Szostak, 2011) would be preserved under ecological competition conditions with which distinct replicator types with different competitive abilities (e.g., replicabilities) inevitably encountered to some extent (Eigen, 1971; Szathmáry, 1991; Maynard Smith & Szathmáry, 1995; Jiménez et al., 2013; Ruiz-Mirazo et al., 2017). This recognition highlighted the importance of deciphering the coexistence mechanisms that can alleviate the competition among independently replicating information carrying modules in such systems. 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, which reflects how easily the template-reaction product complex dissociates, is equal or close to 0.5 (i.e., the dynamics is sub-exponential) because dissociation is rate-limiting for the reaction product formation (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 alleviate 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 selectivity of the dynamics (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; an error rate of the replication mechanism at which the system's ability to propagate genetic information from generation to generation is ceased 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 (Manrubia & Briones, 2007; Leu et al., 2011).

Moreover, practically no study is known to have focused on parabolic replication in a manner that physicochemical details and ecological constraints together with stochastic effects, which 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 embedded into a stochastic dynamics. We individually represent the replicator sequences and copying error as well as energetic constraints in strand separation are also taken into account 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, while sequence effects are also taken into consideration. Then we 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 the function of the copying fidelity and the reproductive superiority of the master sequences relative to the mutants. In this analysis, we also compare the error threshold of parabolic dynamics to that of the exponential dynamics; an extensively used reference model for investigating the genetic information maintaining ability of early replicator systems in the context of copying fidelity. 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, where 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 variant.

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) structure and are assumed to be involved in a template-directed, non-enzymatic replication mechanism. 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 (see 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 higher fitness relative to mutant genotypes, which we refer to as "master" types, 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 1AB*).

Constant population model of parabolic replication

In the constant population model (*Figure 1A*), there is no exchange of materials with the

environment (monomers are not represented explicitly) and a constant total population size is ensured by implementing a Moran process into the replication mechanism (Moran, 1958).

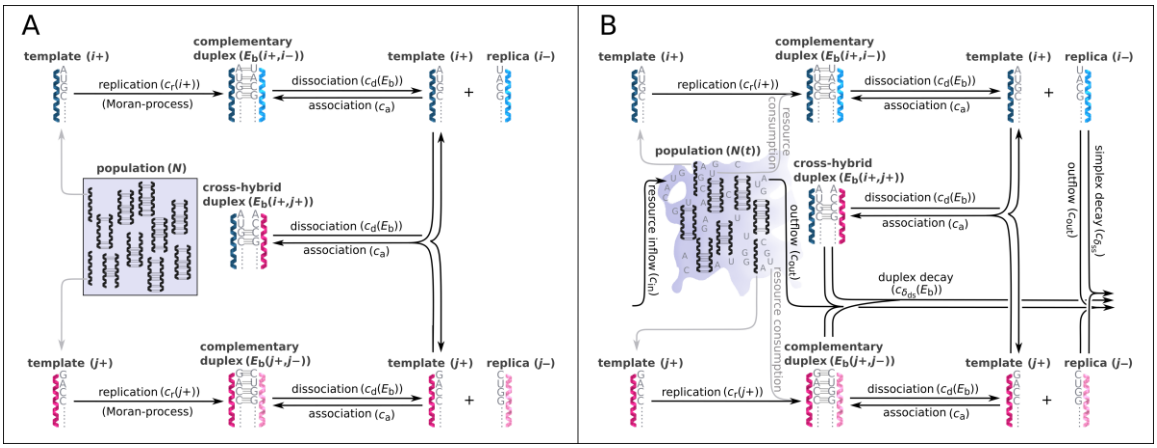


Figure 1. Schematic representation of two parabolically replicating systems. (**A**, **B**) These systems represent spatially constrained or aggregated, but well-mixed (for example by wet-dry cycles on mineral surfaces) replicator populations. Black arrows indicate the possible chemical reactions, i and j denote master types. (**A**) Constant population model representing a Moran-process. N denotes the (fixed) total number of replicator molecules. This model allows us to investigate the inherent dynamical properties of parabolic growth, such as its diversity maintaining ability in face of mutations, under fixed and controlled conditions. (**B**) Chemostat model representing an open system. $N(t)$ denotes the time dependent total number of replicators. This model allows us to investigate the effects of the changes in prebiotic environments on parabolic coexistence. Such alternations in environmental conditions could have involved, for example, variations in resource availability and the consequential shifts in the strength of resource competition. For the explanation of mechanisms and constants, see main text and *Table 1*.

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_iS_j (*Equation 3A*). The rate constants corresponding to these three reactions are $c_\mathrm{r}, c_\mathrm{a}, c_\mathrm{d}$, 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.

$$S_i(+n \cdot R) \xrightarrow{c_\mathrm{r}} S_i S_j \tag{1A}$$

$$S_i + S_j \xrightarrow{c_\mathrm{a}} S_i S_j \tag{2A}$$

$$S_i S_j \xrightarrow{c_\mathrm{d}} S_i + S_j \tag{3A}$$

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_{δ_ss} and c_{δ_ds}, respectively, see *Equation 5B* and *Equation 6B*. The type independent outflow of molecular species is characterized by c_out (*Equation 7B*).

$$S_i + n \cdot R \xrightarrow{c_\mathrm{r}} S_i S_j \tag{1B}$$

$$S_i + S_j \xrightarrow{c_\mathrm{a}} S_i S_j \tag{2B}$$

$$S_i S_j \xrightarrow{c_\mathrm{d}} S_i + S_j \tag{3B}$$

$$\xrightarrow{c_\mathrm{in}} R_{(\mathrm{A,U,G,C})} \tag{4B}$$

$$S_i \xrightarrow{c_{\delta_\mathrm{ss}}} \emptyset \tag{5B}$$

$$S_i S_j \xrightarrow{c_{\delta_{\mathrm{ds}}}} \emptyset \tag{6B}$$

$$S_i S_j, S_i, R_{(\mathrm{A,U,G,C})} \xrightarrow{c_{\mathrm{out}}} \emptyset \tag{7B}$$

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 a source-sink type dynamics.

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

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

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: $c_{\mathrm{r}} = \phi$. Substitution mutations can occur during the replication with the

probability p_{mut} 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

$$E_{\mathrm{b}} = n_{\mathrm{AU}} + 2n_{\mathrm{GC}}, \tag{9}$$

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 $n_{\mathrm{AU}} + n_{\mathrm{GC}} < 2$, 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:

$$p_{\mathrm{diss}}(E_{\mathrm{b}}) = \frac{\exp(-\kappa E_{\mathrm{b}})}{\sum_{E_{\mathrm{b}}=2}^{2L} \exp\left(-\kappa E_{\mathrm{b}}\right)}, \tag{10}$$

where κ incorporates the (inverse) temperature and other possible environmental factors that can have potential effects on duplex separation (e.g., ionic concentration, *p*H, 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., $p_{\mathrm{diss}}(2L) = 0.01$). For sake of simplicity we assume that the rate constant of dissociation is equal to the dissociation probability: $c_{\mathrm{d}} = p_{\mathrm{diss}}$, see *Equation 3A* and *Equation 3B*.

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

$$c_{\delta_{\mathrm{ds}}} = f \cdot c_{\delta_{\mathrm{ss}}} \cdot 0.8^{E_{\mathrm{b}}}, \tag{11}$$

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 $E_{\mathrm{b}} \geq 2$ for every duplex (which is ensured by the $n_{\mathrm{AU}} + n_{\mathrm{GC}} \geq 2$ 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: $c_{\mathrm{a}} = 1$ 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:

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

where $\mu = 1, \ldots, M$, 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:

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

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.

Table 1. Model parameters

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*‡	Duplex decay factor	$\{\mathbf{0.01}; 0.1; 0.5; 1\}$
c_{in}*‡	Kinetic constant for resource influx	$\{0.01; 0.1; 1\}$
$c_{\delta_{\mathrm{ss}}}$‡	Kinetic constant for simplex decay reactions	10^{-4}
$c_{\delta_{\mathrm{ds}}}$‡	Kinetic constant for duplex decay reactions	$[1.15 \cdot 10^{-8}, 6.4 \cdot 10^{-5}]$
c_{out}‡	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	$\{0.05; 0.2; 1\}$
$\phi_{\min}$	Minimum master replicability	0.005
κ	Scale factor for duplex dissociation probability distribution	0.2302

*: model parameters involved in screening, ‡: parameter used only in the chemostat system model

In all simulations, initially each master type is present in equal copy numbers in simple-stranded form and the termination condition is 10^7 replication events, except stated otherwise. During the dynamics, 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 (a type

is considered extinct if its relative frequency drops below 10^{-3}), the simplex-duplex asymmetry and the total number of replicators (in the chemostat system). All simulations were performed in C.

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 2* 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 δs 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), 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* 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*).

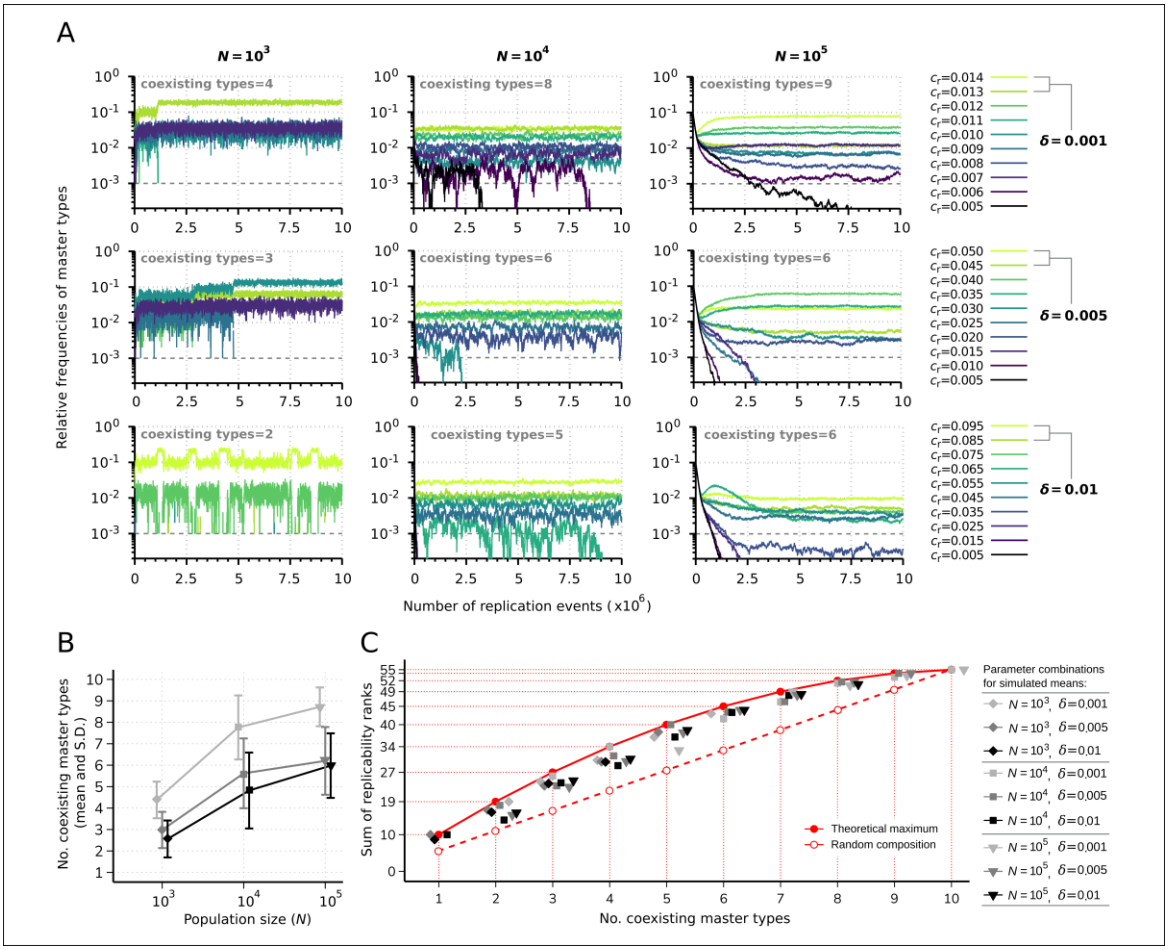


Figure 2. Replicator coexistence in the constant population model. (**A**) Time series of the relative frequencies of master types. Panels corresponding to different rows demonstrate distinct replicability distances (δ), panels corresponding to different columns demonstrate distinct population sizes (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 ranks ($\rho_{\max}$), the dashed red line shows the expected values of the sum of ranks, when the master types are random (ρ_{rand}). The default parameter set (see *Table 1*) was used, unless otherwise indicated.

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.

Compositional effects

The composition of a set of survived master replicator types and thus the selectivity of a parabolically replicating system can be characterized by the replicability rank indices of the coexisting master population. According to *Equation 8*, 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 $\rho_{\max} = \sum_{i=1}^{T_{\mathrm{surv}}} 11 - i = \frac{T_{\mathrm{surv}} \cdot (21 - T_{\mathrm{surv}})}{2}$, $(1 \leq T_{\mathrm{surv}} \leq T)$. The expected value of the sum of ranks, when the masters are distributed randomly with respect to their replicabilities is $\rho_{\mathrm{rand}} = 5.5 \cdot T_{\mathrm{surv}}$. 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*). We found that the key factor leading to this deviation from the theoretical maximums is the GC content (the sum of G and C nucleobases) of the sequences. The direct effects of GC content on the replicator dynamics and thus the selectivity of a parabolically replicating system with regard to sequence effects 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. Finally, random shuffling of the nucleobases among the loci was carried out until the Hamming distance > 3 condition was satisfied for every distinct master-master and master-complementary sequence pair as well as for the 10 master-complementary sequence pair of the same type, thereby excluding palindromes. 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 investigate how parabolic dynamics operates when the above-described sequence effects are excluded (in other words, to indirectly infer the sequence effects on the dynamics from their absence), we examined master sequence abundances under the assumption of equal GC contents, 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. Finally, random shuffling of the nucleobases among the loci was carried out until the Hamming distance > 3 condition was satisfied for every distinct master-master sequence and master-complementary sequence pair as well as for the 10 master-complementary sequence pair of the same type, thereby excluding palindromes. 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 sequence effects 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 not other masters, i.e., the dynamics is "purely" parabolic (cf. Meszéna & Szathmáry, 2002), (*iii*) despite a relatively low total population size ($N = 10^4$), stochastic fluctuations have negligible effects on the dynamics (*Figure 3B*).

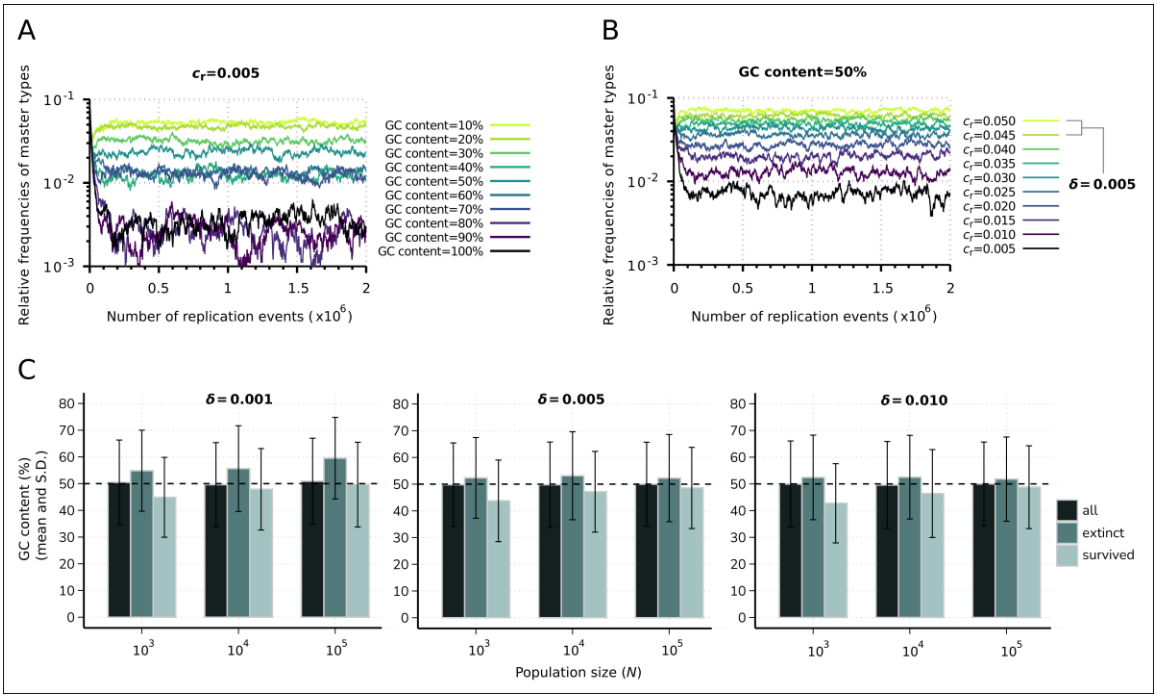


Figure 3. Sequence effects on replicator abundances in the constant population model. (**A**) Demonstration of direct sequence effects on replicator abundances. The master types (indicated by differently colored lines) are defined along a GC-content gradient with identical replicabilities ($c_r = 0{,}005$). (**B**) Demonstration of removal of sequence effects 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%). (**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 2BC*. Columns with the darkest shade represent the average GC content of the randomly generated master sequences constituting the initial populations. Horizontal dashed lines indicate the expected (random) 50% initial GC content. Due to the relatively fast convergence, simulations were terminated after $2 \cdot 10^6$ replication events. Parameters that were used in both simulations: $N = 10^4$, $p_{mut} = 0.01$ (see *Table 1*).

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: $p_{mut} = 0.01, 0.05, 0.1, 0.15$. In accordance with the expectation, 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*.

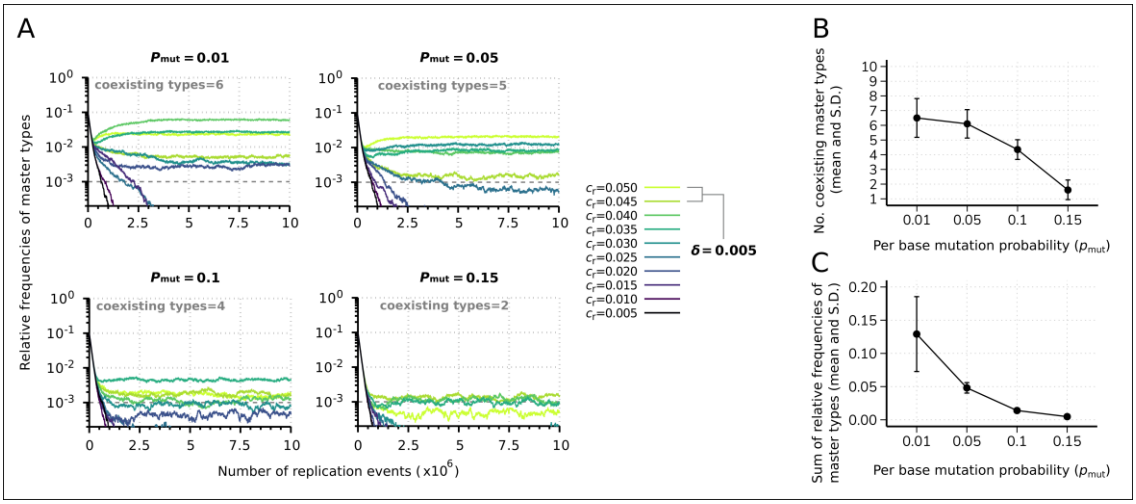


Figure 4. Effects of the mutation rate on the replicator coexistence in the constant population model. (**A**) Time series of the relative frequencies of master types at different mutation rates (p_{mut}). Line colors denote master type replicabilities (values are indicated in the legend). The cut-off 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.

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).

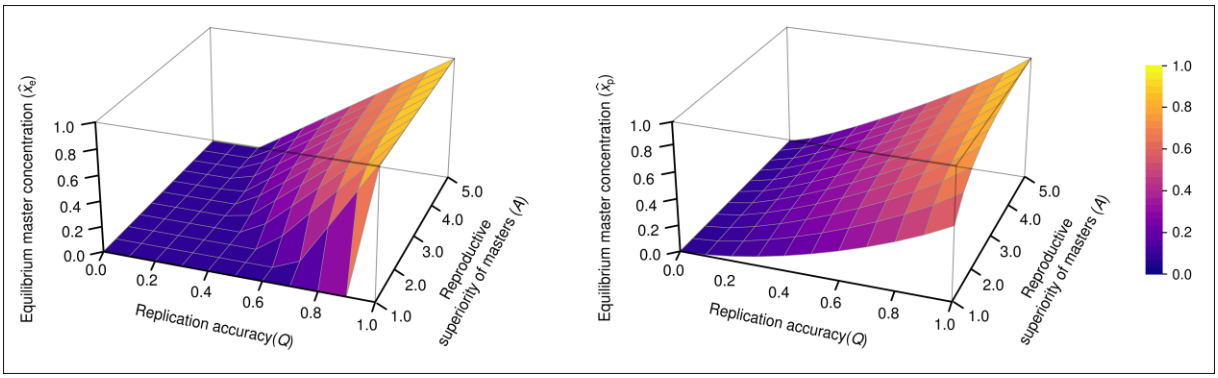


Figure 5. The concentration of master type ($\hat{x}_\mathrm{e}$ and $\hat{x}_\mathrm{p}$) as the function of Q and A in the exponential model (left panel) and in the parabolic replication model (right panel). $a = 1$ (where a stands for the reproductive value of the mutants) was used.

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 $p_\mathrm{B} = [0; 0.1]$ – 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 ($c_{\delta_{\mathrm{ss}}}$) and duplex ($c_{\delta_{\mathrm{ds}}}$) 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 the complete coexistence case, when 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 ($\overline{N}$), 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 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: $\chi = \frac{N_{\mathrm{simplex}} - N_{\mathrm{duplex}}}{N}$, 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 from which one can see 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*).

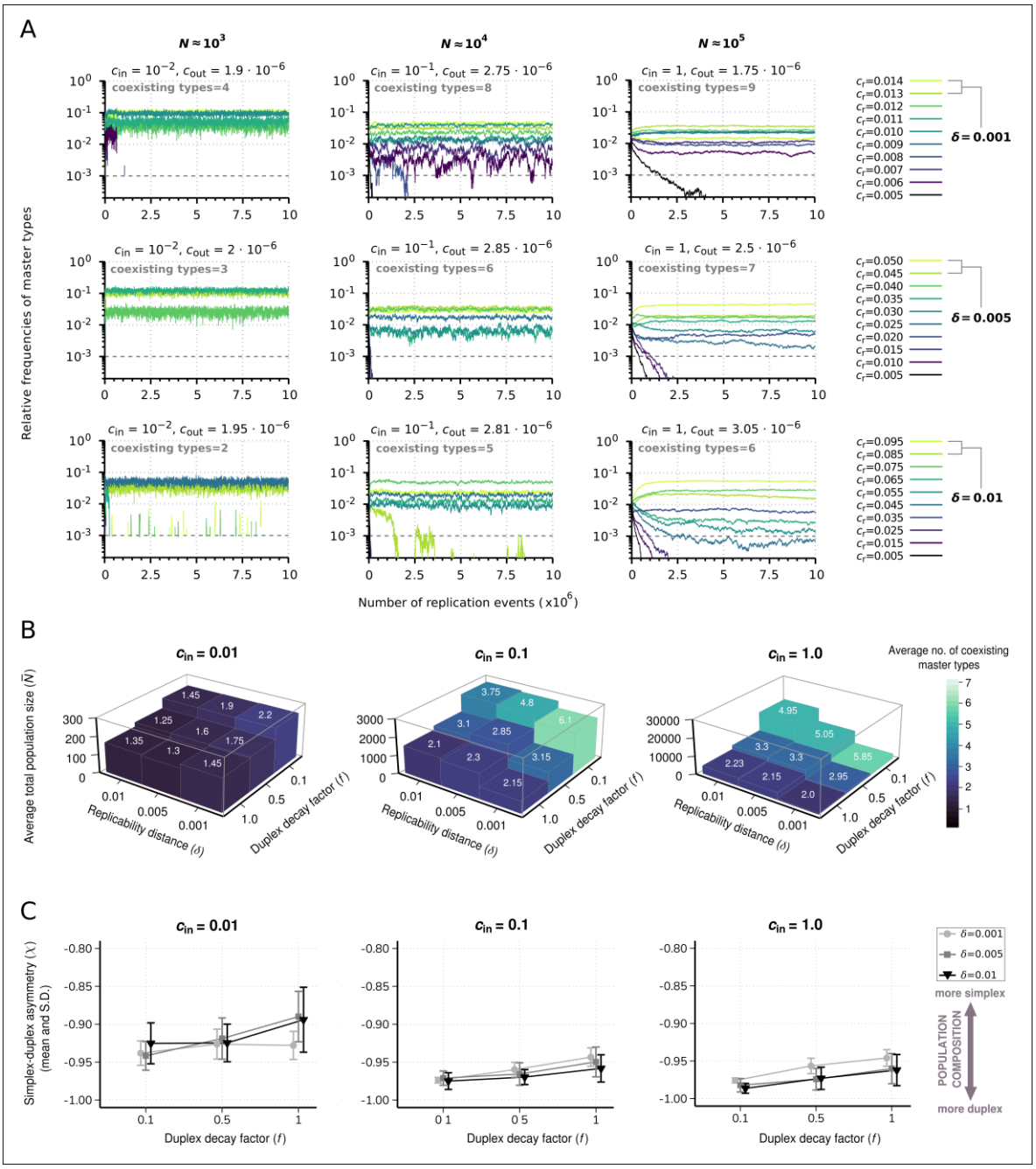


Figure 6. Effects of resource competition on replicator coexistence in the chemostat system model. (**A**) Time series of the relative frequencies of master types. Panels corresponding to different columns demonstrate different population sizes (N). Panels corresponding to different rows demonstrate 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 $\left(\chi = \frac{N_{\text{simplex}} - N_{\text{duplex}}}{N}\right)$ 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.

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 ($c_{\mathrm{in}} = 0.01$), 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 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 (δ = 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 total population size (10^5) and small equidistant difference in the replication rates (δ = 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 ($\delta = 0.001$: $\sum_{i=1}^{T} c_{r(\mathrm{i})} = 0.095$; $\delta = 0.005$: $\sum_{i=1}^{T} c_{r(\mathrm{i})} = 0.275$; $\delta = 0.01$: $\sum_{i=1}^{T} c_{r(\mathrm{i})} = 0.5$) – 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 with population size changes 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 (Meszéna & Szathmáry, 2002). According to this interpretation, 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. Surprisingly, 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 attractor that ensures coexistence towards a Darwinian attractor 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 ($p_{\mathrm{mut}} = 0.15$), the summed relative steady-state master frequencies approach zero (*Figure 4C*) with accelerating, phase transition-like species extinction (*Figure 4B*), indicating that this value (corresponding to the $Q = (1 - 0.15)^L,\ (L = 10) = 0.196$ replication accuracy of a master sequence) 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. Moreover, the constant cut-off relative master frequency for survival (10^{-3}) that we applied makes our empirical error threshold to be a rather conservative approximation. Former studies suggest that the 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 the above outlined studies pertain to

non-enzymatic polymerization reactions that involve 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. 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; Joyce, 2002; Martin et al., 2008). 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; Bada, 2004; Orgel, 2004). 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 these pre-RNA systems as well. It is also worthwhile to mention that 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). Therefore, catalytic aid and/or metabolic cooperation among independently replicating sequences (i.e., *cooperative coexistence*, see e.g., Epstein, 1979; Mizuuchi & Ichihashi, 2018) 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; Li & Chmielewski, 2003; Lincoln & Joyce, 2009) 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). 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 call attention to the fact that a periodically changing environment with alternating resource abundance and shortage can drive, and in the distant past could have driven, an efficient 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.

Additional Information

Competing interests

The authors declare that they have no competing interests.

Funding

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.

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: https://github.com/paczkomatyas/Parabolic_Replicator_2023

References

Attwater, J., Raguram, A., Morgunov, A. S., Gianni, E., & Holliger, P. (2018). Ribozyme-catalysed RNA synthesis using triplet building blocks. *eLife*, *7*, e35255. https://doi.org/10.7554/eLife.35255

Bäck, T., Fogel, D. B., & Michalewicz, Z. (Eds.). (2000). *Evolutionary computation*. New York: Taylor & Francis.

Bada, J. L. (2004). How life began on Earth: A status report. *Earth and Planetary Science Letters*, *226*(1–2), 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*(18), 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*(1), 23. 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*(5), 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*(9), 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*(3), 367–379. https://doi.org/10.1016/0022-2836(68)90392-6

Davis. (1996a). A theory of evolution that includes prebiotic self-organization and episodic species formation. *Bulletin of Mathematical Biology*, *58*(1), 65–97. https://doi.org/10.1007/BF02458282

Davis. (1996b). Darwinian evolution by self-propagating species under fractional order kinetics. *Bulletin of Mathematical Biology*, *58*(1), 99–101. https://doi.org/10.1007/BF02458283

Davis. (2000). Darwinian Aspects of Molecular Evolution at Sublinear Propagation Rates. *Bulletin of Mathematical Biology*, *62*(4), 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*(10), 465–523. https://doi.org/10.1007/BF00623322

Epstein, I. R. (1979). Competitive coexistence of self-reproducing macromolecules. *Journal of Theoretical Biology*, *78*(2), 271–298. 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*(24), 9373–9377. https://doi.org/10.1073/pnas.83.24.9373

Gause, G. F. (1934). *The struggle for existence*. Baltimore: Williams & Wilkins. 163 pp. https://doi.org/10.5962/bhl.title.4489

Gilbert, W. (1986). Origin of life: The RNA world. *Nature*, *319*(6055), 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*(4), 403–434. 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*(25), 2340–2361. https://doi.org/10.1021/j100540a008

Issac, R., & Chmielewski, J. (2002). Approaching Exponential Growth with a Self-Replicating Peptide. *Journal of the American Chemical Society*, *124*(24), 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*(8), 595–605. 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*(37), 14984–14989. https://doi.org/10.1073/pnas.1307604110

Joyce, G. F. (2002). The antiquity of RNA-based evolution. *Nature*, *418*(6894), 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*(5), 469–475. https://doi.org/10.1042/ETLS20190024

Lee, B. D., Neri, U., Roux, S., Wolf, Y. I., Camargo, A. P., Krupovic, M., Simmonds, P., Kyrpides, N., Gophna, U., Dolja, V. V., & Koonin, E. V. (2023). Mining metatranscriptomes reveals a vast world of viroid-like circular RNAs. *Cell*, *186*(3), 646-661.e4. 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*(1), 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*(18), 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*(39), 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*(4), 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*(5918), 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*(1), 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*(11), 805–814. https://doi.org/10.1038/nrmicro1991

Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. Oxford: Oxford Univ. Press.

Meszéna, G., & Szathmáry, E. (2002). Adaptive Dynamics of Parabolic Replicators. *Selection*, *2*(1–2), 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*(10), 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*(1), 60–71. https://doi.org/10.1017/S0305004100033193

Müller, U. F. (2006). Re-creating an RNA world. *Cellular and Molecular Life Sciences*, *63*(11), 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*(3), 381–393. https://doi.org/10.1016/0022-2836(68)90393-8

Orgel, L. E. (1992). Molecular replication. *Nature*, *358*(6383), 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*(2), 99–123. https://doi.org/10.1080/10409230490460765

Patzke, V., & von Kiedrowski, G. (2007). Self replicating systems. *Arkivoc*, *2007*(5), 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*, e17756. https://doi.org/10.7554/eLife.17756

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*(4), 170050. 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*(1), 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'→2') Oligonucleotide System. *Science*, *290*(5495), 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*(1), 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*(4), 629–641. 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*(1), 11. https://doi.org/10.3390/life9010011

Szathmáry, E. (1991). Simple growth laws and selection consequences. *Trends in Ecology & Evolution*, *6*(11), 366–370. 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*(1), 55–58. https://doi.org/10.1016/S0022-5193(89)80177-8

Szilágyi, A., Zachar, I., Scheuring, I., Kun, Á., Könnyű, B., & Czárán, T. (2017). Ecology and Evolution in the RNA World Dynamics and Stability of Prebiotic Replicator Systems. *Life*, *7*(4), 48. 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*(8), e1003193. 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*(1580), 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*(1), 2. https://doi.org/10.1186/1759-2208-3-2

Szostak, J. W., Bartel, D. P., & Luisi, P. L. (2001). Synthesizing life. *Nature*, *409*(6818), 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*(3), 1249–1250. https://doi.org/10.1021/ja00159a057

Varga, Z., & Szathmáry, E. (1997). An extremum principle for parabolic competition. *Bulletin of Mathematical Biology*, *59*(6), 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*(2), 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*(10), 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*(1–3), 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*(4), 423–426. https://doi.org/10.1002/anie.199104231

Wächtershäuser, G. (1988). Before enzymes and templates: Theory of surface metabolism. *Microbiological Reviews*, *52*(4), 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*(6), 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*(12), 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*(6120), 346–347. https://doi.org/10.1038/327346a0

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; pp. 44-49) are as follows:

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

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

while its parabolic counterpart has the following form:

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

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

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:

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

$$(\hat{x}_\mathrm{e}, \hat{y}_\mathrm{e}) = \left(\frac{AQ - a}{A - a}, \frac{A(1 - Q)}{A - a}\right). \tag{16B}$$

If $Q > \frac{a}{A}$ the $(\hat{x}_\mathrm{e}, \hat{y}_\mathrm{e})$ 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):

$$\hat{x}_{\mathrm{p}} = \frac{a^2 + 2A^2Q - \sqrt{a^4 + 4a^2A^2Q(1-Q)}}{2(a^2 + A^2)} \tag{17A}$$

$$\hat{y}_{\mathrm{p}} = \frac{a^2 + 2A^2 - 2A^2Q + \sqrt{a^4 + 4a^2A^2Q(1-Q)}}{2(a^2 + A^2)}. \tag{17B}$$

It is easy to prove that $\hat{x}_{\mathrm{p}}$ is always positive in $0 < Q < 1$, since the expression under the root sign can be estimated as follows:

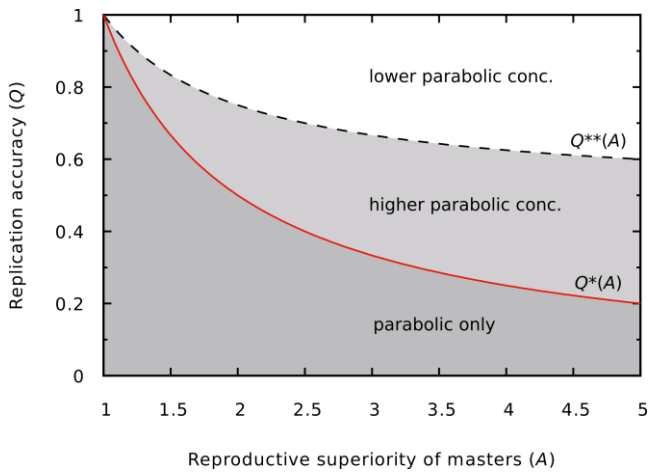
$$\sqrt{a^4 + 4a^2A^2Q(1-Q)} = \sqrt{[a^2 + 2A^2Q(1-Q)]^2 - 4A^4Q^2(1-Q)^2} < a^2 + 2A^2Q.$$

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 $\hat{x}_{\mathrm{p}}$ values, but not otherwise.

When $Q^{**} = \frac{a+A}{2A}$, both regimes have the same master concentrations ($\hat{x}_{\mathrm{p}} = \hat{x}_{\mathrm{e}}$), 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 ($Q^{*} < Q < Q^{**}$):

$$Q^{**} - Q^{*} = \frac{A - a}{2A}. \tag{18}$$

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.



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.

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

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

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

while its parabolic counterpart has the following form:

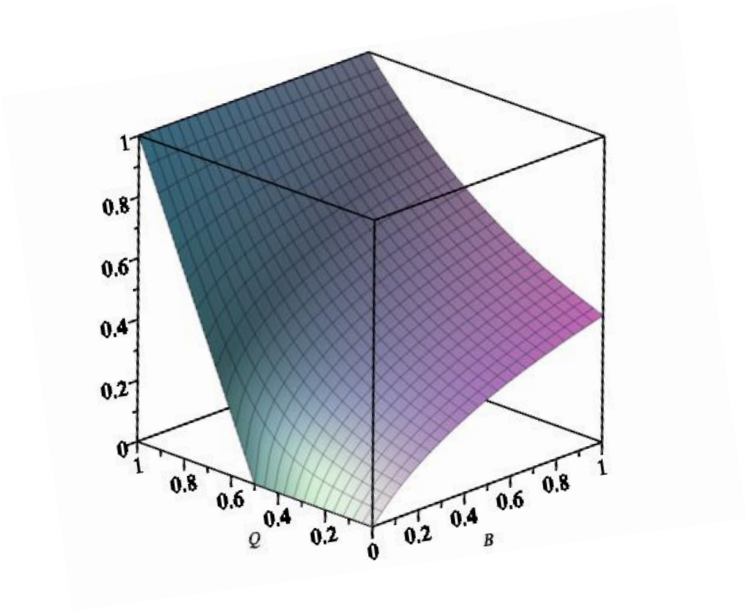
$$\dot{x} = AQ\sqrt{x} + aB\sqrt{y} - x\left(A\sqrt{x} + a\sqrt{y}\right) \tag{20A}$$

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

The non-trivial equilibrium master concentration of the exponential dynamics of *Equation 19A* and *Equation 19B* is

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

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 $\hat{x}_{\mathrm{eB}}$ as the function of Q and B.

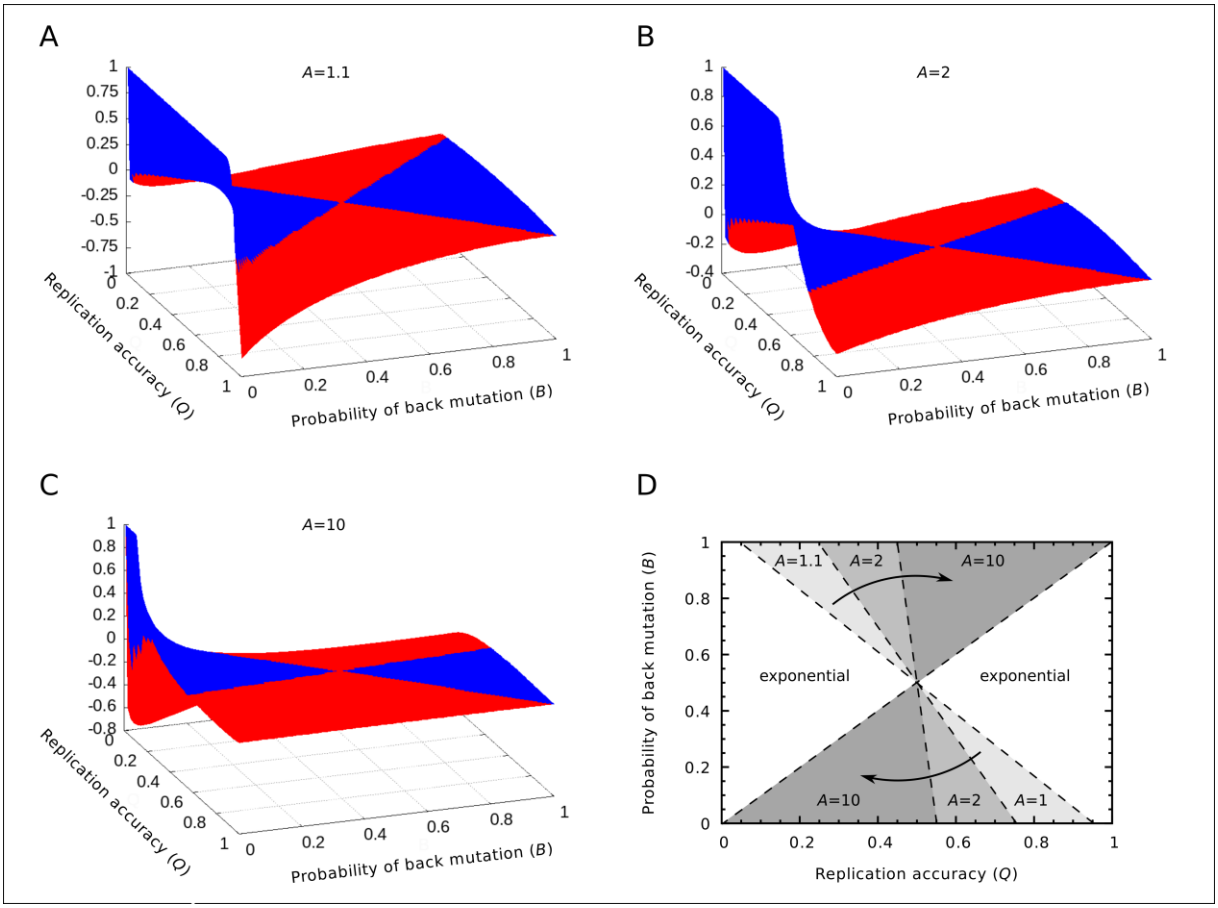


Appendix 1—figure 2. Equilibrium concentration of the master in the exponential system with back mutation ($\hat{x}_{\mathrm{eB}}$) as a function of Q and B. Parameters are: $A = 2, a = 1$.

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 $\hat{x}_{\mathrm{eB}} = \hat{x}_{\mathrm{pB}}$ equation for B with computer algebra ($\hat{x}_{\mathrm{pB}}$ is the relevant fixed point of *Equation 20A* and *Equation 20B*). The result is:

$$B_1(Q) = Q \tag{22A}$$

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



Appendix 1—figure 3. (**A–C**) The relative differences between the equilibrium concentrations of masters in the two regimes: $\frac{\hat{x}_{\mathrm{p}}-\hat{x}_{\mathrm{e}}}{\hat{x}_{\mathrm{p}}}$ 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).