Emergence of linkage between cooperative RNA replicators encoding replication and metabolic enzymes thorough experimental evolution

Authors: Kensuke Ueda[1], Ryo Mizuuchi[2,3*], Norikazu Ichihashi[1,2,4*]

[1]Department of Life Science, Graduate School of Arts and Science, the University of Tokyo, Meguro, Tokyo 153-8902, Japan

[2]Komaba Institute for Science, the University of Tokyo, Meguro, Tokyo 153-8902, Japan

[3]JST, PRESTO, Kawaguchi, Saitama 332-0012, Japan

[4]Universal Biology Institute, the University of Tokyo, Meguro, Tokyo 153-8902, Japan

*Corresponding authors:

mizuuchi@bio.c.u-tokyo.ac.jp and ichihashi@bio.c.u-tokyo.ac.jp

Abstract:

The integration of individually replicating genes into a primitive chromosome is a key evolutionary transition in the development of life, allowing the simultaneous inheritance of genes. However, how this transition occurred is unclear because of the extended size of primitive chromosomes, which replicate slower than unlinked genes. Theoretical studies have suggested that a primitive chromosome can evolve in the presence of cell-like compartments, as the physical linkage prevents the stochastic loss of essential genes upon division, but experimental support for this is lacking. Here, we demonstrate the evolution of a chromosome-like RNA from two cooperative RNA replicators encoding replication and metabolic enzymes. Through their long-term replication in cell-like compartments, linked RNAs emerged with the two cooperative RNAs connected end-to-end. The linked RNAs had different mutation patterns than the two unlinked RNAs, suggesting that they

were maintained as partially distinct lineages in the population. Our results provide experimental evidence supporting the plausibility of the evolution of a primitive chromosome from unlinked gene fragments, an important step in the emergence of complex biological systems.

Author Summary:

The integration of genes into a chromosome is a fundamental genetic organization in all extant life. The assembly of unlinked genes during prebiotic evolution was likely a major evolutionary transition toward the development of a complex cell. Decades of theoretical studies have suggested a plausible evolutionary pathway to a primitive chromosome from replicating RNA molecules that harbor cooperative genes within a protocell structure. However, demonstrating the evolution of a primitive chromosome in an experimental setup is challenging. We previously developed a cooperative RNA replication system in which two types of RNAs co-replicate using their self-encoded replication and metabolic enzymes. Using this system, in the present study, we demonstrate the evolution of a linkage between the two cooperative RNA replicators in compartments. An evolved "linked" RNA harbored the entire region of both genes, accumulated distinct mutations, and retained the ability to replicate using the two proteins translated from itself. These experimental findings support a prebiotic evolutionary scenario, in which unlinked genes assembled into a single genomic structure.

Introduction:

All extant cells have chromosomes that harbor multiple genes and ensure their organized inheritance. In the early evolution of life, such a genome organization may have been absent; unlinked RNA molecules encoding different functions or genes cooperated with each other for replicating the entire system [1–6]. The subsequent appearance of a linkage

between cooperative RNA replicators, or a primitive chromosome, is considered a major evolutionary transition toward complex biological organization [2–4]. Chromosome formation could have been both advantageous and disadvantageous for primitive life. Chromosomes enabled the synchronized replication of essential genes and potentially drove the evolution of efficient enzymes [7]. However, chromosomes were longer than unlinked RNA and hence required more time to replicate and had a higher chance of degradation. It remains unclear whether a primitive chromosome could evolve from individual cooperating RNA molecules despite these disadvantages.

Theoretical studies have suggested that the presence of primitive cell-like compartments (protocells) could have facilitated the selection of a primitive chromosome over individual RNA replicators [8–11]. The random assortment of cooperatively replicating RNAs upon protocell division may have caused the stochastic loss of cooperating partners. In contrast, the formation of a primitive chromosome ensured the co-encapsulation of cooperative genes, and therefore, a chromosome could have had an evolutionary advantage over unlinked replicators in compartments [8,9,11]. Although previous studies have established theoretical frameworks for the evolution of a primitive chromosome, empirical demonstration is lacking.

Previously, we constructed an experimental cooperative RNA replication system consisting of two RNAs and a reconstituted translation system [12]. One of the RNAs encodes a subunit of Qβ replicase, an RNA-dependent RNA polymerase derived from Qβ phage, and the other encodes a metabolic enzyme, nucleotide diphosphate kinase (NDK). We found that the cooperative RNAs can sustainably replicate and evolve in microscale water-in-oil droplets.

In the present study, we examined whether a linked RNA could evolve through the long-term replication of the two cooperative RNAs in water-in-oil droplets. We conducted two successive long-term replication experiments by gradually increasing RNA

concentrations, since higher RNA concentrations raise the likelihood of generating linked RNAs through recombination or ligation, which could be facilitated by the RNA polymerase [13–15] or occur spontaneously [16–18]. We found that linked RNAs comprising the two cooperating RNAs appeared during the experiment. Sequence analyses revealed that they harbored complete genes, as well as accumulated mutations that were distinct from those in the cooperative RNA fragments. Subsequent biochemical analysis confirmed that an emergent linked RNA replicated the entire sequence by expressing both encoded proteins. Our study provides experimental evidence that supports an evolutionary transition scenario of individually replicating genes to a primitive multi-gene chromosome.

Results:

Cooperative RNA replication system

The cooperative RNA replication system consists of a reconstituted cell-free translation system derived from *Escherichia coli* [19] and two types of cooperative RNA replicators (Rep- and NDK-RNAs) (Fig 1A) [12]. Rep-RNA (2041 bp) encodes the catalytic subunit of Qβ replicase (*rep* gene), which becomes active upon association with EF-Tu and EF-Ts in the translation system. NDK-RNA (752 bp) encodes NDK (*ndk* gene), which provides a substrate for RNA replication by converting cytidine diphosphate (CDP) into cytidine triphosphate (CTP). Therefore, Rep- and NDK-RNAs cooperatively replicate through the translation of their encoded enzymes. During replication, mutations occur, and occasional recombination generates a parasitic RNA that loses the gene region of either Rep- or NDK-RNA but replicates by exploiting replicase and NDK translated from other RNAs (Fig 1B). We expected that recombination or ligation between Rep- and NDK-RNAs facilitated by Qβ replicase [13–15] would result in the appearance of a linked RNA that harbors both *rep* and *ndk* genes (Fig 1B).

105 **Long-term replication experiments**

106 To investigate whether linked RNAs with the two genes appear from unlinked Rep- and
107 NDK-RNAs, we conducted long-term replication experiments. We repetitively performed
108 (1) RNA replication by incubating the cooperative RNA replication system at 37 °C for 4
109 or 6 h in water-in-oil droplets, (2) diluted the droplet population typically 5-fold, and (3)
110 induced fusion and division of droplets through vigorous mixing (Fig 1C). In step (3), a
111 fresh translation system was supplied to the RNA population, and the RNA molecules were
112 randomly redistributed among the droplets. We measured Rep- and NDK-RNA
113 concentrations after every replication step using quantitative PCR after reverse
114 transcription (RT-qPCR) with sequence-specific primers. We also measured parasitic RNA
115 concentrations in some rounds using native polyacrylamide gel electrophoresis (PAGE), as
116 their sequences were unknown.

117 In a previous long-term replication experiment, we maintained RNA concentrations
118 below a certain level (1 nM) by changing the dilution rate every round [12]. This was
119 important because high RNA concentrations cause frequent RNA recombination,
120 producing short parasitic RNA that replicates faster than Rep- and NDK-RNAs. The
121 parasitic RNA propagates throughout the compartments by fusion-division of the
122 compartments and disrupts the cooperation between Rep- and NDK-RNAs [12]. However,
123 in the present study, we increased RNA concentrations to induce recombination and
124 ligation, as these processes produce linked RNA. To achieve a higher RNA concentration
125 while repressing the amplification of parasitic RNA as much as possible, we employed a
126 new protocol to determine the dilution rate; the dilution rate was kept low (5-fold) until
127 RNA concentration reached a particular level, around which parasitic RNAs often appear,
128 and then increased to 100-fold for only a few rounds. We expected that this rapid dilution
129 would circumvent the propagation of parasitic RNAs among the compartments and allow
130 both Rep- and NDK-RNAs to replicate continuously while periodically reaching high

concentrations.

In the first long-term replication experiment, we initiated a serial dilution cycle with Rep- and NDK-RNAs obtained in our previous study (hereafter termed e1R0) [12]. In this experiment, we raised the dilution rate from 5- to 100-fold when the RNA concentration reached around 30 nM and successfully continued the replication of both Rep- and NDK-RNAs for 46 rounds (Fig 2A). Parasitic RNA concentrations were measured in four rounds with relatively high Rep- and NDK-RNA concentrations but barely detected (S1A Fig, the detection limit was approximately 10 nM). Next, we evaluated the presence of linked RNA in five rounds with relatively high Rep- and NDK-RNA concentrations by RT-PCR, using primers that could detect RNA with *rep* and *ndk* genes linked in the order 5′-*rep*-*ndk*-3′ (Fig 2B). If the entire region of *rep* and *ndk* genes is retained in the linked RNA, the expected size of the PCR products is between 2136 and 2456 bp. Using agarose gel electrophoresis, we detected approximately 2.5 kbp products in four rounds (Fig 2B), with band intensities roughly correlated with Rep- and NDK-RNA concentrations measured by RT-qPCR (Fig 2C). We also performed RT-PCR using primers for RNA with the two genes linked in the reverse order (5′-*ndk*-*rep*-3′) (S2A Fig) and detected approximately 2.3–2.5 kbp products in only two rounds (S2B Fig), without correlation with Rep- and NDK-RNA concentrations (S2C Fig). These results suggested that linked RNAs appeared during the serial dilution cycle.

Encouraged by the first experiment, we performed a second long-term replication experiment with one of the mutant Rep- and NDK-RNAs obtained at round 46 of the first experiment. These mutant Rep- and NDK-RNAs accumulated six and one mutations, respectively (S1 Table). The replication abilities of the mutant RNA pair (e2R0) were different from those of the original pair (e1R0); Rep- and NDK-RNA replications increased and decreased, respectively (S3 Fig).

In the second long-term replication experiment with the e2R0 pair, we increased the

157 dilution rate from 5- to 100-fold when the RNA concentration reached around 100 nM and
158 successfully continued the replication of both Rep- and NDK-RNAs for 79 rounds (Fig
159 3A). Parasitic RNA concentrations were measured in rounds 5–7, 20–22, 34–36, 54–56,
160 and 60–62, in which Rep- and NDK-RNA concentrations were relatively high, and detected
161 only in rounds 20–22 (Fig 3A, S1B Fig). The dilution rate was not increased around rounds
162 35 and 62 when parasitic RNA was undetectable; however, Rep- and NDK-RNA
163 concentrations decreased and then recovered. As a control, we also conducted a continuous
164 replication experiment without changing the dilution rate when the RNA concentration
165 reached 100 nM. In this case, parasitic RNA appeared and replicated excessively (S1C Fig,
166 S4 Fig), and then the concentrations of Rep- and NDK-RNAs decreased to undetectable
167 levels, confirming the importance of changing the dilution rate.

168 Next, we examined the appearance of a linked RNA by RT-PCR every 1–3 rounds,
169 using primers that could detect 5′-*rep*-*ndk*-3′ (Fig 3B). We detected approximately 2.3–2.5
170 kbp PCR products at rounds 6, 21, 22, 34, 35, 49, 55, 61, 62, and 79 and seemingly smaller
171 (2.0–2.3 kbps) products at rounds 19, 31, and 62. The relative intensities of the bands, if
172 detected, were as strong as those in the first long-term replication experiment (Fig 2B)
173 throughout the rounds. The band intensities were strong in rounds with high Rep- and
174 NDK-RNA concentrations (Fig 3C), with some exceptions (e.g., rounds 22, 37, 54, and 55)
175 where the product bands were undetected or barely detected despite high Rep- and NDK-
176 RNA concentrations. The concentration of a potentially linked RNA in the population at
177 round 79 was estimated to be less than 1 nM, which is less than 10% of either of the
178 unlinked RNAs, by RT-qPCR using primers that could detect the linkage region of 5′-*rep*-
179 *ndk*-3′. We also performed RT-PCR at rounds 6, 19, 34, 54, 62, and 79 using primers for
180 5′-*ndk*-*rep*-3′ (S2D Fig, S2E Fig); we detected approximately 2.3–2.5 kbps products at only
181 round 62.
182

183 **Sequence analysis**

184 Next, we obtained 16 clones of the potentially linked RNA products (5′-*rep*-*ndk*-3′) at round
185 79 of the second long-term replication experiment and analyzed their sequences. All
186 analyzed clones harbored the entire region of both *rep* and *ndk* genes in the 5′-*rep*-*ndk*-3′
187 order, confirming that the obtained long RNAs were the linked products of Rep- and NDK-
188 RNAs (hereafter referred to as RepNDK-RNA). We found two major ways by which Rep-
189 and NDK-RNAs were linked. In 56% of the RepNDK-RNA clones, Rep- and NDK-RNAs
190 were simply connected end-to-end without deletion or insertion. In 38% of the clones, a
191 single C was inserted between the connected Rep- and NDK-RNAs (Fig 4A, linkage
192 region). At the NDK-RNA site around the linkage, we also detected G3A and U12G in 75%
193 and 88% of the RepNDK-RNA clones, respectively.

194 Next, to investigate the origin of RepNDK-RNA at round 79, we compared the
195 sequences of the RepNDK-RNA clones with those of unlinked Rep- and NDK-RNAs. If
196 RepNDK-RNA was continuously replicated during the long-term replication experiment,
197 they could have accumulated mutations different from those found in unlinked Rep- and
198 NDK-RNAs in the same round. However, if RepNDK-RNA appeared from Rep- and NDK-
199 RNAs around round 79 or just as artificial products during the RT-PCR process, they should
200 have a similar set of mutations. To distinguish these possibilities, we obtained 16 clones of
201 unlinked Rep- and NDK-RNAs at round 79 and analyzed their sequences. The Rep- and
202 NDK-RNA clones contained 3−8 and 0−5 mutations, respectively (S2 Table, S3 Table).

203 Fig 4A lists all mutations detected in more than 30% of the clones of any of the three
204 RNA species (unlinked Rep- and NDK-RNAs and linked RepNDK-RNAs). The three types
205 of RNAs accumulated different mutations at their corresponding sites. For example,
206 A825G and U1776C were found in 31% and 25% of the Rep-RNA clones, respectively,
207 whereas both mutations were detected in 100% of the RepNDK-RNA clones. In contrast,
208 U1349C and A1604G were observed in 56% and 63% of the Rep-RNA clones, respectively,

209 but were not detected in the RepNDK-RNA clones. Similarly, C228U was found in only
210 6% of the NDK-RNA clones but in 44% of the RepNDK-RNA clones, whereas U582C was
211 detected in 69% of the NDK-RNA clones but in only 13% of the RepNDK-RNA clones.
212 These different mutation patterns indicate that the RepNDK-RNAs analyzed at round 79
213 were not derived from Rep- and NDK-RNAs of the same round. These results also
214 suggested that RepNDK-RNA was maintained in the population for a certain period, long
215 enough to form its own lineage.

216 We further illustrated the evolutionary relationships between Rep- or NDK-RNA and
217 RepNDK-RNA using phylogenetic trees for each of the Rep- and NDK-RNA regions (Fig
218 4B). The three RNA species are represented by different symbols. For the Rep-RNA region,
219 most of the unlinked Rep-RNA clones (red circles) and linked RepNDK-RNA clones
220 (green squares) were clustered in different branches, while some Rep-RNA clones (clones
221 11, 2, 5, and 6) were present in the RepNDK-RNA cluster. For the NDK-RNA region, the
222 unlinked NDK-RNA clones (blue triangles) and linked RepNDK-RNA clones (green
223 squares) were mixed in the tree. These results suggested that RepNDK-RNA formed
224 lineages that were not completely but partially distinct from the unlinked RNAs.

225

226 **Biochemical characterization of RepNDK-RNA**

227 To understand the biochemical characteristics of the linked RepNDK-RNA at round 79, we
228 selected a clone containing only the most common mutation set (A825G, T963C, U1776C,
229 and -2042C in the Rep-RNA region and G3A and U12G in the NDK-RNA region), which
230 was found in 25% of the 16 clones. For comparison, we also chose a pair of unlinked Rep-
231 and NDK-RNAs containing only the most common mutation sets at round 79. The selected
232 Rep-RNA contained U963C, U1349C, and A1604G, found in 44% of the Rep-RNA clones,
233 whereas the selected NDK-RNA contained A240G, U465C, G498A, and U582C, found in
234 38% of the NDK-RNA clones. The replication of these Rep- and NDK-RNAs (e2R79) was

235 similar to that of the ancestral Rep- and NDK-RNA pair (e2R0) with slightly improved
236 NDK-RNA replication (S3 Fig).

237 RepNDK-RNA was expected to replicate by expressing both encoded replicase and
238 NDK. To test this, we incubated the RepNDK-RNA clone in a reconstituted translation
239 system containing fluorescent-labeled lysyl-tRNA. The newly synthesized proteins were
240 analyzed by sodium dodecyl sulfate (SDS)-PAGE, followed by fluorescent imaging. We
241 found that bands corresponding to both the replicase subunit and NDK were detected for
242 the RepNDK-RNA (Fig 5A). The protein amounts estimated from the band intensities were
243 approximately half of those of the Rep- and NDK-RNAs (Fig 5B). Next, we investigated
244 whether the translation of the two proteins was coupled with RepNDK-RNA replication.
245 During RNA replication, the replicase first synthesizes the minus (complementary) strand,
246 which is then recognized by the replicase for plus-strand synthesis (Fig 5C). We incubated
247 the RepNDK-RNA clone with a translation system in water-in-oil droplets at 37 °C for 4 h
248 and analyzed the synthesis of full-length plus and minus strands by RT-PCR. Agarose gel
249 electrophoresis showed that the intensities of bands corresponding to both strands increased
250 after incubation (Fig 5D), demonstrating the ability of the RepNDK-RNA to replicate
251 without the help of unlinked RNAs. To quantitatively compare the replication of the linked
252 RNA with that of the unlinked RNAs, we measured the replication of the RepNDK-RNA
253 by RT-qPCR. The extent of RepNDK-RNA replication was less than 10% of Rep- and
254 NDK-RNA replications (Fig 5E), which may be caused by the reduced translation activity.
255

256 **Discussion**

257 The integration of distinct genetic information on a primitive chromosome has been
258 considered a major evolutionary transition in the development of life [2–4]. In this study,
259 we demonstrated that such a linkage emerged during the long-term replication of two
260 cooperative RNAs encoding replication and metabolic enzymes in cell-like compartments.

261 The linked RNAs that existed in the final population retained both genes, maintained their
262 translation and replication abilities, and accumulated mutations different from those found
263 in the unlinked RNAs (Fig 4 and 5), suggesting that the linked RNA continuously replicated
264 in the RNA population for at least a certain period. These results provide experimental
265 evidence that individual RNA replicators encoding different genes can assemble into a
266 single RNA molecule through Darwinian evolution. This process may explain the origin of
267 primitive chromosomes during prebiotic evolution.

268 Although the accumulation of unique mutations in RepNDK-RNA indicated their
269 maintenance in the population, at round 79, the linked RNA replicated less efficiently than
270 the pair of Rep- and NDK-RNAs (Fig 5E), possibly because of its lower translational
271 activity (Fig 5B), extended length, and reorganized RNA structures. How the linked RNA
272 was maintained in the population while competing with the unlinked RNAs that replicate
273 faster remains unexplained. One possible reason could be that, unlike unlinked Rep- and
274 NDK-RNAs, RepNDK-RNA can always replicate on its own after redistribution in the
275 droplet population, as suggested by theoretical studies [8,9,11]. This effect is expectedly
276 enhanced when the stochastic mis-encapsulation of Rep- and NDK-RNAs becomes
277 effective at low RNA concentrations [12]. Experiments using lower RNA concentrations
278 for long-term replication could provide more selective advantages for RepNDK-RNA.

279 The generation of a linkage between RNA fragments is not unique to our cooperative
280 RNA replication system. Diverse RNA-dependent RNA polymerases (RdRp), including
281 Qβ replicase, are known to cause intermolecular RNA recombination during replication
282 [20–22], which can link multiple RNA molecules. End-to-end ligation, as detected in the
283 present study (with or without single nucleotide insertion), has also been observed for
284 various RdRp such as through end-to-end template switching [14,23]. Thus, whether linked
285 RNAs can evolve in other *in vitro* RNA replication systems using different RNA replicases
286 [24] needs to be investigated. Furthermore, even before the emergence of proteins, such

recombination and ligation of RNA molecules could have occurred spontaneously [16–18] or with the assistance of catalytic RNAs [25–29]. Therefore, it is reasonable to assume that the linkage of cooperative RNA molecules was prevalent during the early evolution of life.

We also developed a new method that allows the long-term replication of cooperative RNAs, even if the RNA concentrations increase tentatively and parasitic RNA emerges. Our previous study suggested that maintaining a relatively low RNA concentration was crucial for sustained replication of the cooperative system, and continuous RNA replication was achieved only by controlling the dilution rate at every replication round [12]. However, the present study demonstrated that sustainable cooperative RNA replication does not require strict regulation of RNA concentration, as previously believed, supporting the robustness of molecular cooperation. A higher RNA concentration also increases the chance of recombination or ligation of RNAs and may facilitate the evolution of a primitive chromosome.

The evolution of a linked RNA from multiple RNAs that encode different genes may also be useful for developing an artificial replicable RNA genome. To the best of our knowledge, RepNDK-RNA is the first artificial RNA genome that replicates based on the translation of more than one encoded protein. The manual expansion of an RNA genome by introducing new genes is challenging, as it typically disrupts RNA structures and impairs its ability to undergo replication and translation. Our previous study connected two RNAs based on the predicted RNA structures of linked RNAs while maintaining their replication coupled with the translation of one encoded protein [30]. However, further introduction of genes would make genome expansion more difficult as the accuracy of RNA structural production decreases with increasing RNA length. Although further refinement is necessary, the evolutionary technique to integrate genes into a single RNA genome, as demonstrated here, could be utilized to create an optimized RNA genome encoding multiple proteins and contribute to the development of artificial cells that possess a genome replication system

313 [31,32].

314

315 **Materials and Methods**

316 **Plasmids and RNAs**

317 Two plasmids, each encoding Rep- or NDK-RNA "e1R0," were obtained in the previous
318 study as the plasmids encoding Rep- or NDK-RNA "Evo" [12]. The other plasmids, each
319 encoding one of the two Rep- and NDK-RNA clones ("e2R0" or "e2R79") or the RepNDK-
320 RNA clone, were obtained in the present study through cloning, as described below, and
321 site-specific mutagenesis. For the RepNDK-RNA clone, the entire cDNA length was PCR-
322 amplified from the corresponding plasmid. The cDNA of the RepNDK-RNA clone and the
323 plasmids for all other RNA clones were subjected to digestion with Sma I (Takara) and *in*
324 *vitro* transcription with T7 RNA polymerase (Takara). All transcribed RNAs were purified
325 using the RNeasy Mini Kit (QIAGEN).

326

327 **Preparation of the reconstituted translation system**

328 The composition of the reconstituted translation system was as described previously [12]
329 (based on the reconstituted *Escherichia coli* translation system [19]) except that the trigger
330 factor was omitted because it was not essential for translation and contained a high level
331 of NDK activity that could not be reduced by the following purification steps. All protein
332 components of the translation system were purified by two successive affinity column
333 chromatography in a stringent buffer to further reduce the remaining NDK activity derived
334 from *Escherichia coli*. For all proteins except ribosomes, the re-purification procedure was
335 the same as that used for removing tRNA from EF-Tu, described in our previous study [33].
336 Ribosomes were purified as described previously [19] and then washed with another
337 stringent buffer (20 mM Hepes-KOH (pH 7.6), 6 mM magnesium acetate, 7 mM 2-
338 mercaptoethanol, 1% Triton X-100, 1 mM dithiothreitol, and 0.33 M potassium chloride).

339 Briefly, the ribosomes were diluted 20-fold with the stringent buffer and ultracentrifuged
340 at 150,000 g for 2 h at 4 °C. After removing the supernatant, the precipitate was rinsed with
341 70S buffer (20 mM Hepes-KOH (pH 7.6), 6 mM magnesium acetate, 7 mM 2-
342 mercaptoethanol, and 0.03 M potassium chloride). After carefully removing the remaining
343 buffer, the precipitate was dissolved in 70S buffer. Then, the solution was diluted again
344 with the stringent buffer and collected by ultracentrifugation, as described above. To
345 remove the residual stringent buffer, the final ribosome solution was diluted with 70S
346 buffer 10-fold and then concentrated using Amicon Ultra (30 kDa cut, Merck) three times.
347

348 **Assay of remaining NDK activity**

349 The remaining NDK activity in each protein component before and after the purification
350 step was estimated as the activity to convert adenosine diphosphate (ADP) into adenosine
351 triphosphate (ATP). First, each protein component (twice the concentration of that in the
352 translation system) was incubated with 2.5 mM ADP and 1.25 mM CTP in a reaction buffer
353 (100 mM Hepes-KOH (pH 7.6), 70 mM glutamic acid potassium salt, 0.375 mM
354 spermidine, and 11 mM magnesium acetate) at 37 °C for 2 h. Then, an aliquot of the
355 reaction was used to determine the amount of synthesized ATP using ATP Assay Kit
356 (Colorimetric/Fluorometric) (Abcam). The assay was performed at 37 °C, and the
357 fluorescence intensity was measured using Mx3005P Real-Time PCR System (Agilent
358 Technologies) as the indicator of ATP synthesis (S5 Fig).
359

360 **Long-term replication experiment**

361 The experiments were performed as described previously [12] with several modifications.
362 The translation system (10 μL) containing certain concentrations of Rep- and NDK-RNAs
363 "e1R0" (Fig 2A) or "e2R0" (Fig 3A and S4 Fig) were added to 1,000 μL of buffer-saturated
364 oil and mixed vigorously with a homogenizer (POLYTRON PT 1300D, KINEMATICA)

365 at 16,000 rpm for 1 min on ice to obtain water-in-oil droplets. The preparation of the
366 saturated oil was described in the previous study [34]. The droplets were incubated at 37 °C
367 for 6 h (Fig 2A) or 4 h (Fig 3A and S4 Fig) to induce RNA replication through protein
368 translation. After incubation, an aliquot of the droplets was diluted 5-fold or 100-fold with
369 fresh buffer-saturated oil, as described in the main text and shown in the corresponding
370 figures. The solution was mixed with 10 μL of the translation system and homogenized
371 using the same method to obtain a new droplet population, followed by incubation at 37 °C
372 for 4 or 6 h for the next round of RNA replication. In Fig 2A, the original droplet population
373 was diluted 100-fold using the same method as that before incubation to make the initial
374 population (approximately 0.1 nM of Rep- and NDK-RNAs). Rep- and NDK-RNA
375 concentrations were determined at every round by RT-qPCR with primers 1 and 2 (Rep-
376 RNA) or 3 and 4 (NDK-RNA) (S4 Table). The measurement was performed after diluting
377 the droplets 100-fold with 1 mM EDTA (pH 8.0) and using One Step TB Green PrimeScript
378 PLUS RT-PCR Kit (Takara).
379
380 **Measurement of parasitic RNA concentration**
381 In some rounds of the long-term replication experiments, the water phase containing RNA
382 was collected from the droplets (100 μL) by centrifugation (22,000 g, 5 min). The
383 recovered phase was mixed with diethyl ether (40 μL) and centrifuged (11,000 g, 1 min)
384 to remove the diethyl ether phase. Then, RNA was purified with the RNeasy Mini Kit
385 (QIAGEN) and subjected to 8% polyacrylamide gel electrophoresis in 1 × TBE buffer. The
386 fluorescence intensities of bands corresponding to parasitic RNAs were quantified using
387 ImageJ (NIH) after staining with SYBR Green II (Takara). The concentrations of parasitic
388 RNAs were determined from the intensities based on a dilution series of a standard parasitic
389 RNA (s222 [35]).
390

391 **RT-PCR and Sequence analysis**

392 RNA samples were obtained from the long-term replication experiments as described
393 above. Rep- and NDK-RNAs were reverse transcribed with primer 8, PCR-amplified with
394 primers 7 and 8 (S4 Table), and separated using 0.8% agarose gel electrophoresis with E-
395 Gel CloneWell (Thermo Fisher Scientific). The cDNA samples were cloned into a pUC19
396 vector (PCR-amplified with primers 11 and 12) using In-Fusion HD Cloning Kit (Takara).
397 After transformation into *Escherichia coli*, the sequences of randomly selected plasmids
398 were analyzed. Similarly, RepNDK-RNA was reverse transcribed with primer 10, PCR-
399 amplified with primers 9 and 10, size-selected, and cloned using the same method. The 5′
400 and 3′ untranslated regions of the RepNDK-RNA clones could not be retrieved to
401 distinguish them from Rep- and NDK-RNAs in the same RNA population. The sequence
402 information of the linkage region was obtained only for the RepNDK-RNA clones. To
403 detect putative elongated RNAs with *rep* and *ndk* genes linked in this order (Fig 2C and
404 3C) or the reverse order (S2 Fig), RT-PCR was performed with primers 9 and 10 or 13 and
405 14, respectively. PCR products were visualized by agarose gel electrophoresis and stained
406 with SAFELOOK Green Nucleic Acid Stain (FUJIFILM), and band intensities were
407 determined using ImageJ (NIH).

408

409 **Phylogenetic analysis**

410 The sequences of the Rep-, NDK-, and RepNDK-RNA clones obtained at round 79 of the
411 second long-term replication experiment (Fig 3A) were subjected to phylogenetic analysis.
412 The RepNDK-RNA clones were divided into Rep-RNA and NDK-RNA regions and
413 compared with the Rep- and NDK-RNA clones. Only the sequence regions of the
414 RepNDK-RNA clones and their respective unlinked RNAs for which mutation information
415 was commonly available were used for the analysis. Phylogenetic trees were created using
416 the neighbor-joining method in MEGA11 [36] by assuming the same mutation rate for

point mutations, single nucleotide deletions, and single nucleotide insertions. Phylogenetic trees were visualized using Interactive Tree Of Life (iTOL) [37].

Translation-coupled RNA replication experiments

A pair of Rep- and NDK-RNA clones (10 nM each) or the RepNDK-RNA clone (10 nM) was incubated with the translation system in water-in-oil droplets at 37 °C for 4 h. The concentration of each RNA clone was determined by RT-qPCR as described above. Primers 5 and 6 were used for the RepNDK-RNA clone (S4 Table).

Analysis of protein translation

300 nM of Rep-, NDK-, or RepNDK-RNA was incubated at 37 °C for 12 h in a translation system and FluoroTect GreenLys tRNA (Promega), with neither CTP nor CDP to preclude RNA replication. In this experiment, we used a translation system before excessive purification of protein components, as described previously [38], to enhance translation efficiency and obtain measurable amounts of proteins. After translation, an aliquot was treated with 0.1 mg/mL RNase A (QIAGEN) at 37 °C for 15 min to digest fluorescently labeled lysine tRNA. Then, the solution was incubated at 95 °C for 4 min in SDS sample buffer (50 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl, pH 7.4), 2% SDS, 0.86 M 2-mercaptoethanol, and 10% glycerol) and subjected to SDS-PAGE using a 10–20% gradient gel (Funakoshi, Japan). The synthesized proteins that incorporated fluorescently labeled lysine were visualized using FUSION-SL4 (Vilber-Lourmat), and band intensities were determined using ImageJ (NIH).

Acknowledgments

We are grateful to Ms. Nana Kuroda for technical support.

443 **References**

444 1. Eigen M, Schuster P. A principle of natural self-organization - Part A: Emergence
445 of the hypercycle. Naturwissenschaften. 1977;64: 541–565.
446 2. Maynard-Smith J, Szathmáry E. The major transitions in evolution. Oxford
447 University Press, Oxford, UK; 1997.
448 3. Brosius J. Gene duplication and other evolutionary strategies: From the RNA
449 world to the future. J Struct Funct Genomics. 2003;3: 1–17.
450 4. Durand PM, Michod RE. Genomics in the light of evolutionary transitions.
451 Evolution. 2010;64: 1533–1540.
452 5. Higgs PG, Lehman N. The RNA world: Molecular cooperation at the origins of
453 life. Nat Rev Genet. 2015;16: 7–17.
454 6. Adamski P, Eleveld M, Sood A, Kun Á, Szilágyi A, Czárán T, et al. From self-
455 replication to replicator systems en route to de novo life. Nat Rev Chem. 2020;4:
456 386–403.
457 7. Szilágyi A, Kun Á, Szathmáry E. Early evolution of efficient enzymes and genome
458 organization. Biol Direct. 2012;7: 38.
459 8. Maynard Smith J, Szathmáry E. The Origin of Chromosomes I. Selection for
460 Linkage. J Theor Biol. 1993;164: 437–446.
461 9. Santos M. Origin of chromosomes in response to mutation pressure. Am Nat.
462 1998;152: 751–756.
463 10. Ma W, Yu C, Zhang W. Circularity and self-cleavage as a strategy for the
464 emergence of a chromosome in the RNA-based protocell. Biol Direct. 2013;8: 21.
465 11. Szilágyi A, Kovács VP, Szathmáry E, Santos M. Evolution of linkage and genome
466 expansion in protocells: The origin of chromosomes. PLoS Genet. 2020;16:
467 e1009155.
468 12. Mizuuchi R, Ichihashi N. Sustainable replication and coevolution of cooperative

469 RNAs in an artificial cell-like system. Nat Ecol Evol. 2018;2: 1654–1660.
470 13. Biebricher CK, Luce R. In vitro recombination and terminal elongation of RNA by
471 Qβ replicase. EMBO J. 1992;11: 5129–5135.
472 14. Chetverin AB, Chetverina H V., Demidenko AA, Ugarov VI. Nonhomologous
473 RNA recombination in a cell-free system: Evidence for a transesterification
474 mechanism guided by secondary structure. Cell. 1997;88: 503–513.
475 15. Chetverin AB. Thirty Years of Studies of Qβ Replicase: What Have We Learned
476 and What Is Yet to Be Learned? Biochem. 2018;83: S19–S32.
477 16. Chetverina H V., Demidenko AA, Ugarov VI, Chetverin AB. Spontaneous
478 rearrangements in RNA sequences. FEBS Lett. 1999;450: 89–94.
479 17. Mutschler H, Taylor AI, Lightowler A, Houlihan G, Abramov M, Herdewijn P, et
480 al. Innate potential of random genetic oligomer pools for recombination. Elife.
481 2018;7: e43022.
482 18. Smail BA, Clifton BE, Mizuuchi R, Lehman N. Spontaneous advent of genetic
483 diversity in RNA populations through multiple recombination mechanisms. RNA.
484 2019;25: 453–464.
485 19. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, et al. Cell-
486 free translation reconstituted with purified components. Nat Biotechnol. 2001;19:
487 751–5.
488 20. Lai MMC. RNA recombination in animal and plant viruses. Microbiol Rev.
489 1992;56: 61–79.
490 21. Chetverin AB, Kopein DS, Chetverina H V., Demidenko AA, Ugarov VI. Viral
491 RNA-directed RNA polymerases use diverse mechanisms to promote
492 recombination between RNA molecules. J Biol Chem. 2005;280: 8748–8755.
493 22. Bujarski JJ. Genetic recombination in plant-infecting messenger-sense RNA
494 viruses: Overview and research perspectives. Front Plant Sci. 2013;4: 68.

495 23. Kim MJ, Kao C. Factors regulating template switch in vitro by viral RNA-
496 dependent RNA polymerases: Implications for RNA-RNA recombination. Proc
497 Natl Acad Sci U S A. 2001;98: 4972–4977.
498 24. Weise LI, Heymann M, Mayr V, Mutschler H. Cell-free expression of RNA
499 encoded genes using MS2 replicase. Nucleic Acids Res. 2019;47: 10956–10967.
500 25. Ekland EH, Szostak JW, Bartel DP. Structurally Complex and Highly Active RNA
501 Ligases Derived from Random RNA Sequences. Science. 1995;269: 364–370.
502 26. Robertson MP, Ellington AD. In vitro selection of an allosteric ribozyme that
503 transduces analytes to amplicons. Nat Biotechnol. 1999;17: 62–66.
504 27. Shippy R, Lockner R, Farnsworth M, Hampel A. The hairpin ribozyme:
505 Discovery, mechanism, and development for gene therapy. Appl Biochem
506 Biotechnol - Part B Mol Biotechnol. 1999;12: 117–129.
507 28. Rogers J, Joyce GF. The effect of cytidine on the structure and function of an RNA
508 ligase ribozyme. RNA. 2001;7: 395–404.
509 29. A.Riley C, Lehman N. Generalized RNA-Directed Recombination of RNA. Chem
510 Biol. 2003;10: 1233–1243.
511 30. Ueda K, Mizuuchi R, Matsuda F, Ichihashi N. A Fusion Method to Develop an
512 Expanded Artificial Genomic RNA Replicable by Qβ Replicase. ChemBioChem.
513 2019;20: 2331–2335.
514 31. Forster AC, Church GM. Towards synthesis of a minimal cell. Mol Syst Biol.
515 2006;2: 45.
516 32. Jewett MC, Forster AC. Update on designing and building minimal cells. Curr
517 Opin Biotechnol. 2010;21: 697–703.
518 33. Miyachi R, Shimizu Y, Ichihashi N. Transfer RNA Synthesis-Coupled Translation
519 and DNA Replication in a Reconstituted Transcription/Translation System. ACS
520 Synth Biol. 2022;10: 3507–3517.

34. Ichihashi N, Usui K, Kazuta Y, Sunami T, Matsuura T, Yomo T. Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment. Nat Commun. 2013;4: 2494.

35. Hosoda K, Matsuura T, Kita H, Ichihashi N, Tsukada K, Yomo T. Kinetic analysis of the entire RNA amplification process by Qβ replicase. J Biol Chem. 2007;282: 15516–15527.

36. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38: 3022–3027.

37. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49: W293–W296.

38. Mizuuchi R, Ichihashi N, Yomo T. Adaptation and diversification of an RNA replication system under initiation- or termination-impaired translational conditions. ChemBioChem. 2016;17: 1229–1232.

Supporting information captions

S1 Fig. Detection of parasitic RNAs during the long-term replication experiments.

S2 Fig. Detection of putative linked RNAs harboring *rep* and *ndk* genes in the reverse order.

S3 Fig. Translation-coupled cooperative RNA replication experiment.

S4 Fig. Long-term replication experiment without high temporal dilution.

S5 Fig. Contamination levels of NDK in the translation system.

S1 Table. The list of mutations in the Rep- and NDK-RNA clones obtained at round 46 of the long-term replication experiment shown in Fig 2A.

S2 Table. The list of mutations in the Rep- and RepNDK-RNA clones obtained at round 79 of the long-term replication experiment shown in Fig 3A.

S3 Table. The list of mutations in the NDK- and RepNDK-RNA clones obtained at round 79 of the long-term replication experiment shown in Fig 3A.

S4 Table The list of primers (from 5′ end to 3′end).

Author Contributions

Conceptualization: Kensuke Ueda, Ryo Mizuuchi, Norikazu Ichihashi.

Data curation: Kensuke Ueda.

Formal analysis: Kensuke Ueda, Ryo Mizuuchi.

Funding acquisition: Ryo Mizuuchi, Norikazu Ichihashi.

Investigation: Kensuke Ueda, Ryo Mizuuchi, Norikazu Ichihashi.

Methodology: Kensuke Ueda, Ryo Mizuuchi, Norikazu Ichihashi.

Project Administration: Ryo Mizuuchi, Norikazu Ichihashi.

Supervision: Ryo Mizuuchi, Norikazu Ichihashi.

Validation: Kensuke Ueda, Ryo Mizuuchi, Norikazu Ichihashi.

Visualization: Kensuke Ueda, Ryo Mizuuchi.

Writing – original draft: Kensuke Ueda, Ryo Mizuuchi.

Writing – review & editing: Kensuke Ueda, Ryo Mizuuchi, Norikazu Ichihashi.

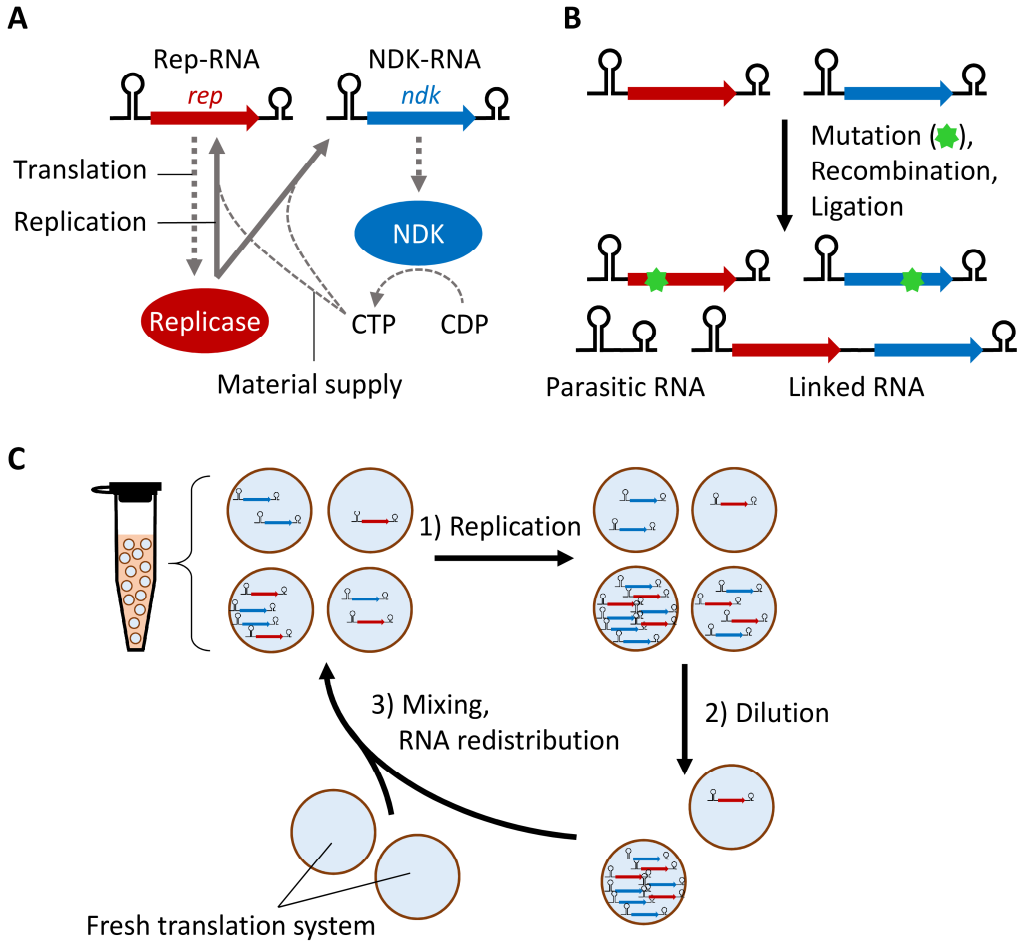
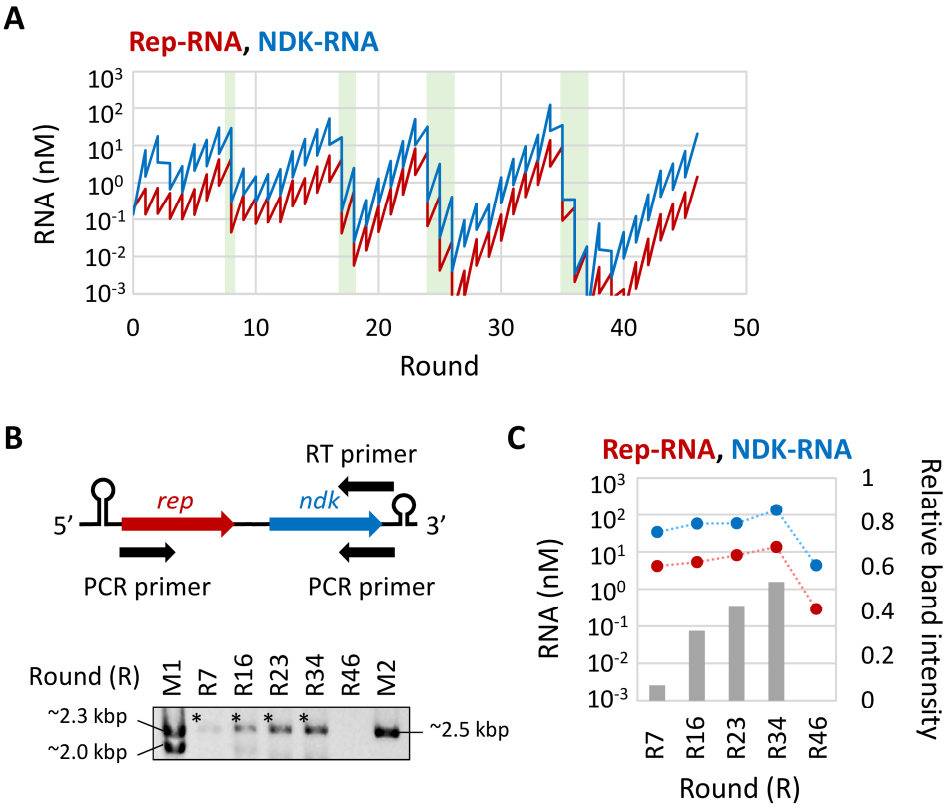


Fig 1. Cooperative RNA replication system. (**A**) Schematic representation of cooperative RNA replication. NDK is translated from NDK-RNA and converts CDP into CTP, a material for RNA replication, whereas replicase is translated from Rep-RNA and replicates both RNAs using the synthesized CTP. (**B**) During replication, mutant Rep- and NDK-RNAs and parasitic RNA that loses gene regions are generated by random mutation and recombination. Similar recombination or ligation could also generate linked RNA that harbors both *rep* and *ndk* genes. (**C**) Schematic representation of long-term replication experiments. 1) Cooperative RNA replication was performed at 37 °C in water-in-oil droplets. 2) Droplets were diluted with new droplets containing a translation system. The dilution rate was typically 5-fold, but 100-fold dilution was used in some rounds. 3) Droplets were vigorously mixed to facilitate their random fusion and division, supplying the translation system to RNA, and RNA was randomly redistributed in the droplets.



578

579 **Fig 2. The first long-term replication experiment.** (**A**) Changes in Rep-RNA (red) and
580 NDK-RNA (blue) concentrations during the long-term replication experiment, measured
581 by RT-qPCR. The experiment was initiated with Rep- and NDK-RNAs "e1R0" and
582 conducted with high (100-fold) temporal dilution. The replication step was performed at
583 37 °C for 6 h. Parasitic RNAs were visualized by native PAGE at rounds 8, 17, 24, and 34,
584 but detected only at round 34 (S1A Fig). The parasitic RNA concentration at the round was
585 33 nM (not plotted). The light green areas highlight the rounds with high dilution. (**B**) RNA
586 samples at the indicated rounds of the long-term replication experiment were subjected to
587 RT-PCR, using primers that could detect 5′-*rep*-*ndk*-3′ (top), and PCR products were
588 analyzed by agarose gel electrophoresis (bottom). M1 and M2 are size markers. Asterisks
589 indicate analyzed bands. (**C**) Relative band intensities of the RT-PCR products to M2 (grey
590 bars, right axis), in comparison with Rep- and NDK-RNA concentrations (red and blue
591 plots, left axis). Dotted lines are plotted for visibility.

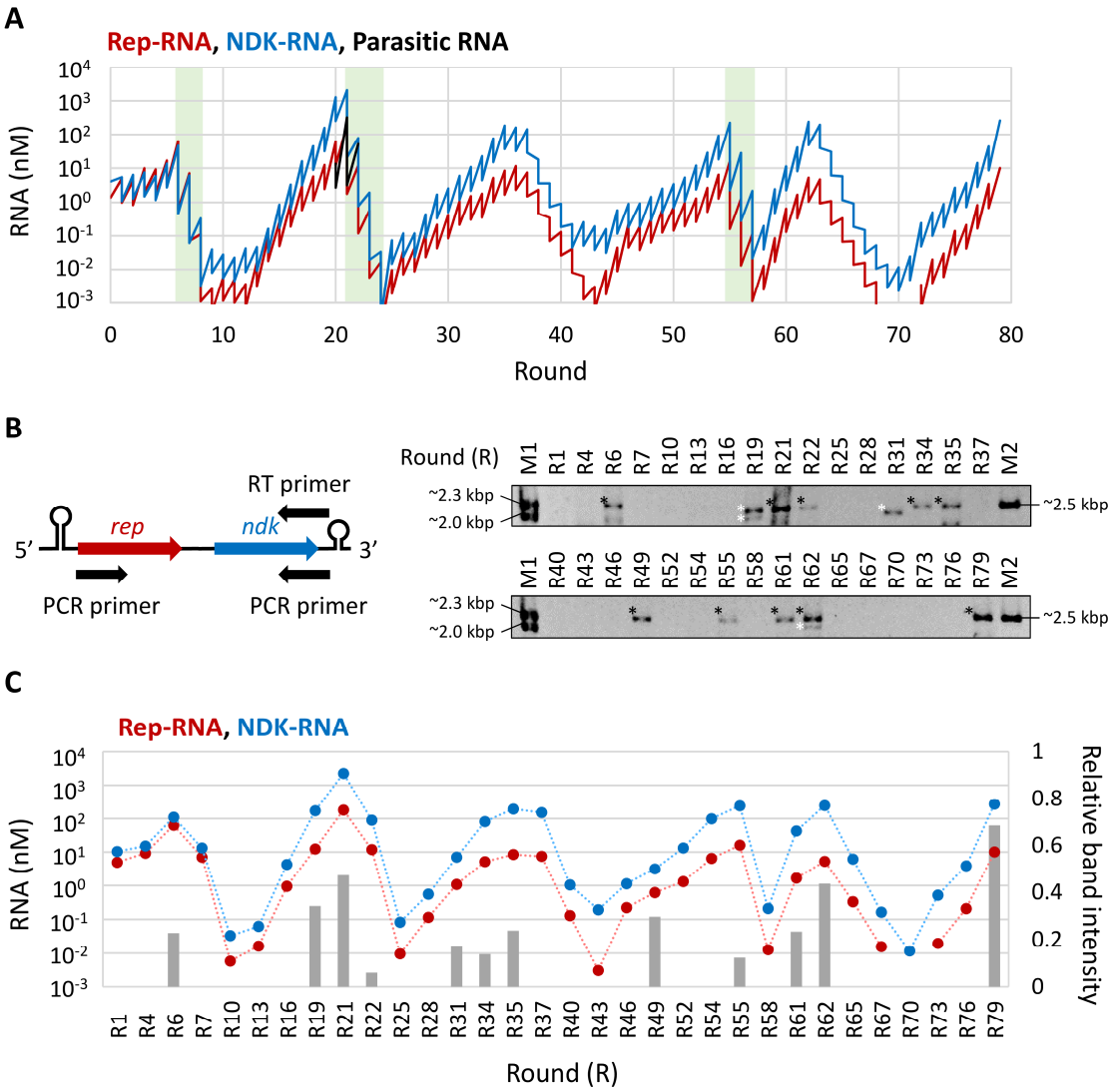


Fig 3. The second long-term replication experiment. (**A**) Changes in Rep-RNA (red), NDK-RNA (blue), and parasitic RNA (black) concentrations during the long-term replication experiment. The experiment was initiated with Rep- and NDK-RNAs "e2R0" and conducted with high (100-fold) temporal dilution. The replication step was performed at 37 °C for 4 h. RNA concentrations were measured by RT-qPCR (Rep- and NDK-RNAs) or based on native PAGE (parasitic RNAs, S1B Fig). Detection of parasitic RNAs was attempted at the rounds described in the main text. Rounds with high dilution are highlighted in light green. (**B**) Detection of putative linked RNAs that harbor both *rep* and *ndk* genes (5′-*rep*-*ndk*-3′) in the RNA samples in every 1–3 rounds of the long-term replication experiment. RT-PCR was performed using the indicated primers (left), and PCR products were analyzed by agarose gel electrophoresis (right). M1 and M2 are size markers.

603 Asterisks indicate analyzed bands; white ones indicate shorter sizes. (**C**) Relative band
604 intensities of RT-PCR products to M2 (grey bars, right axis), in comparison with Rep- and
605 NDK-RNA concentrations (red and blue plots, left axis). Dotted lines are plotted for
606 visibility.
607

A

Rep-RNA region

Base	A825G	T963C	U1349C	A1604G	U1776C
Amino acid	I200V	C246R	V374V	Q459Q	C517R
Rep-RNA	31	81	56	63	25
RepNDK-RNA	100	100	0	0	100

Percentage
100
0

NDK-RNA region

Base	C228U	A240G	T465C	G498A	U582C
Amino acid	5'UTR	E4E	V79V	L90L	G118G
NDK-RNA	6	44	44	44	69
RepNDK-RNA	44	13	38	16	13

Linkage region

sites:	Junction	NDK-RNA	
Base	-C	G3A	U12G
Amino acid	-	5'UTR	5'UTR
RepNDK-RNA	38	75	88

B

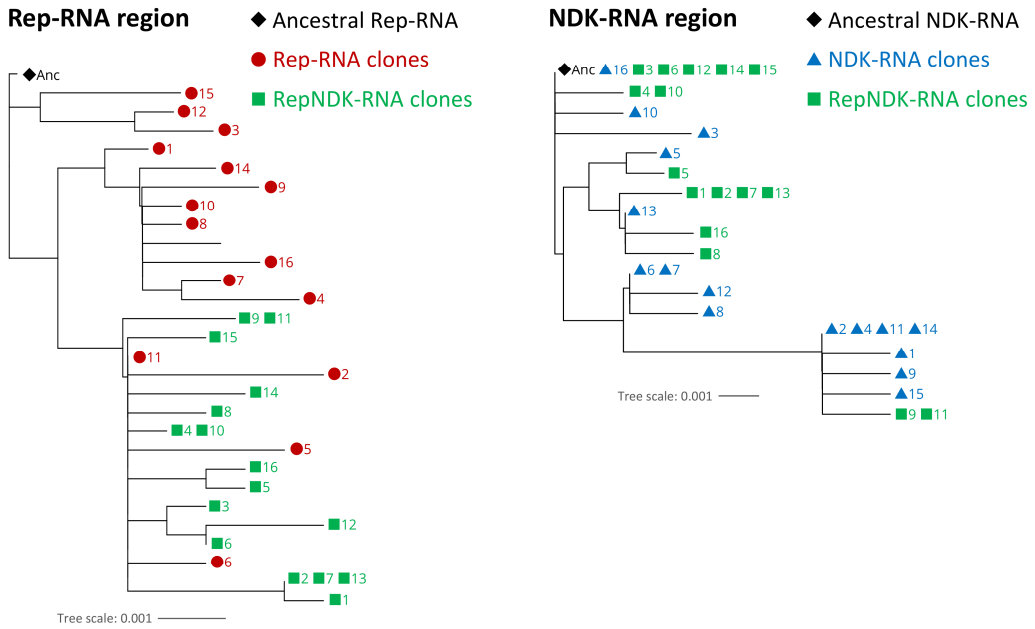


Fig 4. Accumulated mutations in the linked and unlinked RNAs. (**A**) Mutations detected in more than 30% of the 16 clones of any RNA species analyzed at round 79 are listed with their percentages. The sequence of the RepNDK-RNA clones was divided into Rep-RNA, NDK-RNA, and linkage regions for comparison with the Rep- and NDK-RNA clones. All detected mutations are listed in S2 Table and S3 Table. (**B**) Phylogenetic trees for each Rep- (left) and NDK- (right) RNA region were constructed from all the analyzed clones. Leaves representing different RNA species (Rep-, NDK-, or RepNDK-RNA) are marked with different colored symbols, followed by clone numbers. The ancestral Rep- and NDK-RNAs (Anc) were analyzed and displayed together.

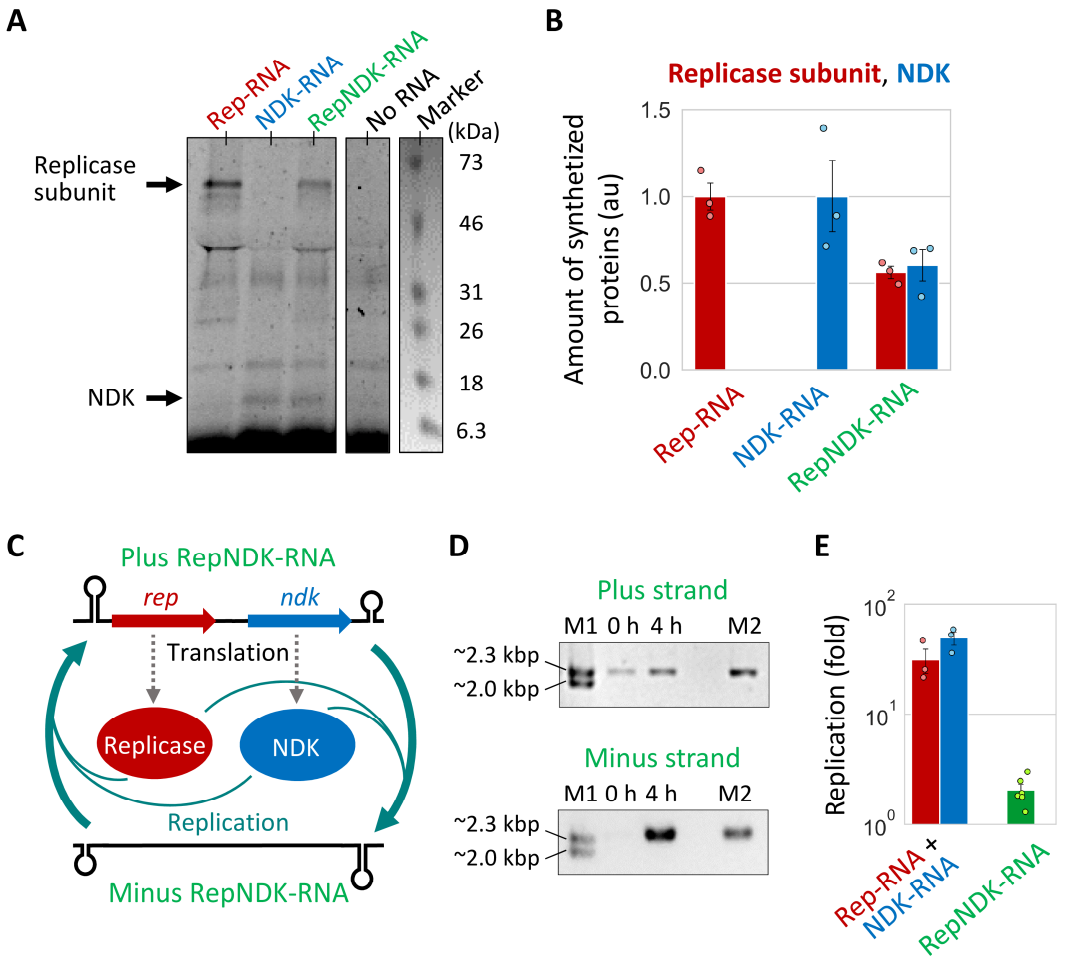


Fig 5. Biochemical properties of the RNA clones at round 79. (**A**) SDS-PAGE of the translation products after incubation of the Rep-, NDK-, and RepNDK-RNA clones (300 nM) at 37 °C for 12 h in a modified translation system containing a fluorescently labeled lysine tRNA. Neither CTP nor CDP was present to prevent RNA replication. An analyzed fluorescent gel image is shown, in parallel to a trimmed white-light image of the same gel to visualize a pre-stained molecular weight marker. The expected bands of the replicase subunit (~64 kDa) and NDK (~15 kDa) are indicated by the black arrows. (**B**) The amounts of synthesized proteins, normalized to those of the Rep- or NDK-RNA clones (replicase subunit or NDK, respectively). Error bars indicate standard errors (n=3). (**C**) The expected replication scheme of RepNDK-RNA. The translation of the two encoded proteins induced RNA replication, completed with minus strand (complementary strand) synthesis from a plus strand (template) and following plus strand synthesis from the minus strand. (**D**) The RepNDK-RNA clone (10 nM) was incubated with the translation system in water-in-oil droplets at 37 °C for 4 h, and the replications of entire plus and minus strands were analyzed

632 by RT-PCR and agarose gel electrophoresis. M1, size marker. M2, control PCR product of
633 the RepNDK-RNA clone. (**E**) The replication amount of the RepNDK-RNA clone was
634 measured by RT-qPCR in comparison with that of the Rep- and NDK-RNA pair (10 nM
635 each). Error bars indicate standard errors (n=3–6).