Title:

Sustainable replication and coevolution of cooperative RNAs in an artificial cell-like system

Authors:

Ryo Mizuuchi[a], Norikazu Ichihashi[a,b]

Affiliations:

[a] Department of Bioinformatics Engineering, Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan

[b] Graduate School of Frontier Biosciences, Osaka Universit, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan

Corresponding Author: Norikazu Ichihashi

Address for all authors:

Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan

Tel: 81-6-6879-4151; Fax: 81-6-6879-7433; ichihashi@ist.osaka-u.ac.jp

Keywords

cooperation, RNA replicator, parasite, evolution, coevolution, compartment

Abstract

Cooperation among independently replicating molecules is a key phenomenon that allowed the development of complexity during the early evolution of life. Generally, this process is vulnerable to parasitic or selfish entities, which can easily appear and destroy such cooperation. How this fragile cooperation process appeared and was sustained during the evolution is one of the largest mysteries. Theoretical studies indicated that spatial structures, such as compartments, allow a sustainable replication and the evolution of cooperative replication, although this has yet to be confirmed experimentally. In this study, we constructed a molecular cooperative replication system, in which two types of RNAs, encoding replication or metabolic enzymes, cooperate for their replication in compartments, and performed long-term replication experiments to examine the sustainability and evolution of the RNAs. We demonstrated that the

cooperative relationship of the two RNAs could be sustained at a certain range of RNA concentrations, even experiencing the appearance of parasites. We also found that a more efficient cooperative RNA replication evolved during long-term replication through seemingly selfish evolution of each RNA. Our results represent the first experimental evidence supporting the sustainability and robustness of molecular cooperation at the evolutionary timescale.

Introduction

Cooperation is a key phenomenon that allows for major transitions from simpler to more complex levels of biological organization[1,2]. In these transitions, independently replicating entities, such as molecules, cells, and organisms, started to divide their roles and cooperate with each other for the replication of the entire system. In this study, we focused on one of the earliest cases, the cooperation of RNAs encoding different functions or genes, which allowed the transition from a simple RNA replicator to a more complex multifunctional RNA replication system in the RNA world or the RNA-protein world[3-5]. A considerable obstacle to establishing this cooperation is the appearance of parasitic or selfish RNAs (i.e., cheaters), which can easily appear if one of the cooperative RNAs lose the function or gene that supports the replication of other RNAs. These parasitic RNAs replicate predominantly since they need not pay any cost for cooperation and eventually inhibit the cooperative replication of the original RNAs[4,6].

It remains unclear how such fragile cooperation among RNAs could be sustained with spontaneous appearance of parasitic RNAs. Theoretical studies predicted that a possible solution for the parasite appearance is the introduction of spatial structures, such as cell-like compartments. Cooperative molecular replication systems (*e.g.*, hypercycles) can be sustained if they are compartmentalized[3,4,6,7] or located in the equivalent spatial structures[8-14], owing to group-level selection. Other theoretical studies predicted that mutant replicators could evolve under certain conditions without losing cooperative activity[7,15-18]. Despite decades of theoretical studies, experimental verification of the sustainability and evolution of molecular cooperation remains to be performed.

To date, several types of replication systems have been constructed using RNAs[19,20], DNAs[21,22], and peptides[23]. However, in these systems, mutations are not spontaneously introduced[19-21,23] or they are introduced only in a limited region[22], and thus parasitic

replicators did not appear. The sustainability of these systems in the presence of parasitic replicators has not been investigated.

In our previous study, we constructed a simple non-cooperative RNA-protein replication system that consists of an RNA molecule (artificial genomic RNA) and a cell-free translation system of *Escherichia coli*. This replication system can be used as an experimental model for a primitive replication system in the RNA-protein world. In the previous system, the genomic RNA replicated by the RNA replicase translated from itself. When we repeated the replication, mutations were introduced spontaneously in the RNA by replication error, and a series of RNA mutants that replicate faster successively dominated the population (i.e., evolution occurred)[24,25]. We also found that a parasitic RNA that lost the replicase gene appeared spontaneously through RNA recombination and competitively inhibited the replication of genomic "host" RNA. This genomic RNA continued to replicate even after the appearance of parasitic RNA, when the system was encapsulated in water-in-oil droplets[26], which functions as a shelter for the genomic RNA to escape from parasitic RNA.

We thought that construction of a cooperative RNA replication system consisting of two types of RNAs, by expanding the previous RNA-protein replication system, would be a good experimental model to investigate the persistence of cooperation among RNA species with the spontaneous appearance of parasitic RNAs. In the previous single RNA replication system, RNA replication was sustained even after the appearance of parasitic RNAs if the system was compartmentalized in water-in-oil droplets[26]. This is because genomic RNA can eventually restart replication in parasite-free compartments after sufficient dilution of parasitic RNA; however, cooperative RNA replication is not guaranteed because in this case, both of the cooperating RNAs should exist in the same parasite-free compartment, by chance, in order to restart replication. Therefore, it is still unknown under what conditions cooperative RNA replication can be sustained with spontaneous appearance of parasite, even if it is compartmentalized. Moreover, it is also unknown how the functions of the cooperating RNAs change during long-term replication with continuous mutagenesis.

In this study, we investigated the sustainability of RNA cooperation with the appearance of parasites, both theoretically and experimentally. We first identified the critical parameters for this sustainability, i.e., the average RNA concentration, by using computer simulations based on a simple theoretical model. Then, we constructed a

cooperative RNA replication system composed of two types of cooperative RNAs by expanding the previous RNA-protein replication system. Through long-term replication of the RNAs in water-in-oil droplets, we found that parasitic RNAs actually appeared at high average RNA concentrations and significantly inhibited the cooperative replication of RNAs, while at lower average concentrations, the cooperative RNAs replicated sustainably and, furthermore, coevolved.

Results

Cooperative replication system analysis using simulation

Prior to construction of an experimental system, we conducted a theoretical analysis to understand the basic nature of cooperative replication systems. This theoretical model mimics a primitive RNA replication system that could exist in the RNA world or RNA-protein world[3-5], and also can be an abstract model of the experimental cooperative system we constructed below. This model consists of two RNA replicators, RNA-1 and RNA-2. These two RNAs are assumed to have different functions (e.g., encoding different genes) and both of them are required for the replication of both RNAs (Fig. 1a). We also assumed the appearance of a mutant RNA (parasitic RNA) from both of the RNAs by losing encoded functions while retaining the activity to be replicated. Such parasitic RNA can easily appear by the deletions of or deleterious mutations in the encoded genes, as we have observed in the previous non-cooperative RNA replication system[27]. The parasitic RNA rapidly replicates without supporting the replication of other RNAs and competitively inhibits the cooperative replication between RNA-1 and -2 (Fig. 1b). This replication system was encapsulated in uniformly-sized compartments, which could repress the amplification of the parasitic RNA[26]. These compartments mimic primitive cell-like structures, such as vesicles of amphiphilic molecules synthesized abiotically[28].

Long-term replications were simulated through the following three steps: replication, dilution, and fusion-division (Fig. 1c). At the replication step, the parasitic RNAs appeared and all RNA species (RNA-1, -2, and parasitic RNAs) were replicated. At the dilution step, a certain number of compartments were removed based on the dilution rate, and vacant compartments were added instead. The dilution rate was adjusted to maintain the average RNA concentrations in pre-determined ranges. At the fusion-division step, the contents of the compartments were mixed through a number of fusion-division processes independent of internal RNA replication. It should be noted

that these processes, such as a constant dilution of droplets, and stochastic fusion and division of compartments, might not be realistic processes that actually occurred on the early Earth; nevertheless, we attempted to emulate simplified processes that allow continuous RNA replication in cell-like structures. By using this simplified model, we can investigate the conditions that allow the sustainable replication of cooperating RNAs, a prerequisite for evolution from a simple RNA replicator to a more complex and multifunctional replication system.

We found that two issues can lead to the collapse of cooperative RNA replication in compartments, amplification of the parasitic RNAs and stochastic mis-encapsulation of cooperating RNAs, which are referred to as "the mutational reef" and "fluctuation abyss", respectively in previous literatures[4,29], and both are directly related to the average concentrations of the RNAs. When we initiated the replication at a high RNA concentration, the average concentrations of both RNA-1 and -2 gradually decreased after a few rounds of replication and never recovered (Fig. 1d, High); this was caused by the immediate appearance and rapid amplification of the parasitic RNA. However, for simulations initiated using low RNA concentrations, the parasitic RNA did not appear, yet the average concentrations of both RNA-1 and -2 decreased and never recovered; this was caused by the overly small chance for both RNAs to be encapsulated in the same compartment (*i.e.*, stochastic mis-encapsulation) (Fig. 1d, Low). In contrast, in the middle concentration range, both RNA-1 and -2 continuously replicated due to the repression of the parasitic RNA amplification and high enough chance for both RNA-1 and -2 to be encapsulated in the same compartment (Fig. 1d, Middle). Additional simulations at various concentrations revealed that the sustainable replication could be achieved only in the middle RNA concentration range (Supplementary Fig. 1). These results are fundamentally consistent with the results of a previous theoretical study[4].

Construction of an empirical cooperative RNA replication system

To verify the effect of RNA concentration on the sustainability of cooperative RNA replication and to further investigate how cooperative relationships change during a long-term replication, we experimentally constructed a cooperative molecular replication system. The replication system (scheme presented in Fig. 2a) was based on the translation-coupled RNAs replication system we previously constructed[24]. The original system consisted of the *E. coli* translation system and an artificial genomic RNA (Rep-RNA) encoding the core subunit of an RNA replicase (Qβ replicase), in which genomic RNA was replicated by the replicase translated from itself. In this study, we

introduced another RNA, called NDK-RNA, which encodes *E. coli* nucleotide diphosphate kinase (NDK) and contains replicase recognition sites on the termini. To enable the mutualistic replication of the RNAs, we substituted cytidine triphosphate (CTP) in the original system with cytidine diphosphate (CDP). Therefore, RNA replication required the NDK activity that converts CDP to CTP. In this system, both Rep- and NDK-RNAs could replicate only when both the replicase and NDK are translated from Rep- or NDK-RNAs. We designated this system as the translation-coupled cooperative RNA replication (TcCRR) system, and encapsulated it into the micro-scale water-in-oil droplets for compartmentalization.

A large obstacle to the construction of the TcCRR system was the creation of NDK-RNA that could replicate efficiently. The first NDK-RNA we constructed barely replicated because it lacked strong secondary structures throughout the molecule, essential for replication by the replicase[30,31]. Recently, we established a method to create replicable RNA sequences by introducing mutations that modify RNA structures[32]. Using this method, we introduced 52 mutations in three steps and obtained a NDK-RNA mutant (Mod2-CE-X) with strong secondary structures throughout the molecule (Supplementary Table 1 and Supplementary Fig. 2), and could replicate 250-fold better than the original sequence (Supplementary Fig. 3). In the subsequent sections, we refer to this mutant as NDK-RNA.

We next examined the cross-dependency of NDK-RNA and Rep-RNA during their replication processes. Each or both RNAs were incubated in the TcCRR system for 4 h at 37 °C, and the replication levels were determined. In the absence of the other RNA, both RNAs replicated negligibly, whereas, in the presence of both RNAs, they replicated 7 - 13-fold (Fig. 2b), indicating that the two RNAs replicate in a cooperative manner. In addition, the complementary strands were synthesized only in the presence of the respective partner (Supplementary Fig. 4).

Long-term replication experiments

To investigate the sustainability of this cooperative RNA replication, we performed long-term replication experiments starting at different RNA concentrations. The schematic representation of this process is shown in Fig. 3a. Initially, the TcCRR reaction was performed with Rep- and NDK-RNAs at defined concentrations at 37 °C for 4 h in the water-in-oil droplets, and the average RNA concentrations in all droplets were determined. During this replication process, mutations were introduced into both

RNAs by replication errors (approximately 10^{-5} per replication per nucleotide[33]). Next, the droplets were diluted with new droplets that contain the mixture for TcCRR reaction without Rep- and NDK-RNAs, and these diluted droplets were mixed with a homogenizer to induce fusion and division among droplets for the next round of TcCRR reaction. The dilution rates were adjusted to maintain the average RNA concentrations in certain ranges with five-fold as the minimum value; therefore, if the average RNA replication values are lower than five-fold, the average RNA concentrations should decrease as replication rounds proceed.

When high initial Rep- and NDK RNA concentrations (300 nM) were used, their replication immediately stopped and high levels of parasitic RNAs appeared (Fig. 3b, High). Afterwards, both RNAs decreased to undetectable concentrations. When the initial concentrations of Rep- and NDK RNAs were low (0.03 nM), both RNAs replicated only slightly and, subsequently, they decreased to undetectable levels (Fig. 3c, Low). In contrast, when the initial concentrations of Rep- and NDK-RNAs were in the mid-range (10 nM), and later maintained in the lower range (0.003 – 1 nM), their replications continued for at least 50 rounds, corresponding to approximately 160 generations, although relatively lower parasite levels were detected in the first several rounds (Fig. 3d, Middle). We confirmed the reproducibility of this sustainable replication in another lineage that was separated after the round 18 (Supplementary Fig. 8a). We also found that if average concentrations were increased from the range, then parasitic RNA appeared and cooperative replications stopped in two cases (Supplementary Figs. 8b and 8c). These results show that suitable RNA concentration is crucial for the sustainability of the experimental cooperative system, consistent with predictions obtained using the simple theoretical model described above.

One of the unexpected results of the long-term replication experiments was the observation that once cooperative replication started, the replication became sustainable at lower concentrations. When the replication started at mid-range concentrations (10 nM), it continued sustainably in the lower range (0.003 – 1 nM) (Fig. 3d), although RNA replication could not be initiated from this range (Fig. 3c, 0.03 nM). Furthermore, we found that the replication continued when we further diluted the compartments more than 100-fold in the lower range (Figs. 3e and 3f), indicating robustness of the cooperative replication system against compartment dilution. This may be explained by an uneven RNA distribution in droplets caused by incomplete mixing of the internal solution among droplets, which helped maintain sufficient RNA

concentrations in some compartments. This point was confirmed using simulations (Supplementary Fig. 6).

Analysis of RNA populations

In the long-term replication system, random mutations were introduced by replication error, with most of them expected to be deleterious or neutral and only a small fraction expected to be beneficial. The mutant RNAs containing these mutations change their frequencies depending on their replication ability. Therefore, the composition of the RNA population inevitably changes during long-term replication. To investigate how these mutations alter the sequence composition of sustainably replicating RNAs, we obtained 32 clones at round 50 (Fig. 3d) and analyzed their sequences. Rep- and NDK-RNAs accumulated an average of 5.8 and 3.2 mutations, respectively (Supplementary Table 2). Most mutations were unique for each clone (Supplementary Tables 3 and 4), representing a typical pattern of quasi-species, a population of mutants containing a small number of random mutations[34].

Next, we analyzed the replicative ability of these clones. The cooperative replication-associated activities of Rep- and NDK-RNAs can be divided into two types: those associated with replicase-mediated replication (template activity) and those associated with assisting replication of other RNA (cooperation activity). We analyzed these activities for all Rep- and NDK-RNA clones. The template activity was determined as the RNA levels replicated by the purified original replicase. The cooperation activity was determined using the following two-step reactions. Initially, the replicase or NDK were expressed from each clone, and then used to induce replication of the other original RNA. The replicated RNA level was then measured as an index of cooperation activity (detailed methods are described in Supplementary Fig. 7 and Materials and Methods). For the majority of Rep-and NDK-RNA clones, template activities were similar or higher than those of the originals (Fig. 4a). In contrast, cooperation activities of all Rep-RNA clones were lower than those of the original molecules, and half of these clones had almost no activity (Fig. 4b, left). Similarly, cooperation activities of approximately one-third of NDK-RNA clones were lower than those of the original molecules, although more than 50% of these clones had higher cooperation activities (Fig. 4b, right).

One of the possible explanations for decreased cooperation activities of most of the Rep-RNAs is the change in template specificity, whereby the Rep-RNAs might replicate

themselves specifically. To examine this possibility, we chose some Rep-RNA clones and measured the replication of own RNA as described in the Method section. The own replication showed a similar tendency to the replication of NDK-RNA (i.e., cooperation activity) (Supplementary Fig. 9), indicating that the template specificity of the replicase did not significantly change and cannot explain the decrease of cooperation activities.

Analysis of evolution

The template and cooperation activities of some clones were shown to be higher than those of the original molecules, which indicate that beneficial mutations should be enriched in the population during long-term replication. Our analyses showed that some mutations were enriched in the 32 clones selected at round 50 (Supplementary Tables 3 and 4). The most common mutation sets were A206G, U746C, and G975A, found in 31% of clones in Rep-RNA, and C86U and U234G, found in 75% of the NDK-RNA clones (Fig. 5a). In the subsequent experiments, we used Rep- or NDK-RNAs containing only these mutations as the evolved RNAs.

We initially performed the TcCRR reactions with all combinations of original (Ori) or evolved (Evo) Rep- and NDK-RNAs. The evolved pair (Evo × Evo) exhibited higher replication levels than those of the original pair (Ori × Ori, Fig. 5b), indicating that cooperative replication improved with the mutations. Interestingly, both RNAs were shown to replicate efficiently only if paired with the evolved partner RNAs; when each of the evolved RNAs were paired with the original partner RNAs (Ori × Evo or Evo × Ori), the evolved RNAs replicated to a level similar to that of the evolved pairs, while the original partner RNA replication was lower than that in original pairs, indicating that Rep- and NDK-RNAs adapted to each other.

To understand this process, we determined the template and cooperation activities of the evolved RNAs as described above. Template activities were found to be increased approximately two-fold for both Rep- and NDK-RNAs (Fig. 5c), while the cooperation activity of NDK-RNA alone was increased two-fold (Fig. 5d).

Discussion

One of the long-standing questions important for the understanding of the early evolution of life is how molecular cooperation was sustained and developed. Here, we constructed a cooperative replication system comprising two types of RNAs encoding

replication and metabolic enzymes. Using this system, we initially confirmed that the cooperative system has two important problems, parasite amplification and stochastic mis-encapsulation of cooperating RNAs, as theoretically predicted[4,29], and thus the system can only be sustainable in the middle RNA concentration range (Figs. 3b, c, d). We showed that initial RNA concentrations are important, however, following the initiation of the replication, the cooperation system shows considerable robustness against compartment dilution (Figs. 3e, f). Furthermore, we found that after the initiation of replication, two cooperating RNAs spontaneously accumulate mutations that allow a larger amount of replication (Fig. 5). Our experiments demonstrated that, although a molecular cooperation is difficult to sustain as it is easily disrupted by parasitic molecules and stochastic mis-encapsulation, once the initially required conditions are satisfied, it can develop more easily than we previously expected.

The mutations introduced in the long-term replication system seemingly allowed both Rep- and NDK-RNAs to replicate "selfishly" because the evolved RNAs acquired higher template activities (Fig. 5c) and also inhibited activities of other RNAs when paired with the original RNAs (Ori × Evo or Evo × Ori, Fig 5b). Nevertheless, the evolved Rep- and NDK-RNAs that were paired (Evo × Evo) replicated by a greater amount compared to the original pair (Ori × Ori), indicating that the cooperative replication became more efficient by the introduced mutations. These seemingly contradictory results could be attributed to the combination of the selfish evolution of each RNA and the compensation by adaptation of both RNAs to each other. These results indicate that the development of cooperative replication can be compatible with selfish evolution of the components at least in this experimental setup, supporting the plausibility of sustainable molecular cooperation.

For more discussion, see Supplementary Information.

Materials & Methods

Plasmids and RNAs

Plasmid encoding the original Rep-RNA was described in our previous study as the plasmid encoding R11 mutant[35]. The plasmid encoding the first NDK-RNAs (pUC-NDK) was prepared by ligating two DNA fragments, a PCR fragment amplified using pUC-N96(+) as template and primers 8 and 9 (Supplementary Table 5), and a PCR

fragment amplified using the NDK expressing plasmid[36] as template and primers 10 and 11. Mutations (C182A, C184A, and G insert into 219-220) were introduced into pUC-NDK to increase NDK expression and the resulting plasmid, pUC-NDK-HT, was used for the production of original NDK-RNA molecules. To synthesize plasmids encoding modified NDK-RNAs, Mod2, Mod2-CE, and Mod2-CE-X, synonymous mutations (Supplementary Table 1) were introduced by artificial DNA synthesis (FASMAC) or PCR reaction with mutagenized primers. The RNAs were prepared by *in vitro* transcription using T7 RNA polymerase as described previously[37].

NDK contamination-less translation system

The translation system used in this study is based on the reconstituted *E. coli* translation system[36]. The composition of the translation system used for the TcCRR reaction is the same as that used in our previous study[38], except that NDK and myokinase were omitted, creatine kinase concentration was 25 nM, and CTP was substituted with CDP. The purification of the protein components was based on a previously reported method[39], with the exception of EF-Tu and ribosome. To eliminate contaminated NDK activity, EF-Tu and ribosome fractions were washed with stringent buffer (50 mM Hepes-KOH (pH 7.6), 1 m potassium chloride, 10 mM magnesium chloride, 15% glycerol, 1 mM dithiothreitol, and 1% Triton X-100), and purified twice using a nickel column chromatography or ultracentrifugation, respectively, as previously described[39]. The composition of the original translation system used to analyze replicase and target replication is the same as previously described[38].

TcCRR reaction

The standard reaction mixture contained 10 nM Rep-RNA, 10 nM NDK-RNA, and the translation system for the TcCRR reaction. Reaction mixture (10 μL) was vigorously mixed in 1 mL of saturated oil using a homogenizer (Polytron PT-1300D, KINEMATICA) at 15 krpm for 1 min on ice. The saturated oil was prepared as described previously[24]. The average diameter of the water-in-oil droplets was approximately 2 μm[26]. After incubating the emulsion at 37 °C, RNA concentrations were determined using quantitative PCR after reverse-transcription (quantitative RT-PCR), as described previously[24]. For measurement of the original NDK-RNA concentrations before evolution, we used primer 1 for reverse transcription and primers 3 and 4 for PCR. For the complementary strand, we used primer 2 for the reverse transcription and primers 5 and 6 for PCR. For the evolved NDK-RNAs, we used primer 7 instead of primer 3 and primer 14 instead of primer 2, since a mutation (U234G) was introduced

at the original primer regions. In the experiments shown in Supplementary Fig. 3, we used NDK-RNAs labeled with Sp-GTP-α-S (BIOLOG Life Science Institute) and degraded the template RNA after reactions as described previously[24]. The TcCRR reaction was performed in water-in-oil emulsion.

Long-term replication experiments

The TcCRR reaction was conducted in 1 mL emulsion using different Rep- and NDK-RNA concentrations, as described. After 4-h incubation at 37 °C, an aliquot of the emulsion was diluted with 1 mL of new saturated oil containing 10 μL of the translation system, and the mixture was homogenized as described. The amount of the aliquot varied. The emulsion was incubated for 4 h at 37 °C for the next round of replication. Rep- and NDK-RNA concentrations were determined after the replication step by using quantitative RT-PCR as described above. NDK-RNA concentrations after round 19 (Fig. 3) were determined using primer 7 instead of primer 3.

Parasitic RNA determination

Water phase was recovered from 300 μL of the emulsion at each round by centrifugation (15,000 rpm, 5 min). The recovered solution was mixed with four volumes of diethyl ether and centrifuged (10,000 rpm, 1 min). After removing the diethyl ether phase, RNA was purified with PureLink RNA Mini Kit (Thermo Fisher) and subjected to 8% polyacrylamide gel electrophoresis as previously described[27]. The gel was stained with SYBR Green II (Takara) and quantified using the band of a commonly appearing parasitic RNA (s222)[40] as a standard.

Sequence analysis

Both Rep- and NDK-RNAs at round 50 were amplified using primer 12 for reverse-transcription and primers 12 and 13 for the subsequent PCR, and ligated into a vector as described previously[24]. Thirty-two clones of each RNA were randomly picked and the plasmid sequences were analyzed. The same primers and vector were used for both Rep- and NDK-RNAs.

Template activity assay

Rep- or NDK-RNAs (1 nM) were mixed with the purified replicase (10 nM)[41] in the original translation system without amino acids to stop translation. The sample was incubated at 37 °C for 15 min and concentrations of the synthesized complementary strands were determined by quantitative RT-PCR as described above. For some Rep and

NDK-RNA clones, we used primers 15 or 17, while primers 16 and 6 or 18 and 6, respectively, were used for the subsequent PCR, because some clones had mutations in the primer regions.

Cooperation activity assay for Rep-RNA

The cooperation activity of Rep-RNA was determined in the following reactions. In the first reaction, replicase was translated from each Rep-RNA (10 nM) at 37 °C for 2 h in the original translation system, without UTP, to stop replication. In the second reaction, an aliquot (1/10 volume) of the first reaction was mixed with the original translation system containing Mod2-CE-X NDK-RNA (10 nM) as the replication template, and 30 μg/mL streptomycin to inhibit further translation. After incubation at 37°C for 20 min, the synthesized complementary strand of NDK-RNA was quantified as an indicator of the cooperation activity, by using quantitative RT-PCR. For the assay of six Rep-RNA clones for own replication, we used each Rep-RNA (10 nM) as the replication template instead of Mod2-CE-X NDK-RNA in the second reaction.

Cooperation activity assay for NDK-RNA

The cooperation activity of NDK-RNA was determined in the following reactions. In the first reaction, NDK was translated from each NDK-RNA (100 nM) at 37 °C for 2 h in the original translation system. In the second reaction, an aliquot (1/10000 volume) of the first reaction was mixed with a replication solution containing 100 nM original Rep-RNA, 1 μM purified original replicase[41], 1.25 mM ATP, 1.25 mM GTP, 1.25 mM UTP, 1.25 mM CDP, 25 μg/mL streptomycin, 125 mM Tris-HCl (pH 7.8), 10 mM magnesium acetate, and 0.01% BSA. Following incubation at 37 °C for 45 min, the synthesized complementary strand of Rep-RNA was quantified as an indicator of the cooperation activity, by using quantitative RT-PCR as described.

Simulation

$[R_1]$: RNA-1 concentration in each compartment
$[R_2]$: RNA-2 concentration in each compartment
[P]: Parasite concentration in each compartment
k_1: RNA-1 replication rate constant (0.15)
k_2: RNA-2 replication rate constant (0.15)
k_p: Parasite replication rate constant (0.3)
A: Average concentrations of RNA-1 and RNA-2 in all compartments
T: Target average RNA concentrations

L_1: Length of RNA-1 (10)

L_2: Length of RNA-2 (4)

L_p: Length of parasite (1)

C: Carrying capacity (2000)

D_{min}: Minimum dilution rate (5)

M: Number of compartments (2000)

p: Probability of parasite generation per RNA-1 or RNA-2 molecule per round (0.00005)

F: Fusion-division number (10)

The values in the parentheses were used, unless otherwise specified.

We considered one of the simplest schemes of the cooperative replication of two types of RNAs, where the replication rates of RNA-1, RNA-2, and parasite depend on both RNA-1 and RNA-2 concentrations ([R1] and [R2]), length of each RNA (L1, L2, Lp) and the carrying capacity (C) as follows:

$$\frac{dR_1}{dt} = k_1[R_1][R_2]\left(1 - \frac{L_1[R1] + L_2[R2] + L_p[P]}{C}\right), \qquad [S1]$$

Supplementary Information

Sustainable replication and coevolution of cooperative RNAs in an artificial cell-like system

Ryo Mizuuchi, Norikazu Ichihashi

Supplementary discussion

Sustainability of molecular cooperation

Computer simulation based on the simple model (Fig. 1d) qualitatively explained the importance of RNA concentration for sustainable replication that we demonstrated experimentally (Fig. 3), but there are several experimental results that the simulation could not explain. First is the RNA concentration range in which cooperative replication is sustainable. This did not change in the simulation as the round of replication proceeded, but in the experiment, the region shifted downward after a few round of replications (Fig. 3d). This discrepancy can be explained by the incomplete mixing of droplets contents in the experiment (Supplementary Fig. 6). The second result that cannot be explained by simulation based on the simple theoretical model is the heterogeneity of RNA populations. Most Rep-RNAs lost cooperation activity, yet still RNA populations replicated as a whole (Fig. 4b). This heterogeneity in cooperative activity is similar to the decrease in catalytic activity of one of the complementary RNA strands (i.e., symmetry braking) observed in recent theoretical studies[1-3]. It would be an important next step to investigate the effect of heterogeneous populations on the sustainability and evolution of cooperative replication systems by introducing mutational effects in the theoretical model.

One of the important consequences of cooperative replication is more weakness to parasite amplification than non-cooperative RNA replication. In our previous non-cooperative single RNA replication system, even if parasitic RNAs amplified enormously, the host genomic RNA replication can recover after sufficient dilution of parasitic RNAs[4]. However, in the cooperative system, the cooperating RNAs are also diluted after sufficient dilution of parasitic RNAs, and thus there is little chance for the two RNAs to be encapsulated in the same compartment. This could explain why the RNA replication did not recover once parasitic RNAs amplified enormously in this cooperative system (Fig. 3b). To avoid parasite amplification and achieve sustainable replication, we had to control the RNA concentrations in this study. The next important challenge is to find a condition that allows sustainable cooperative replication without any artificial control of RNA concentration, which would be a prerequisite for a molecular cooperation to be sustainable during early Earth. A possible method to avoid the appearance of parasitic RNA is to reduce the RNA replication substrate, which decreases the amplification of the parasites and eases their negative effects, as reported in our previous study[5].

Evolution of cooperative systems

To date, little is known about how a cooperative molecular system changes with mutations at an evolutionary timescale. A previous theoretical study predicted that the cooperative replication can be sustained with continuous mutations producing quasi-species[6]. Other studies investigated possible evolutionary processes and found that some mutant replicators that have different properties can dominate the population, which is highly dependent on the constraints and parameters of the model[7-10]. In this study, we performed long-term replication experiments with continuous mutations, and observed that a considerable fraction of clones lost their cooperation activities (Fig. 4b), while maintaining cooperative replication as a whole. This result demonstrates the robustness of cooperative replication to random mutations, consistent with the results of a previous theoretical study[6]. Furthermore, we found that through a long-term replication system corresponding to at least 160 generations, NDK-RNAs evolved not only template activity for their own replication but also cooperation activity that supported the replication of other RNAs (Figs. 5c, d). These results demonstrate that the cooperative relationship can develop in compartments even with the continuous appearance of random mutations that are likely to abolish cooperation activity, supporting the development of cooperative molecular systems during the early evolution of life.

It should be noted that the cooperative structure of our cooperation system is unique, in that both gene products from Rep- and NDK-RNAs support the replication of themselves as well as that of other RNAs. Therefore, the cooperative replication can be more efficient as a consequence of increasing the expression of gene products for the purpose of own replication (i.e., selfish evolution). This special cooperative structure might be important for a certain level of compatibility of selfishness with cooperation observed in this study. This type of cooperative scheme is not unrealistic if we consider RNA-protein replication systems in the RNA-protein world. Any cooperation through the translation of metabolic enzymes, such as NDK, can be equally beneficial for both RNAs. This special type of cooperative scheme might have been common and allowed the development of cooperation in the RNA-protein world.

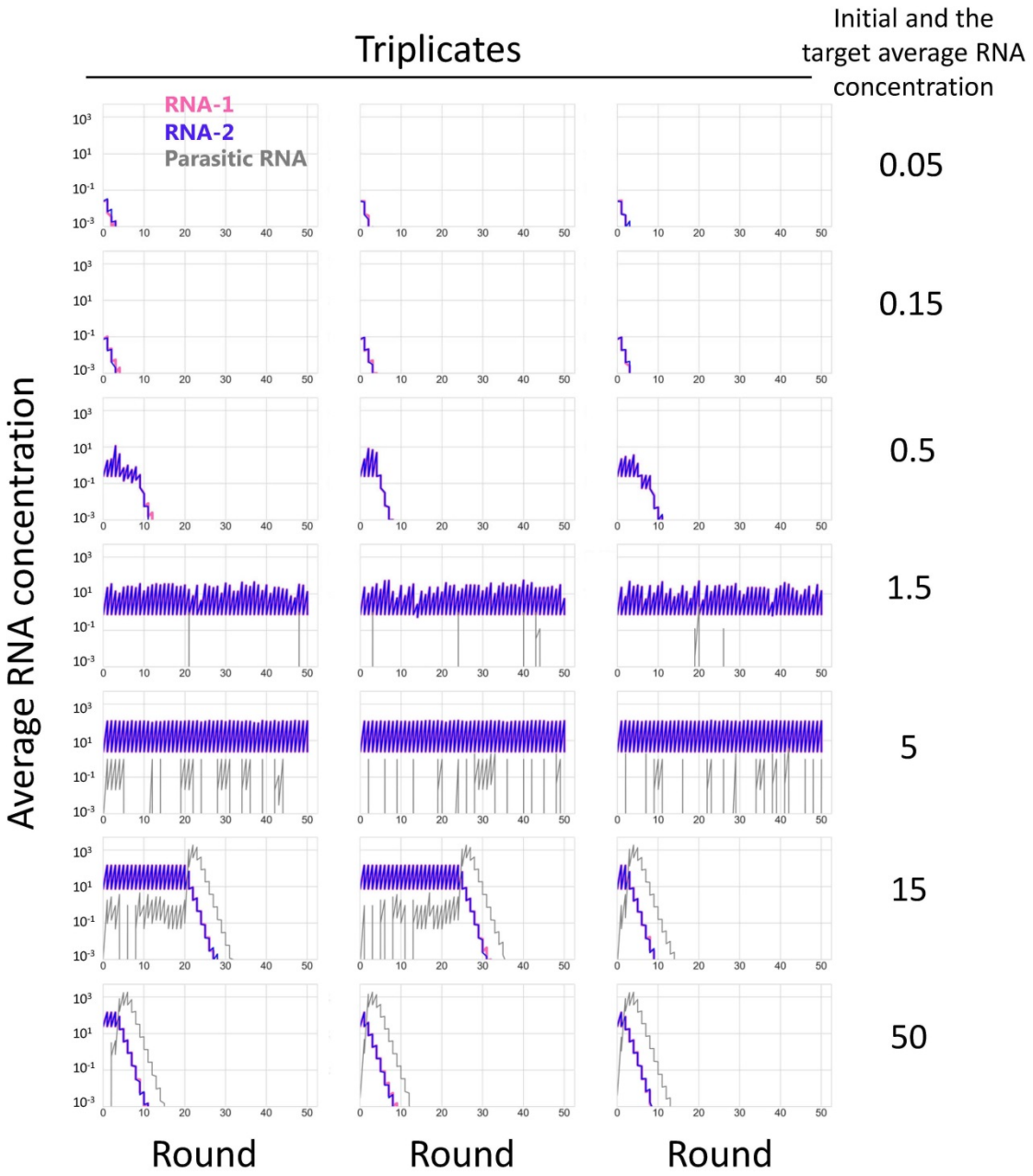
Differences between our mathematical model and previously published models

Previous theoretical models of cooperative replication differ from our simple model, based on our previous model without cooperation[5], in several ways. First, we made the

assumption of compartments with a rigid boundary, which are different from spatial structures with no clear boundaries[8,10-14]. Additionally, the timing of the compartment divisions in previous studies depended on the internal molecular concentrations or compartment size[6,8,10,15,16], while in this study, it is independent of the internal reaction, similar to the package model[17]. The occasional division occurring in our model and in the package model is apparently relevant to a primitive cellular structure that occasionally divides due to physical forces. Furthermore, we assumed variations in the RNA numbers in each compartment, while, in the previous package model, such numbers were considered constant in every compartment[17].

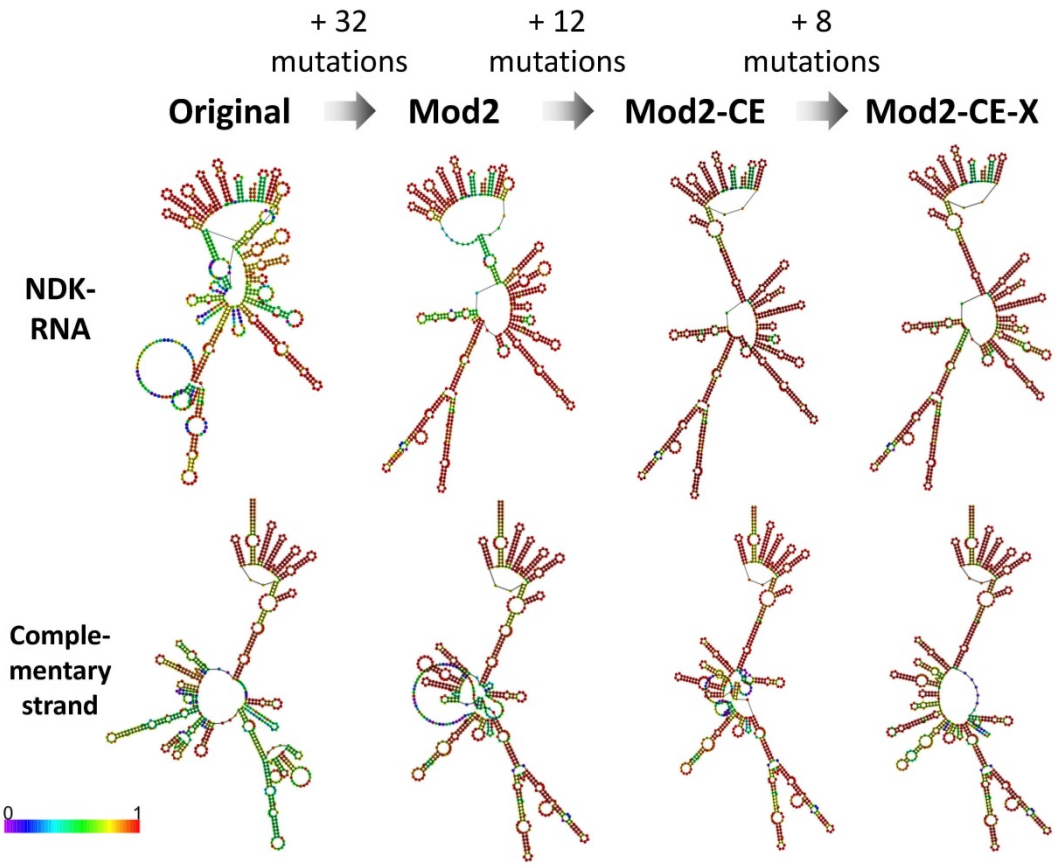
Future directions

We have not investigated here whether the two cooperating RNAs evolve into a single long RNA. The fusion of cooperating replicators into a single replicator is considered to be associated with the origin of chromosomes[18,19]. At least until round 50, we did not observe any indications that support the appearance of long RNA, although replicase is known to facilitate RNA linking through non-homologous RNA recombination[20]. One of the next important challenges would be to identify the conditions under which a single long RNA is selected rather than two distinct RNAs. The cooperative RNA replication system we constructed may represent a useful experimental model for the understanding of primitive replication systems and the development of complexity.



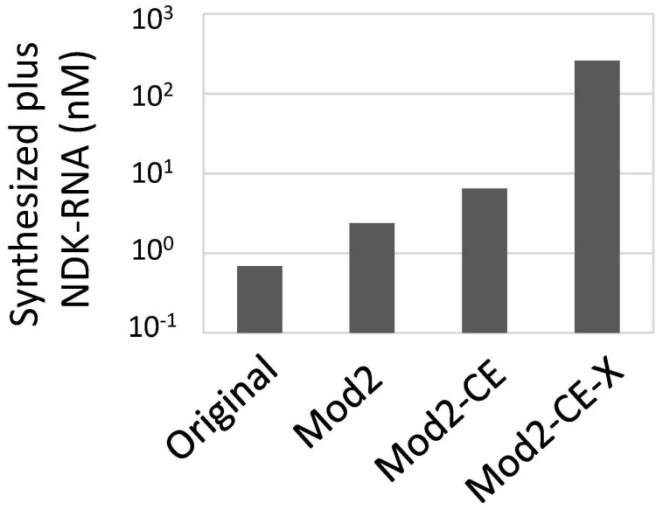
Supplementary Figure 1. Simulation of the cooperative replication at different RNA concentrations.

Simulations performed as those presented in Fig. 1d, using different initial RNA concentrations. Each simulation was performed three times independently.



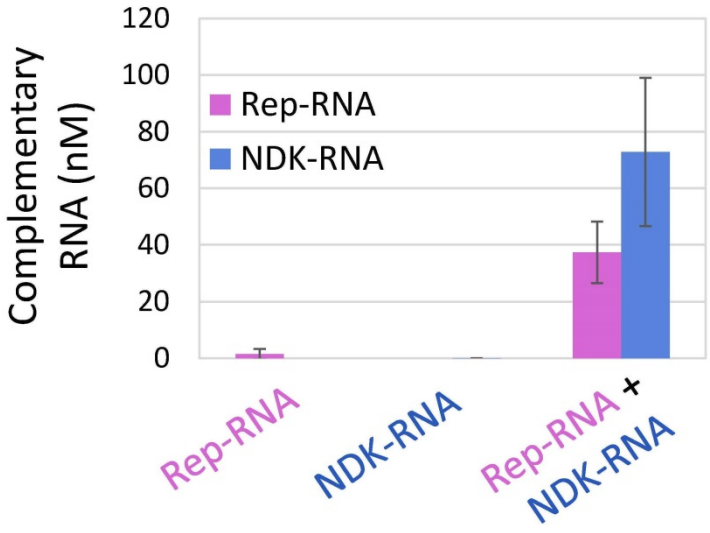
Supplementary Figure 2. Design of replicable NDK-RNAs.

Synonymous mutations were introduced in three steps, based on the rule of a replicable RNA (less GC number in loops) according to our previous study[21]. Secondary structures predicted using Vienna RNA (centroid structure)[22] were shown. The colors represent the probability of forming the local structures.

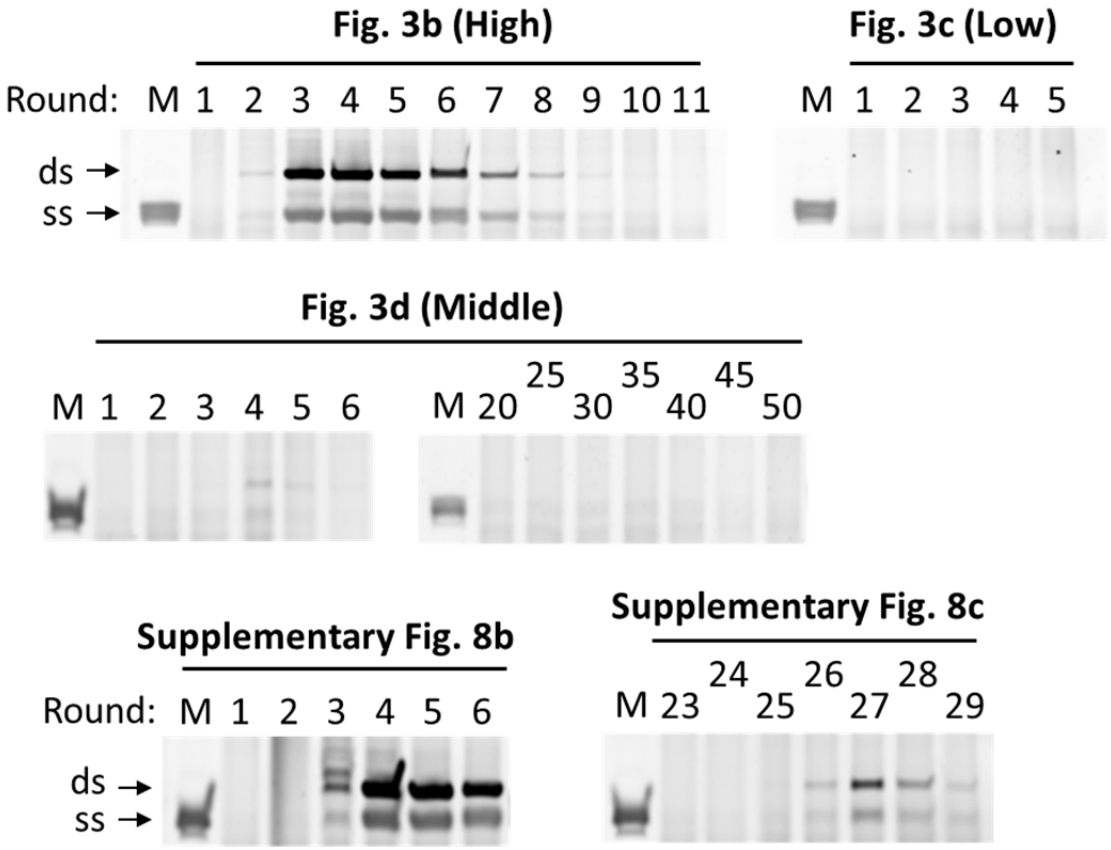


Supplementary Figure 3. TcCRR reactions with the designed NDK-RNAs.

The TcCRR reactions were conducted with 10 nM Rep- and NDK-RNAs at 37 °C for 4 h. Synthesized NDK-RNA concentrations were determined by using quantitative RT-PCR after selectively degrading the initial NDK-RNAs, labeled with α-S ribonucleotide analogue.

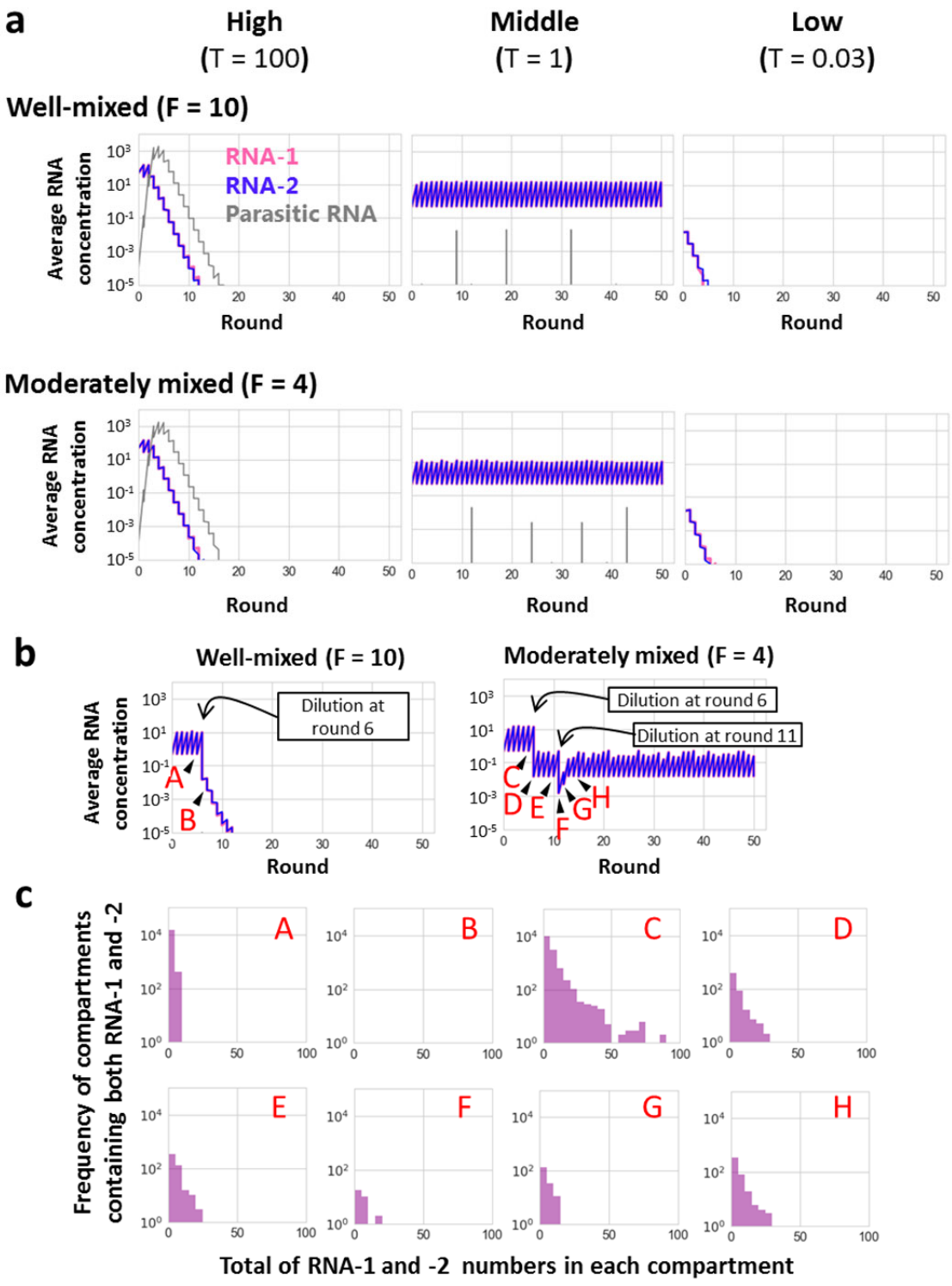


Supplementary Figure 4. Complementary strands synthesis during TcCRR reaction.
TcCRR reaction was conducted as described in the Fig. 2b legend and the complementary RNA concentration was determined by quantitative RT-PCR. The error bars indicate standard deviations (n=3). Measurements were taken from distinct samples.



Supplementary Figure 5. Detection of the parasitic RNA during the long-term replication experiments.

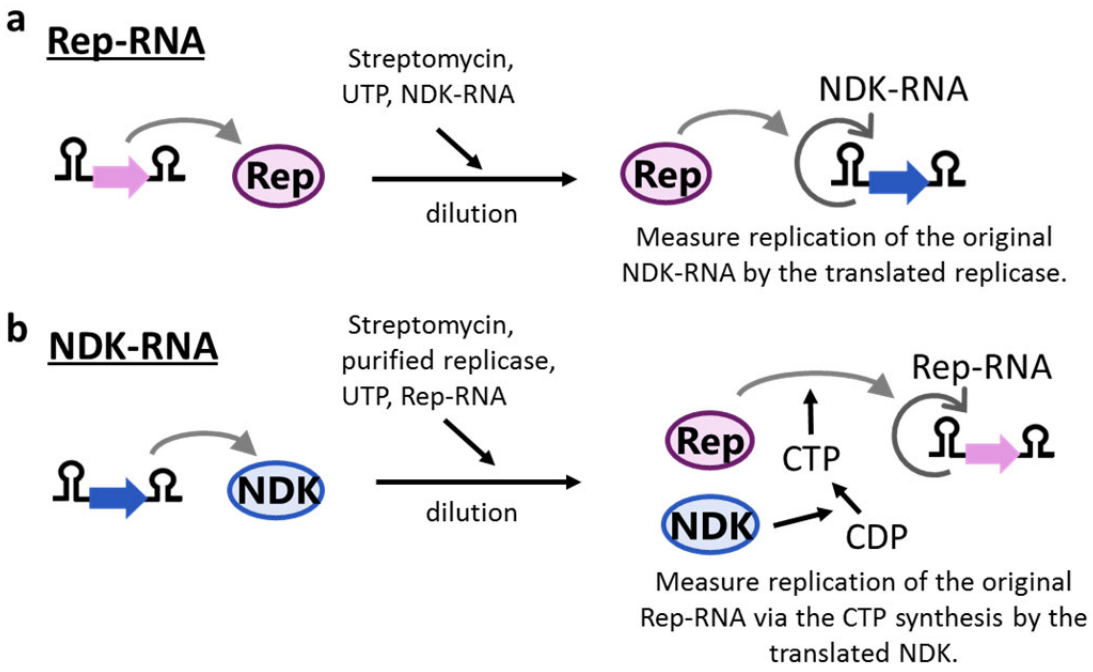
The reaction mixtures were subjected to native polyacrylamide-gel electrophoresis and stained with SYBR green II (Takara). ss and ds indicate the single- or double-strand of the parasitic RNA. s222 RNA, a typical parasitic RNA was used as the standard (M). Band intensities were quantified and plotted in Figs. 3b, and 3d, and Supplementary Figs. 8b, and 8c.



Supplementary Figure 6. The effect of the fusion-division rate (F) on the sustainability of cooperative replication.

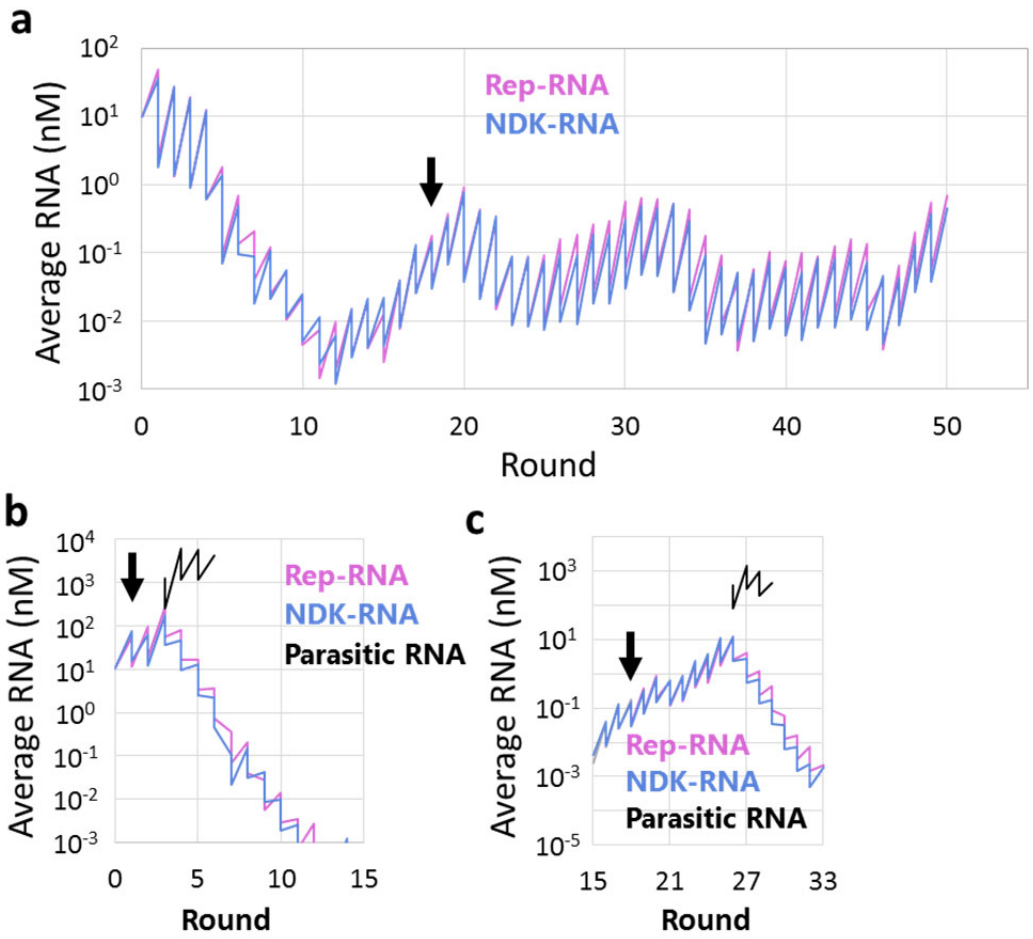
(a) Computer simulations of the long-term replication were performed using two different fusion-division rates (F). RNAs were shown to sustainably replicate only in the

mid-RNA concentration range, independent of the F values. (b) The effect of F values on the sustainability after compartment dilution. A simulation of the long-term replications was initiated at mid-range concentrations (T = 1), showing sustainable replications, after which the compartments were diluted approximately 30-fold at round 6 (indicated with arrows). Under the well-mixed conditions (F = 10), sustainable replication was not obtained (left), while under the moderately-mixed condition (F = 4), the replication continued (right). We further diluted the compartments 10-fold at round 11 (indicated with an arrow) under the moderately-mixed condition, and the replication continued sustainably. (c) The distribution of the RNAs in each compartment at points (A) – (H). At each point, we selected all compartments containing both RNA-1 and -2, and determined their concentrations before replication. At (A), in which the RNAs sustainably replicated under well-mixed conditions, the RNAs were almost evenly distributed and a sharp peak in compartment frequency was observed. After the dilution at round 6, no compartment was shown to contain both RNA-1 and -2 (B). Therefore, under the well-mixed condition, the RNAs could not replicate sustainably. In contrast, under the moderately mixed conditions, the distribution of RNA concentrations was broad (C) before the dilution. Even after the dilution at round 6, some compartments still contain both RNAs (D), which was maintained for several rounds (E). After the further dilution at round 11, some compartments still contained both RNAs (F) and their frequency increased with time (G and H), suggesting that the sustainability of cooperative RNA replication increases even after massive compartment dilution under the moderately-mixed conditions, since RNAs are distributed unevenly and therefore some compartments have a chance to contain both RNAs. This simulation explains the sustainable replication at lower concentrations observed in results presented in Figs. 3d, 3e and 3f. These simulations were performed using: M = 100000, p = 0.000001.



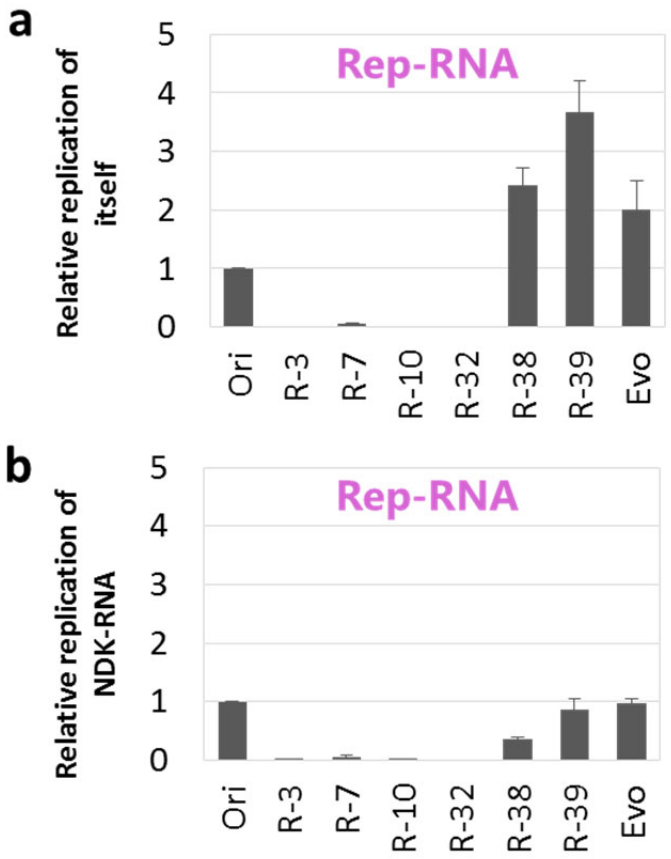
Supplementary Figure 7. Determining the cooperation activity of molecules.

(a) To determine Rep-RNA cooperation activity, replicase was translated from a Rep-RNA clone (10 nM) at 37 °C for 2 h without UTP to avoid replication. An aliquot was then diluted with a solution containing streptomycin, UTP, and the original NDK-RNA (10 nM) to stop translation and start replication. The mixture was incubated at 37 °C for 20 min and the synthesis of the complementary NDK-RNA strand was determined using quantitative RT-PCR. (b) To determine the cooperation activity of NDK-RNA, NDK was translated from a NDK-RNA clone (100 nM) at 37 °C for 2 h without UTP to avoid replication. An aliquot was then diluted with a solution containing streptomycin, UTP, the original Rep-RNA (100 nM), and 1 μM purified replicase to stop translation and initiate replication. The mixture was incubated at 37 °C for 45 min and the synthesis of the complementary strand of Rep-RNA was determined by using quantitative RT-PCR.



Supplementary Figure 8. Trajectories of RNA concentrations of other linages

(a) Trajectory of RNA concentrations of another lineage of a long-term replication. We separated the RNA population at round 18 (indicated with an arrow) of Fig. 3d and continued replication for an additional 32 rounds while maintaining the average concentrations at 0.005 – 1 nM independently. (b, c) Trajectories of RNA concentrations when the concentrations were increased to a higher range. The average concentration was increased by decreasing the dilution rate at round 1 (b) or 18 (c) (indicated with arrows). The Rep-RNA (pink line) and NDK-RNA (blue line) concentrations were measured by quantitative PCR after reverse-transcription. Parasitic RNA concentrations (black) were quantified after polyacrylamide gel electrophoresis (Supplementary Fig. 5).



Supplementary Figure 9. The activities of some of the Rep-RNA clones at round 50 to support own or NDK-RNA replication.

(a) The activity of the Rep-RNA clones to support the replication of itself. We chose six Rep-RNA clones out of 32 clones used in Fig. 4, and measured the activity to support own replication together with the original (Ori) and the evolved (Evo) RNAs. The experimental procedure was almost the same as that to measure the cooperation activity as schematically described in Supplementary Fig. 7a, but we used each Rep-RNA as a template instead of the original NDK-RNA. The measured activity levels were normalized to the levels of the original Rep-RNA. (b) The activity to support the original NDK-RNA activity to support NDK-RNAs replication (i.e., cooperation activity). The same data as Fig. 4b is shown again for comparison. Error bars represent standard errors (n = 3). Measurements were taken from distinct samples.

Supplementary Table 1. List of the mutations introduced during the NDK-RNA design.

All mutations are synonymous or located in the untranslated regions.

Replication improvement

Original → mod2		mod2 → mod 2-CE	mod2-CE → mod2-CE-X
C258U	C426U	G179U	C496U
G264A	C453U	A576U*	C501U
C267U	U483C	C603A	G504U
G270A	C519U	C609U	U534C
C282U	A522G	U618C	G537A
C285U	C549U	U621C	C540G
U294C	G561U	U624C*	U546C
C297U	C576A*	C630A	U552A
G303U	U585C	C636U	
C306U	C591U	G681C	
C324U	C594U	A691U	
C339U	U600G	U692A	
C357U	C624U*		
A363G	A639G		
A369U	C651A		
U420C	C662U		

*Mutations introduced again at the previously mutated sites.

Supplementary Table 2. Number of mutations at round 50.

	Rep-RNA	NDK-RNA
Total mutations	5.8 (1.9)	3.2 (1.4)
Untranslated region	1.6 (0.94)	1.3 (0.57)
Synonymous	1.8 (1.2)	1.5 (0.80)
Nonsynonymous	2.4 (1.2)	0.47 (0.84)

The standard deviations are shown in the parentheses.

Supplementary Table 3. List of mutations in 32 Rep- and NDK-RNA clones at round 50.

Rep-RNA: R-	1	3	4	5	6	7	8	10	11	14	15	17	18	19	20	21	22	23	24	25	27	29	30	31	32	33	35	36	37	38	39	40
C54T 5'UTR																+		+														
A92G 5'UTR																											+			+		
A93G 5'UTR																						+		+	+	+						
U164C 5'UTR																											+			+		
U166C 5'UTR																						+		+	+	+						
A184U 5'UTR					+			+					+			+		+			+								+			
A206G 5'UTR	+	+		+					+	+	+	+								+			+				+				+	+
A224C 5'UTR																	+											+				
U357C S44P					+																							+				
U386C I53I								+											+													
U390C F55L																						+		+	+	+				+		
A401G R58R																						+		+	+	+						
U657C S144P												+		+																		
U746C F173F	+	+		+					+	+	+												+				+				+	+
C923U F232F																									+	+						
G975A D250N	+	+		+					+	+	+	+								+			+	+			+				+	+
U1550C P441P																						+		+	+	+						
U1637C G470G																															+	+
1678-9A insert							+																		+							
A1817G P530P																								+		+						
U1832C F535F				+																			+									
G1891C R555T				+			+													+						+						+
U2005C 3'UTR											+									+												
unique mutations	5	1	4	4	4	8	5	2	3	1	2	3	2	5	3	4	5	2	1	3	3	3	2	2	2	0	2	3	4	2	0	2

NDK-RNA: N-	1	2	4	5	7	8	9	10	12	13	17	18	19	20	22	23	24	25	26	27	29	31	32	33	34	35	38	39	40	41	42	43
C86T 5'UTR	+	+		+	+	+					+	+		+		+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+
A193T 5'UTR															+								+									
T234G A2A	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+
T465C G79G							+	+												+												
T524C L99P			+										+																			
G550T A108S															+								+									
A612G E128E															+								+									
C615T I129I															+								+									
unique mutations	0	0	2	0	2	0	1	1	6	0	0	2	2	3	1	0	2	2	0	0	4	3	0	1	0	0	0	2	2	0	0	0

Supplementary Table 4. Clustering of the 32 Rep- and NDK-RNA clones at round 50.

Rep-RNA: R-	1	3	11	14	35	15	30	39	40	5	25	17	21	23	6	10	18	27	37	32	33	31	29	38	8	36	22	19	24	4	7	20
G975A D250N	+	+	+	+	+	+	+	+	+	+	+	+										+										
A206G 5'UTR	+	+	+	+	+	+	+	+	+	+	+	+																				
U746C F173F	+	+	+	+	+	+	+	+	+	+																						
A184U 5'UTR													+	+	+	+	+	+	+													
G1891C R555T									+	+	+										+				+							
U390C F55L																				+	+	+	+	+								
A93G 5'UTR																				+	+	+	+									
U166C 5'UTR																				+	+	+	+									
A401G R58R																				+	+	+	+									
U1550C P441P																				+	+	+	+									
U1637C G470G								+	+																							
U1832C F535F							+			+																						
U2005C 3'UTR						+					+																					
C54U 5'UTR													+	+																		
C923U F232F																				+	+											
A1817G P530P																					+	+										
A92G 5'UTR					+																			+								
U164C 5'UTR					+																			+								
A224C 5'UTR																										+	+					
U357C S44P															+											+						
U386C I53I																+													+			
U657C S144P												+																+				
1678-9A insert																				+					+							

NDK-RNA: N-	1	2	5	7	8	17	18	20	23	24	25	26	29	31	33	34	35	38	39	40	41	42	43	27	9	10	4	19	12	13	22	32
U234G A2A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
C86U 5'UTR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+								
U465C G79G																								+	+	+						
A193U 5'UTR																															+	+
U524C L99P																											+	+				
G550U A108S																															+	+
A612G E128E																															+	+
C615U I129I																															+	+

Supplementary Table 5. List of primers used in this study.

Primer 1: GCAAGTGACTCAGGATTCGTACGGTTTTCCATCGTGTTCAGC
Primer 2: TAAGCGAATGTTGCGAGCACAGGAGGATATACACATGGCTAT
Primer 3: AGGAGGATATACACATGGCTAT
Primer 4: GCAAGTGACTCAGGATTCGTAC
Primer 5: GGTTTTCCATCGTGTTCAGC
Primer 6: TAAGCGAATGTTGCGAGCAC
Primer 7: CCATCATTAAACCAAATGCAGTAGC
Primer 8: TAGGCTTGCGGCCGCAC
Primer 9: CATGTGTATATCTCCTTCTTAGAGTTAAAC
Primer 10: GGAGATATACACATGGCTATTGAACGTACTTTTTCC
Primer 11: GCGGCCGCAAGCCTAACGGGTGCGCGGGCAC
Primer 12: CCGGAAGGGGGGGACGAGG
Primer 13: GGGTCACCTCGCGCAGC
Primer 14: TAAGCGAATGTTGCGAGCACCCATCATTAAACCAAATGCAGTAGC
Primer 15: TAAGCGAATGTTGCGAGCACGTCACCGTATAGTGAGTCCTGC
Primer 16: ATGTTGCAGCTGTTTAGGCAG
Primer 17: TAAGCGAATGTTGCGAGCACCTCGTTTTGAAGCTGCAGGG
Primer 18: CTTCCAGCACAGAAACCACG

References

1 Takeuchi, N., Hogeweg, P. & Koonin, E. V. On the origin of DNA genomes: evolution of the division of labor between template and catalyst in model replicator systems. *PLoS Comput Biol* **7**, e1002024, doi:10.1371/journal.pcbi.1002024 (2011).

2 Takeuchi, N., Hogeweg, P. & Kaneko, K. The origin of a primordial genome through spontaneous symmetry breaking. *Nat Commun* **8**, 250, doi:10.1038/s41467-017-00243-x, 10.1038/s41467-017-00243-x [pii] (2017).

3 von der Dunk, S. H. A., Colizzi, E. S. & Hogeweg, P. Evolutionary Conflict Leads to Innovation: Symmetry Breaking in a Spatial Model of RNA-Like Replicators. *Life (Basel)* **7**, doi:E43 [pii], 10.3390/life7040043, life7040043 [pii] (2017).

4 Bansho, Y., Furubayashi, T., Ichihashi, N. & Yomo, T. Host-parasite oscillation dynamics and evolution in a compartmentalized RNA replication system. *Proc Natl Acad Sci U S A* **113**, 4045-4050, doi:10.1073/pnas.1524404113, 1524404113 [pii] (2016).

5 Furubayashi, T. & Ichihashi, N. Sustainability of a Compartmentalized Host-Parasite Replicator System under Periodic Washout-Mixing Cycles. *Life (Basel)* **8**, doi:E3 [pii], 10.3390/life8010003, life8010003 [pii] (2018).

6 Fontanari, J. F., Santos, M. & Szathmáry, E. Coexistence and error propagation in pre-biotic vesicle models: A group selection approach. *Journal of Theoretical Biology* **239**, 247-256, doi:10.1016/j.jtbi.2005.08.039 (2006).

7 Frank, S. A. The origin of synergistic symbiosis. *Journal of theoretical biology* **176**, 403-410, doi:10.1006/jtbi.1995.0208 (1995).

8 Hogeweg, P. & Takeuchi, N. Multilevel selection in models of prebiotic evolution: compartments and spatial self-organization. *Orig Life Evol Biosph* **33**, 375-403 (2003).

9 Zintzaras, E., Santos, M. & Szathmáry, E. Selfishness versus functional cooperation in a stochastic protocell model. *Journal of Theoretical Biology* **267**, 605-613, doi:10.1016/j.jtbi.2010.09.011 (2010).

10 Takeuchi, N. & Hogeweg, P. Multilevel selection in models of prebiotic evolution II: a direct comparison of compartmentalization and spatial self-organization. *PLoS Comput Biol* **5**, e1000542, doi:10.1371/journal.pcbi.1000542 (2009).

11 Boerlijst, M. C. & Hogeweg, P. Spiral Wave Structure in Pre-Biotic Evolution - Hypercycles Stable against Parasites. *Physica D* **48**, 17-28, doi:Doi 10.1016/0167-2789(91)90049-F (1991).

12 Szabo, P., Scheuring, I., Czaran, T. & Szathmary, E. In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity.

Nature **420**, 340-343, doi:10.1038/nature01187, nature01187 [pii] (2002).

13 Takeuchi, N. & Hogeweg, P. The role of complex formation and deleterious mutations for the stability of RNA-like replicator systems. *J Mol Evol* **65**, 668-686, doi:10.1007/s00239-007-9044-6 (2007).

14 Branciamore, S., Gallori, E., Szathmary, E. & Czaran, T. The Origin of Life: Chemical Evolution of a Metabolic System in a Mineral Honeycomb? *Journal of Molecular Evolution* **69**, 458-469, doi:10.1007/s00239-009-9278-6 (2009).

15 Szathmary, E. & Demeter, L. Group selection of early replicators and the origin of life. *J Theor Biol* **128**, 463-486 (1987).

16 Serra, R. a. V., Marco. *Modelling Protocells: The Emergent Synchronization of Reproduction and Molecular Replication*. (Springer, 2017).

17 Niesert, U., Harnasch, D. & Bresch, C. Origin of life between Scylla and Charybdis. *J Mol Evol* **17**, 348-353 (1981).

18 Maynard Smith, J. M., Szathmáry, E., Smith, J. M. & Szathmáry, E. in *Journal of theoretical biology* Vol. 164 437-446 (1993).

19 Santos, M. Origin of chromosomes in response to mutation pressure. *The American Naturalist* **152**, 751-756, doi:10.1007/BF00832686 (1998).

20 Chetverin, A. B., Chetverina, H. V., Demidenko, A. A. & Ugarov, V. I. Nonhomologous RNA recombination in a cell-free system: evidence for a transesterification mechanism guided by secondary structure. *Cell* **88**, 503-513 (1997).

21 Usui, K., Ichihashi, N. & Yomo, T. A design principle for a single-stranded RNA genome that replicates with less double-strand formation. *Nucleic Acids Research* **43**, 8033-8043, doi:10.1093/nar/gkv742 (2015).

22 Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26, doi:10.1186/1748-7188-6-26, 1748-7188-6-26 [pii] (2011).