Evolutionary arms-races between host and parasitic RNA replicators in an artificial cell-like system

Taro Furubayashi [a], Kensuke Ueda [b], Yohsuke Bansho [c], Daisuke Motooka [d], Shota Nakamura [d], Norikazu Ichihashi* [b,c,e,f,g]

[a] Laboratoire Gulliver, CNRS, ESPCI Paris, PSL Research University, Paris, France
[b] Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan
[c] Graduate School of Frontier Bioscience, Osaka University, Osaka, Japan
[d] Research Institute for Microbial Diseases, Osaka University, Osaka, Japan
[e] Komaba Institute for Science, The University of Tokyo, Tokyo, Japan
[f] Department of Life Science, Graduate School of Arts and Science, The University of Tokyo, Tokyo, Japan
[g] Universal Biology Institute, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo, Japan
* Corresponding author : ichihashi@bio.c.tokyo-u.ac.jp

A crucial problem for primitive replicators before the origin of life is the appearance of parasitic or selfish replicators, which destabilize molecular cooperation and prevent the development of complexity. To date, theoretical and experimental studies have indicated that spatial structures, such as cell-like compartments, support sustainable replication of primitive replicators even in the presence of parasites. However, it is still a mystery how these host and parasitic replicators can evolve when they undergo long-term co-replication. Here, we investigated the coevolutionary process between an artificial RNA replicator (host RNA) and spontaneously appearing parasitic replicator (parasitic RNA) in artificial cell-like compartments. We performed a long-term replication experiment and found that the population dynamics of the host and parasitic RNAs gradually changed and novel types of hosts and parasites continued to appear. Competitive replication assays confirmed that parasite-resistant evolution of the host RNA and counter-adaptive evolution of the parasitic RNA occurred one after another. These results demonstrated that evolutionary arms races occur in this simple molecular system and generate a continuously-evolving molecular ecosystem. This coevolutionary process between host and parasitic molecules might play an important role in the open-ended evolution and development of complexity for primitive self-replication systems.

Introduction

Parasitic entities are ubiquitous in the biosphere. Viruses, for example, are the most prosperous biological replicators outnumbering cellular organisms and imposing ceaseless selection pressures on host organisms [1-4]. The competition between parasite infectivity and host resistance, often referred to as “arms race”, can accelerate evolution and even allow the development of a sophisticated biological system, such as adaptive immunity [5-7]. Accumulating evidence highlights the key roles of parasites, including mobile genetic elements, plasmids, and viruses, in diversification of life and the major evolutionary transitions [8-14].

Understanding evolutionary processes between host and parasite is also crucial to understand the origin of life [15-18] and to construct artificial cells capable of evolution [19-21]. The emergence of parasitic replicators are believed to be inevitable and crucial in an evolving replicator system [6] as experimentally demonstrated in an artificial RNA replication system we have reported [22]. Theoretical studies have predicted that spatial structures, such as cell-like compartments, can allow survival of host replicators in the presence of parasites [23-26]. Recent experimental studies using the RNA replication system confirmed this; compartmentalization actually supported the survival of functional RNA replicators in the presence of spontaneously emerging parasitic RNA replicators [27-29]. Although these theoretical and experimental advances are revealing how simple molecular replicators survive under the threats of parasites, their coevolutionary process still remains to be elucidated.

In a previous study, we constructed an RNA replication system consisting of an artificial genomic RNA,

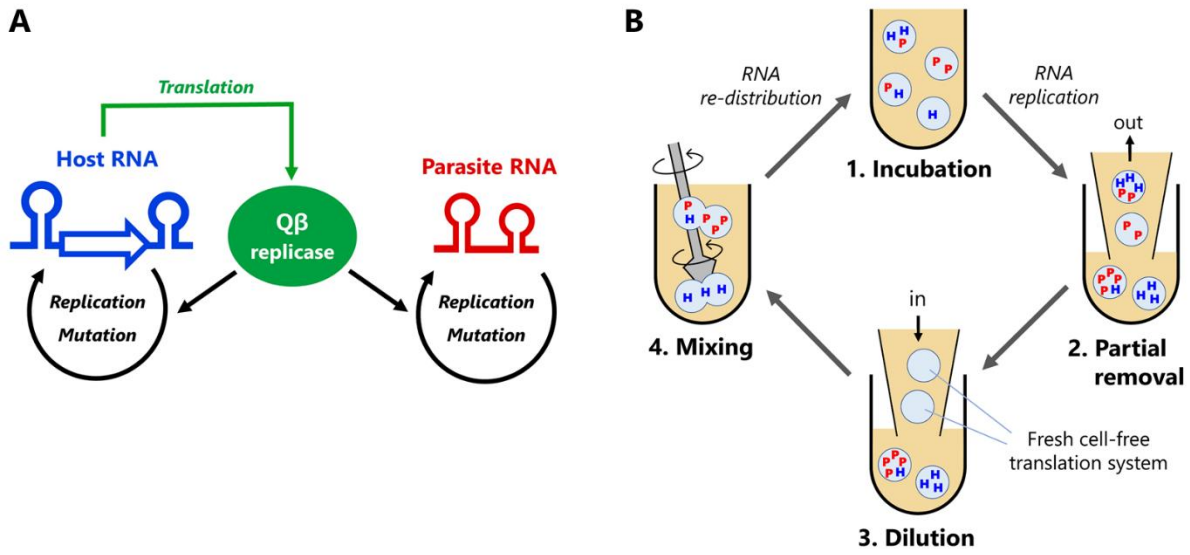


Fig.1 | Host and parasitic RNA replication system. (A) Replication scheme of the host and parasitic RNAs. The host RNA encodes Qβ replicase subunit, while the parasitic RNA does not. Both RNAs are replicated by the translated Qβ replicase in the reconstituted translation system of *E. coli*. (B) Replication-dilution cycle for a long-term replication experiment. The host RNA is encapsulated in water-in-oil droplets with approximately 2 μm diameter. The parasitic RNA spontaneously appears. (1) The droplets are incubated at 37 °C for 5 h for translation and replication. (2) 80 % of droplets are removed and (3) diluted with new droplets containing the translation system (i.e., 5-fold dilution). (4) Diluted droplets are vigorously mixed to induce random fusion and division among the droplets. We repeated this cycle for 120 rounds. Reaction volume was 1 mL with 1% aqueous phase, corresponding to approximately 10^8 droplets.

which encodes the catalytic subunit of an RNA replicase originated from bacteriophage Qβ, and a reconstituted translation system of *Escherichia coli* encapsulated in water-in-oil droplets. In this system, the artificial genomic RNA (host RNA) replicates through the translation of the self-encoded replicase subunit. During replication, a deletion mutant of the host RNA (parasitic RNA), which lost the encoded replicase subunit gene, spontaneously appears and replicates by freeriding the replicase provided from the host RNA. In this replication system, host and parasitic RNAs compete for the host-provided RNA replicase in water-in-oil droplets and undergo repeated error-prone replication and natural selection processes [30]. In another study, we reported that the host and parasitic RNA replicators showed oscillating population dynamics, and a new type of the host RNA acquiring a certain level of parasite-resistance appeared through 43 rounds of replication experiment [27]. However, we have not observed the host-parasite coevolutionary process, e.g. counter-adaptative evolution of the parasitic RNA., in that short period of time.

In this study, to investigate co-evolutionary dynamics of the host and parasitic RNA replicators, we extended the replication experiment from 43 to 120 rounds. Oscillating population dynamics of the host and parasitic RNA replicators changed gradually, and novel types of host and parasitic RNAs continue to appear during the experiment. Competitive replication assays confirmed that parasite-resistant evolution of the host RNA and the counter-adaptive evolution of the parasitic RNAs occurred one after another, i.e. evolutionary arms-races occurred even in this simple molecular replication system.

Results

RNA replication system. The RNA replication system used in this study consists of two types of single-stranded RNAs (host and parasitic RNAs) and a reconstituted translation system of *E. coli* [31] (Fig.1A). A distinctive feature between the host and parasitic RNAs is the capability of providing an RNA replicase (Qβ replicase). The host RNA provides the catalytic β-subunit of the replicase via translation, which forms active

replicase by associating with EF-Tu and EF-Ts subunits in the translation system, while the parasitic RNA lacks the intact gene. The host RNA replicates using the self-provided replicase, whereas the parasitic RNA replicates relying on the host-provided replicase.

In this system, parasitic RNAs spontaneously emerges from the host RNA by deleting the internal replicase gene plausibly through nonhomologous recombination [27, 32]. The parasitic RNAs reported previously have similar sizes (~200 nt). We refer to this size of parasitic RNA species as "parasite-α". The parasite-α replicates much faster than the original host RNA (~2040 nt) due to its smaller size, and thus inhibits the host replication through competition for the replicase. The replication with Qβ replicase is error-prone, about 1.0×10^{-5} per base [33], and mutations are randomly introduced into the host and parasitic RNAs during the replication reaction.

Long-term replication experiment. In the previous study and also in this study, we performed a long-term replication experiment of the host and parasitic RNAs. The replication reaction was performed in water-in-oil droplets through repeating a fusion-division cycle with the supply of new droplets containing the translation system (Fig.1B). The single round of the experiment consists of four steps: 1) incubation, 2) partial removal, 3) dilution, and 4) mixing. 1) In the incubation process, the water-in-oil droplets were incubated at 37 °C for 5 h to induce the internal translation and RNA replication reactions. Note that we started with 1 nM of the host RNA without the parasite-α, but the parasite-α was detected within two rounds in most of the cases. 2) In the partial removal process, we randomly removed 80% of the water-in-oil droplets and 3) substituted with new water-in-oil droplets containing the cell-free translation system in the dilution process (i.e., 5-fold dilution). 4) In the mixing process, droplets were vigorously mixed with a homogenizer to induce random fusion and division among droplets to allow the mixing of RNAs and other components among droplets. We repeated this cycle for 43 rounds (215 h) in the previous study [27] and continued for additional 77 rounds (385 h) in this study.

Population dynamics of the host and parasitic RNAs. We measured the concentrations of the host and parasitic RNAs after every incubation process (Fig.2A). The host RNA was measured using quantitative PCR after reverse transcription (RT-qPCR) and the parasitic RNA was measured from the band intensity after polyacrylamide gel electrophoresis (Supplementary Figure 1). In some rounds (7-12, 16-22, and 75-84), the parasitic RNAs were under the detection limit (less than ~30 nM). Note that, in the previous study [27], we measured the parasitic RNA concentration from the replication kinetics using a purified Qβ replicase and the detection limit was lower, but this method cannot be employed here because it cannot distinguish different types of parasitic RNAs appeared in this study.

The population dynamics of the host and parasitic RNAs gradually changed throughout rounds (Fig.2A). In the early stage (round 1 to ~35), the host RNA and the parasite-α showed relatively regular oscillation pattern caused by competition between the host and parasitic RNAs in compartments as reported previously [27]. In this regime, the concentrations of the parasite-α were higher than those of the host RNA where detected. In the middle stage (round ~35 to ~75), the concentrations of the host RNA increased and the oscillation pattern became irregular, suggesting that the host RNA acquired parasite resistance, partially confirmed in the previous study [27]. In the later stage (round ~95 to ~116), the concentrations of the host and the parasite-α further increased and the oscillation pattern became further unclear. In this regime, we observed the appearance of new parasitic RNA species of different sizes and classified them as parasites-β (~1000 nt, green squares) and -γ (~500 nt, purple diamonds) according to their sizes. We termed these new RNAs "parasites" because each clone of these RNAs does not replicate alone (Supplementary Figure 2). These continuously changing population dynamics can be caused by successive adaptation processes between host and parasitic RNA species.

Sequence analysis. To investigate the sequences of evolving RNA populations, we recovered RNA mixtures from 17 points (rounds 13, 24, 33, 39, 43, 50, 53, 60, 65, 72, 86, 91, 94, 99, 104, 110, and 115) and subjected them to reverse transcription followed by deep sequencing with PacBio RS II for long RNA species (host, parasite-β and -γ) and MiSeq for short RNA species (parasite-α). For the sequencing using PacBio RS II, we

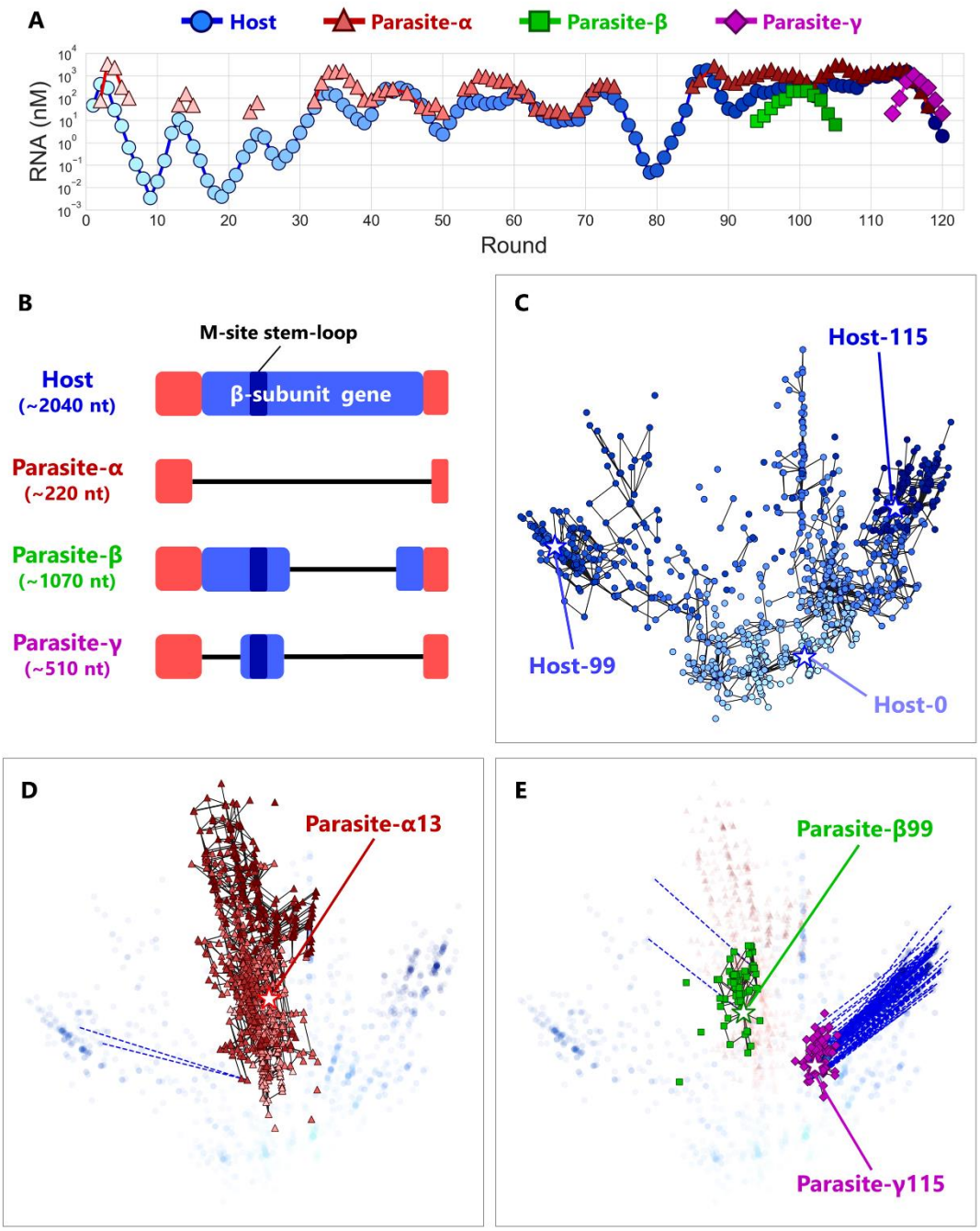


Fig.2 | Coevolutionary dynamics of the host and parasitic RNAs. (A) Population dynamics of the host and parasitic RNAs through the long-term replication experiment. In the regions without points, parasitic RNA concentrations were under the detection limits (< 30 nM). Three different parasitic species (α, β, and γ) are classified based on their sizes. (B) Schematic representation of the sequence alignments of the host and parasitic RNA species. The terminal regions (red) of all the RNA species are derived from a small replicable RNA by the replicase, MDV-1 [50]. β-subunit encoding region are shown in blue and the branched stem-loop of the M-site, one of the binding site for Qβ replicase, is also indicated. Deleted regions are shown in black lines. (C, D, E) 2D maps of the dominant RNA genotypes for the host RNA (C), parasites-α (D), -β and (E) -γ. The top-90 dominant genotypes were plotted for each round. A point represents each genotype. The color depths are consistent with those in (A). Black lines connect pairs of points one Hamming distance apart in the same RNA species. A broken blue line connects a pair of points zero Hamming distance apart (perfect match) in the different RNA species, ignoring the large deletion between host and parasitic RNAs. Stars represent the genotypes of the evolved RNAs used for the competitive replication assay in Fig.3A. The original host RNA is Host-0. Round-by-round data are shown in Supplementary Figure 3.

obtained 365-4143 reads for each RNA species at each round. For the sequencing using MiSeq, we obtained ~5000 reads for each round (Supplementary Table 1).

Sequence analysis revealed that four major RNA species of different sizes existed in the long-term replication experiment, consistent with the band positions observed in the polyacrylamide gels: the host (~2040 nt), the parasite-α (~220 nt), the parasite-β (~1070 nt) and the parasite-γ (~510 nt). The sequences of all the parasitic RNA species shared a high degree of similarity with those of host RNAs but lacked a large part of the replicase subunit gene (Fig.2B). The parasite-α lacks the entire gene. The parasite-β lacks approximately 3' half of the gene and the parasite-γ further lost approximately 1/4 of the remaining 5' region of the gene. Both the parasite-β and -γ retain a part of the M-site sequence, one of the recognition sites for Qβ replicase, in the middle of the gene [34-35].

All four RNA species (host, parasite-α, -β, and -γ) contain various mutations that accumulated during the long-term replication experiment. We next investigated dominant genotypes of each RNA species at each round. Although the RNAs replication by Qβ replicase is error-prone and introduces many random mutations to produce quasi-species for each genotype, we focused on the consensus sequences that consist of mutations commonly found in multiple RNAs. We first identified 74 dominant (i.e., common) mutations that have accumulated more than 10% of the population in any sequenced rounds in any RNA species, including 60 base substitutions, 4 insertions, and 10 deletions in total (Supplementary Table 2). Next, we measured the frequencies of all genotypes composed of the combination of these 74 dominant mutations in each round for each RNA species. All the genotypes and their frequencies are shown in the supplementary data (Genotype frequency.xlsx).

We next investigated the relationships of the detected genotypes. To visualize evolutionary trajectories, we calculated Hamming distances between all combinations of the top-90 genotypes of all the host and parasitic RNA species in each round, and then plotted them in a single two-dimensional (2D) map by using Principle Coordinate Analysis (Fig. 2C-E). Note that we assigned zero distance for the large deletions between host and parasitic RNAs. A point represents each genotype, and the colors of points represent the rounds of their appearance, consistent with the colors of the markers in Fig. 2A. A black line connects a pair of genotypes one Hamming distance apart in the same RNA species. Round-by-round data are shown in Supplementary Figure 3. The host RNA genotypes gradually became distant from the original genotype (Host-0) as rounds proceed (Fig. 2C and Supplementary Figure 3). From round 0 to round 43, sequences diversified around the original genotype. Then, until round 72, most of the genotypes move towards the upper-right branches. But at round 86, where a certain fraction of the genotypes shifted to the left branch and dominated until round 99. At round 104, most of the genotypes moved back to the right branch again and stayed until round 115. These frequent changes of dominant lineages imply that the fittest genotype changes frequently during the long-term replication experiment.

The population of the parasite-α represented a distinct cluster from host RNA populations (Fig.2D) and most of the genotypes are connected with lines (i.e., one Hamming distance apart). The parasite-α did not show clear directionality throughout the long-term replication experiment (round-by-round data is shown in Supplementary Figure 3). We also identified 18 unique mutations specific to the parasite-α (Supplementary Table 2), not found in the corresponding region of the host sequence, indicating that most of the new parasite-α genotypes were not newly generated from evolving host RNAs but the parasite-α underwent independent evolution.

The populations of the parasite-β and -γ formed distinct clusters (Fig. 2E) and most of the genotypes were closely related (connected with one Hamming distant lines) within each species. Sequences of some parasite-β and -γ were perfectly matched with some dominant host RNAs in the same rounds as connected with broken blue lines, suggesting that these parasitic RNAs originated from these host RNAs. This sequence similarity of the parasite-β and -γ to the coexisting host RNAs suggest that newly emerged parasite species possess similar phenotypic traits of evolved host species.

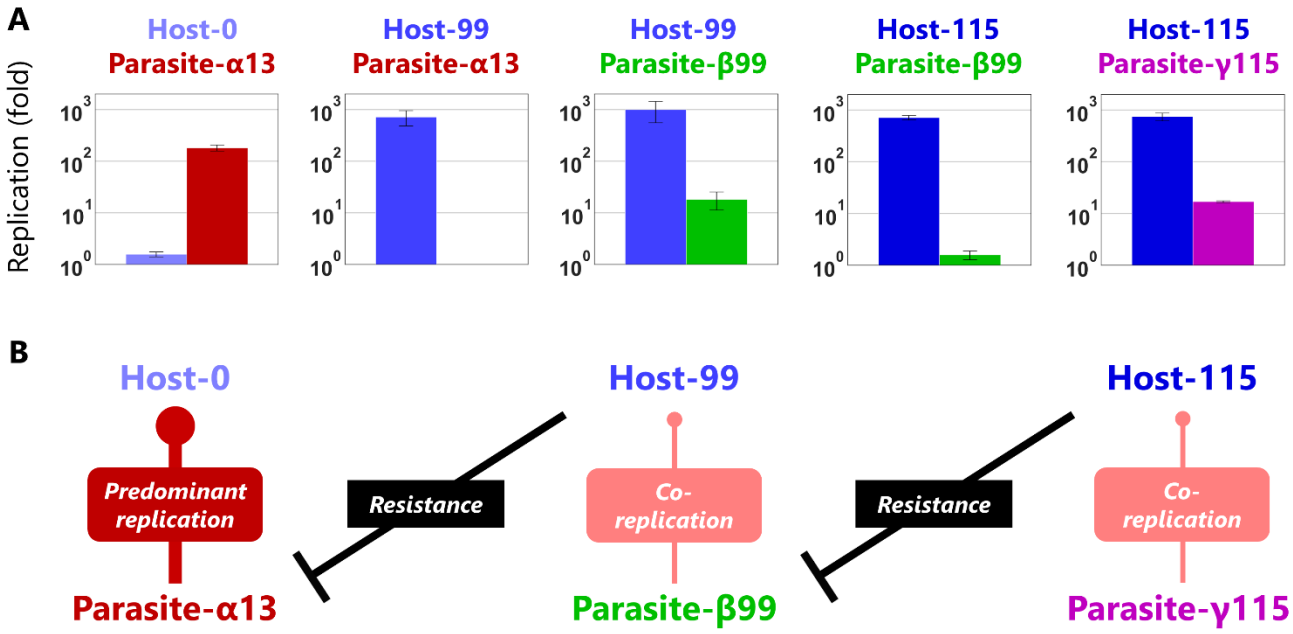


Fig.3 | Evolutionary arms races between host and parasitic RNAs. (A) Competitive replication assay of each pair of the evolved host and parasitic RNA clones. The RNA replication reactions were performed with 10 nM of the host and parasitic RNAs for 3 h and each concentration was measured by sequence-specific RT-qPCR. Error bars represent standard errors of three independent competition assays. (B) Schematic representation of the host-parasite relationships among the RNA clones.

Competitive replication assay of the host and parasitic RNAs. The continuous appearance of new host genotypes and new parasite species can be a consequence of the coevolution between the hosts and the parasites to adapt to each other. To test this possibility, we performed a series of competitive replication assays by using three representative host and parasitic RNAs. As the representatives, we chose the most dominant genotypes at rounds 0, 99, and 115 for host RNA (Host-0, Host-99, and Host-115, respectively). For the parasite-α, -β, and -γ, we chose the most dominant genotype for each at round 13, 99, and 115 (Parasite-α13, Parasite-β99, Parasite-γ115), respectively (sequences are available in Supplementary Text). We mixed a pair of the host and parasitic RNAs according to their order of appearance at an equivalent molar and performed competitive replication. The concentrations of replicated RNAs were measured by sequence-specific RT-qPCR (Fig. 3A). In the first pair (Host-0 vs Parasite-α13), Host-0 hardly replicated (less than 2-fold) and Parasite-α13 predominantly replicated (~200-fold), indicating that Parasite-α13 severely inhibits the original host replication, whereas in the second pair (Host-99 vs Parasite-α13), Host-99, efficiently replicated (~700-fold) with negligible replication of Parasite-α13, indicating that Host-99 acquired the resistance to Parasite-α13. In the third pair (Host-99 vs Parasite-β99), Host-99 still replicated efficiently (~1000-fold) but Parasite-β99, also replicated up to approximately 20-fold, indicating that Parasite-β99 acquired the ability to co-replicate with Host-99. In the fourth pair (Host-115 vs Parasite-β99), Host-115 repressed the replication of Parasite-β99 to less than 2-fold, indicating that Host-115 acquired the ability to evade co-replication of Parasite-β99. In the final pair (Host-115 vs Parasite-γ115), Parasite-γ115 again acquired the ability to replicate up to approximately 20-fold with Host-115. These results demonstrated that successive counter-adaptive evolution (i.e., evolutionary arms-races) occurred among the host and parasitic RNAs as schematically illustrated in Fig. 3B.

Discussion

The emergence of parasitic replicators is a large obstacle for a simple molecular replication system in the early stages of life forms. Here, we investigated the coevolutionary process of simple host and parasitic RNA replicators for the first time and found that the new types of host and parasitic RNAs continuously appeared one after another to produce diversified molecular ecosystem through evolutionary arms-races. This process stands in sharp contrast to the previous unidirectional and rapidly slowing down evolution in the absence of parasitic RNAs [30]. The results of this study implies that the existence of evolving parasites provides dynamically changing selection pressure and allows the host RNAs to continuously evolve and diversify, consistent with the consequence of coevolution between natural host and parasite organisms [8-14] and also with the theoretical study of primitive replicators [36]. This study demonstrated that even a simple molecular replicator has the ability to evolve continuously and diversify through coevolutionary processes with parasitic replicators. Coevolution and evolutionary arms-races with parasitic entities might play a key role in open-ended evolution of primitive self-replicators before the origin of life.

The molecular mechanism for the resistance of the evolved host RNAs to the parasitic RNAs and the following adaptation by parasites (Fig. 3) still remains unclear. One of the possible mechanisms is the change of sequence specificity of the encoded replicase. We previously reported that the replicase encoded in a host RNA at round 43 does not replicate a parasite-α [27] and here we observed that Host-99 and Host-115 with a different set of non-synonymous mutations (Supplementary Table 3) exhibited distinct parasite resistance in the competitive replication assays (Fig.3). The parasite resistance possibly comes from the specificity of the evolved replicases to additional recognition sites only existing in host sequences, which might explain why new parasitic RNAs (parasites-β and -γ) harbor a part of the host sequences. A probable recognition site is a branched stem-loop in the M-site sequence, an important but not indispensable recognition site for Qβ replicase [34-35] and included both in parasite-β and -γ sequences (Fig. 2B). Although further investigation is required, these parasites-β and -γ were able to adapt to the parasite-α-resistant host RNA by acquiring the important part of the M-site from the evolved host sequence.

Coevolution with parasitic entities has been considered as one of the fundamental processes to allow the highly diverse living world. To understand coevolutionary processes in nature, many theoretical and experimental studies have been reported (reviewed in [5, 37-40]). A famous phenomenon in host-parasite coevolution is Red Queen dynamics, in which host and parasite populations oscillate due to persistent replacement of dominant hosts and parasites [41-43]. Similar oscillation dynamics were observed in our experiment (Fig. 2A) with a remarkable feature of damped-oscillation like behavior. A study on *Daphnia* and its parasite [44] also reported damped long-term host-parasite Red Queen coevolutionary dynamics and suggested that the increased host diversity as a consequence of coevolution could decrease fluctuations in host-parasite Red Queen dynamics. Likewise, damped fluctuations and coexistence of the host-parasite population in the later stage of the replication experiment here might be partially attributed to the increased diversity (Fig. 2C-E, Supplementary Figure 3), which allows competition among various types of host and parasitic species to average the population dynamics.

A remaining interesting question is what happens at the end of this evolutionary arms race between the host and parasitic RNAs. Coevolution has been proposed as a driving force for the evolution of complexity [14, 36]. Also, a theoretical study predicted that the mutualism between hosts and parasites can appear during coevolution [45]. The RNA replication system used here provides a useful model system to experimentally demonstrate the possible contribution of host-parasite coevolution to the development of complexity in primitive life-forms.

Methods

A long-term replication experiment. In this study, we additionally performed 77 rounds of replication using the RNA population at round 43 of the previous experiment by the same method [27]. In this method, initially,

10 μL of the reconstituted *E. coli* translation system [31] containing 1 nM of the original host RNA (Host-0, the round 128 clone in [30] was mixed with 1 mL of the buffer-saturated oil prepared as described previously [30] using a homogenizer (POLYTRON PT-1300D; KINEMATICA) at 16,000 rpm for 1 min on ice. The water-in-oil droplet was incubated at 37 °C for 5 h to induce the protein translation reaction and the RNA replication reaction. For the next round of RNA replication, a fraction of the water-in-oil droplets (200 μL) was transferred and mixed with the new buffer-saturated oil (800 mL) and the new translation system (10 μL) using the homogenizer at 16,000 rpm for 1 min on ice, then incubated at 37 °C for 5 h. The average diameter of the water-in-oil droplets was approximately 2 μm [27]. After the incubation step in each round, RNA concentrations were measured as described below. The composition of the translation system is described in the previous study [46].

Measurement of host RNA concentrations. The water-in-oil droplets after the incubation step were diluted 10000-fold with 1 mM EDTA (pH 8.0) and subjected to RT-qPCR (PrimeScript One Step RT-PCR Kit (TaKaRa)) with primer 1 and 2 after heating at 95 °C for 5 min. These primers specifically bind to the host RNA. To draw a standard curve in RT-qPCR, dilution series of the water-in-oil droplets containing the original host RNA diluted 10000-fold with 1 mM EDTA were used.

Measurement of parasitic RNA concentrations. To determine the concentrations of the parasitic RNAs that appeared during the long-term replication experiment (Fig. 2A), polyacrylamide gel electrophoresis was performed followed by quantification of the fluorescent intensities of the parasitic RNA bands using ImageJ. The water phases were collected from the water-in-oil droplets after the incubation step at each round and RNAs were purified with spin columns (RNeasy, QIAGEN). The purified RNA samples and dilution series of the standard parasitic RNA (S222 RNA [47]) were subjected to 8% polyacrylamide gel electrophoresis with 0.1% SDS in TBE buffer (pH 8.4) containing Tris(hydroxymethyl)aminomethane (100 mM), boric acid (90 mM), and EDTA (1 mM), followed by staining with SYBR Green II (Takara). The fluorescent intensities of the parasitic RNA bands were quantified and the concentrations were determined based on the standard curve drawn with the dilution series of the standard parasitic RNA bands.

Sequence analysis. The RNA mixture of round-13, 24, 33, 39, 43, 50, 53, 60, 65, 72, 86, 91, 94, 99, 104, 110, and 115 in the long-term replication experiment were purified with spin columns (RNeasy, QIAGEN). The purified RNAs were reverse-transcribed using PrimeScript reverse transcriptase (Takara) and primer 3 and then PCR-amplified using primer 3 and 4. The PCR products were subjected to agarose gel electrophoresis and the bands corresponding to the host cDNA and the parasitic cDNA were separately extracted by using E-gel CloneWell (Thermo Fisher Scientific). The host and the parasite-β and -γ were sequenced using PacBio RS II with C4/P6 chemistry (Pacific Biosciences), and the parasite-α was sequenced using MiSeq (Illumina). To reduce read errors in the PacBio RS II sequencing, we used circular consensus sequencing (CCS) reads at least comprising five and ten reads for the host and the parasites, respectively. All the sequence reads were subjected to sequence alignment with a reference sequence (the original host sequence) separately for each molecular species (i.e. the host, the parasite-α, -β, and -γ) using MAFFT v7.294b with FFT-NS-2 algorithm [48]. Frequencies of mutations were calculated for each round sample and 74 dominant mutations were identified (Supplementary Table 2). These 74 dominant mutations are located in 72 sites (i.e., a few mutations were introduced in the same sites). In the following analysis, we focused on only the genotypes consist of these 72 sites with 74 mutations.

Mapping dominant genotypes on a two-dimensional space. From the genotypes consisting of the 72 sites with the 74 mutations, the top-90 most dominant genotypes were identified for each host and parasitic species at each round. Hamming distances between all the pairs of the genotypes were calculated and a square distance matrix D, whose i,j-th component d_{ij} represents the square of the Hamming distance between the i-th genotype

and the *j*-th genotype, was constructed. Using principal coordinate analysis on the square distance matrix *D*, the positions of each genotype were determined. The matrix *D* was transformed into a kernel matrix $K = -1/2CDC$, where *C* is the centering matrix. Let λ_k and $e_k \equiv (e_{k1}, e_{k2}, \ldots, e_{kM})$ denote the *k*-th eigenvalue and the *k*-th normalized eigenvector, where $\lambda_1 > \lambda_2 > \ldots > \lambda_M$ and $| e_k | = 1$ for all *k* and *M* is the dimension of the kernel matrix *K*. The eigenvalues and eigenvectors of the kernel matrix *K* were calculated and the *i*-th genotype was plotted in the two-dimensional space with a coordinate described as follows:

$$(X(i), Y(i)) = (\sqrt{\lambda_1} e_{1i}, -\sqrt{\lambda_2} e_{2i})$$

Competitive replication assay of host and parasitic RNAs. The six plasmids, each containing T7 promoter and following cDNA sequences of Host-0, Host-99, Host-115, Parasite-α13, Parasite-β99, and Parasite-γ115, were constructed using the gene synthesis service of Eurofins Genomics. Each RNA were synthesized by *in vitro* transcription with T7 RNA polymerase (TaKaRa) from the plasmids digested with SmaI according to a previous study [49]. We mixed 10 nM of each host and parasitic RNAs in the cell-free translation system and incubated 37 °C for 3 h. The concentrations of the host and the parasitic RNAs were measured by RT-qPCR (PrimeScript One Step RT-PCR Kit (TaKaRa)) with sequence-specific primers (Supplementary text).

References

1. Ø. Bergh, K. Y. Børsheim, G. Bratbak, M. Heldal, High abundance of viruses found in aquatic environments. *Nature* (1989), doi:10.1038/340467a0.
2. F. Rohwer, Global phage diversity. Cell (2003), doi:10.1016/S0092-8674(03)00276-9.
3. C. A. Suttle, Marine viruses - Major players in the global ecosystem. Nat. Rev. Microbiol. (2007), doi:10.1038/nrmicro1750.
4. E. V. Koonin, V. V. Dolja, A virocentric perspective on the evolution of life. *Curr. Opin. Virol.* (2013) , doi:10.1016/j.coviro.2013.06.008.
5. R. Dawkins, J. R. Krebs, Arms races between and within species. *Proc. R. Soc. London - Biol. Sci.* (1979), doi:10.1098/rspb.1979.0081.
6. E. V. Koonin, Y. I. Wolf, M. I. Katsnelson, Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. *Biol. Direct* (2017), doi:10.1186/s13062-017-0202-5.
7. V. Müller, R. J. de Boer, S. Bonhoeffer, E. Szathmáry, An evolutionary perspective on the systems of adaptive immunity. *Biol. Rev.* (2018), doi:10.1111/brv.12355.
8. B. Koskella, M. A. Brockhurst, Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiol. Rev. (2014), doi:10.1111/1574-6976.12072.
9. P. L. Deininger, J. V. Moran, M. A. Batzer, H. H. Kazazian, Mobile elements and mammalian genome evolution. Curr. Opin. Genet. Dev. (2003), doi:10.1016/j.gde.2003.10.013.
10. P. Forterre, The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* (2006), doi:10.1016/j.virusres.2006.01.010.
11. J. Iranzo, A. E. Lobkovsky, Y. I. Wolf, E. V. Koonin, Virus-host arms race at the joint origin of multicellularity and programmed cell death. *Cell Cycle* (2014), doi:10.4161/15384101.2014.949496.
12. R. A. Elbarbary, B. A. Lucas, L. E. Maquat, Retrotransposons as regulators of gene expression. *Science (80-.).* (2016), doi:10.1126/science.aac7247.
13. P. Forterre, D. Prangishvili, The major role of viruses in cellular evolution: Facts and hypotheses. *Curr. Opin. Virol.* (2013), doi:10.1016/j.coviro.2013.06.013.
14. E. V. Koonin, Viruses and mobile elements as drivers of evolutionary transitions. *Philos. Trans. R. Soc. B Biol. Sci.* (2016), doi:10.1098/rstb.2015.0442.
15. W. Gilbert, Origin of life: The RNA world. Nature (1986), doi:10.1038/319618a0.
16. L. E. Orgel, Molecular replication. Nature (1992), doi:10.1038/358203a0.
17. J. Attwater, P. Holliger, Origins of life: The cooperative gene. Nature (2012), doi:10.1038/nature11635.
18. G. F. Joyce, J. W. Szostak, Protocells and RNA self-replication. Cold Spring Harb. Perspect. Biol. (2018), doi:10.1101/cshperspect.a034801.
19. J. W. Szostak, D. P. Bartel, P. L. Luisi, Synthesizing life. *Nature* (2001), doi:10.1038/35053176.
20. A. C. Forster, G. M. Church, Towards synthesis of a minimal cell. Mol. Syst. Biol. (2006), doi:10.1038/msb4100090.
21. P. Stano, P. L. Luisi, Semi-synthetic minimal cells: Origin and recent developments. *Curr. Opin. Biotechnol.* (2013),

doi:10.1016/j.copbio.2013.01.002.
22. Y. Bansho et al., Importance of parasite RNA species repression for prolonged translation-coupled RNA self-replication. Chem. Biol. (2012), doi:10.1016/j.chembiol.2012.01.019.
23. C. Bresch, U. Niesert, D. Harnasch, Hypercycles, parasites and packages. *J. Theor. Biol.* (1980), doi:10.1016/0022-5193(80)90314-8.
24. E. Szathmáry, L. Demeter, Group selection of early replicators and the origin of life. *J. Theor. Biol.* (1987), doi:10.1016/S0022-5193(87)80191-1.
25. N. Takeuchi, P. Hogeweg, Multilevel selection in models of prebiotic evolution II: A direct comparison of compartmentalization and spatial self-organization. *PLoS Comput. Biol.* (2009), doi:10.1371/journal.pcbi.1000542.
26. T. Furubayashi, N. Ichihashi, Sustainability of a Compartmentalized Host-Parasite Replicator System under Periodic Washout-Mixing Cycles. Life (2018), doi:10.3390/life8010003.
27. Y. Bansho, T. Furubayashi, N. Ichihashi, T. Yomo, Host–parasite oscillation dynamics and evolution in a compartmentalized RNA replication system. *Proc. Natl. Acad. Sci.* (2016), doi:10.1073/pnas.1524404113.
28. S. Matsumura *et al.*, Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science (80-.).* (2016), doi:10.1126/science.aag1582.
29. R. Mizuuchi, N. Ichihashi, Sustainable replication and coevolution of cooperative RNAs in an artificial cell-like system. *Nat. Ecol. Evol.* (2018), doi:10.1038/s41559-018-0650-z.
30. N. Ichihashi *et al.*, Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment. *Nat. Commun.* (2013), doi:10.1038/ncomms3494.
31. Y. Shimizu et al., Cell-free translation reconstituted with purified components. Nat. Biotechnol. (2001), doi:10.1038/90802.
32. A. B. Chetverin, H. V. Chetverina, A. A. Demidenko, V. I. Ugarov, Nonhomologous RNA recombination in a cell-free system: Evidence for a transesterification mechanism guided by secondary structure. *Cell* (1997), doi:10.1016/S0092-8674(00)81890-5.
33. L. García-Villada, J. W. Drake, The three faces of riboviral spontaneous mutation: Spectrum, mode of genome replication, and mutation rate. PLoS Genet. (2012), doi:10.1371/journal.pgen.1002832.
34. F. Meyer, H. Weber, C. Weissmann, Interactions of Qβ replicase with Qβ RNA. *J. Mol. Biol.* (1981), doi:10.1016/0022-2836(81)90411-3.
35. D. Schuppli, G. Miranda, S. Qiu, H. Weber, A branched stem-loop structure in the M-site of bacteriophage Qβ RNA is important for template recognition by Qβ replicase holoenzyme. *J. Mol. Biol.* (1998), doi:10.1006/jmbi.1998.2123.
36. N. Takeuchi, P. Hogeweg, Evolution of complexity in RNA-like replicator systems. Biol. Direct (2008), doi:10.1186/1745-6150-3-11.
37. M. E. J. Woolhouse, J. P. Webster, E. Domingo, B. Charlesworth, B. R. Levin, Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat. Genet.* (2002), doi:10.1038/ng1202-569.
38. D. Ebert, Host-parasite coevolution: Insights from the Daphnia-parasite model system. *Curr. Opin. Microbiol.* (2008), doi:10.1016/j.mib.2008.05.012.
39. B. J. M. Bohannan, R. E. Lenski, Linking genetic change to community evolution: Insights from studies of bacteria and bacteriophage. *Ecol. Lett.* (2000), doi:10.1046/j.1461-0248.2000.00161.x.
40. JN Thompson, Specific Hypotheses on the Geographic Mosaic of Coevolution. *The American Naturalist.* (1999), doi: 10.1086/303208
41. L. Van Valen, A new evolutionary theory. *Evol. Theory* (1973), doi:Evolutinary Theory.
42. R. V. Solé, R. Ferrer, I. G. G. J. Quer, E. Domingo, Red queen dynamics, competition and critical points in a model of RNA virus quasispecies. *J. Theor. Biol.* (1999), doi:10.1006/jtbi.1999.0901.
43. J. F. Rabajante *et al.*, Red Queen dynamics in multi-host and multi-parasite interaction system. *Sci. Rep.* (2015), doi:10.1038/srep10004.
44. E. Decaestecker, H. De Gersem, Y. Michalakis, J. A. M. Raeymaekers, Damped long-term host-parasite Red Queen coevolutionary dynamics: A reflection of dilution effects? *Ecol. Lett.* (2013), doi:10.1111/ele.12186.
45. T. Ikegami, K. Kaneko, Computer symbiosis-emergence of symbiotic behavior through evolution. *Phys. D Nonlinear Phenom.* (1990), doi:10.1016/0167-2789(90)90077-3.
46. T. Yoshiyama, T. Ichii, T. Yomo, N. Ichihashi, Automated in vitro evolution of a translation-coupled RNA replication system in a droplet flow reactor. *Sci. Rep.* (2018), doi:10.1038/s41598-018-30374-0.
47. K. Hosoda *et al.*, Kinetic analysis of the entire RNA amplification process by Qβ replicase. *J. Biol. Chem.* (2007), doi:10.1074/jbc.M700307200.
48. K. Katoh, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* (2002), doi:10.1093/nar/gkf436.
49. M. Yumura *et al.*, Combinatorial selection for replicable RNA by Qβ replicase while maintaining encoded gene function. *PLoS One* (2017), doi:10.1371/journal.pone.0174130.
50. D. R. Mills, F. R. Kramer, S. Spiegelman, Complete nucleotide sequence of a replicating RNA molecule. *Science (80-.).* (1973), doi:10.1126/science.180.4089.916.

Acknowledgements

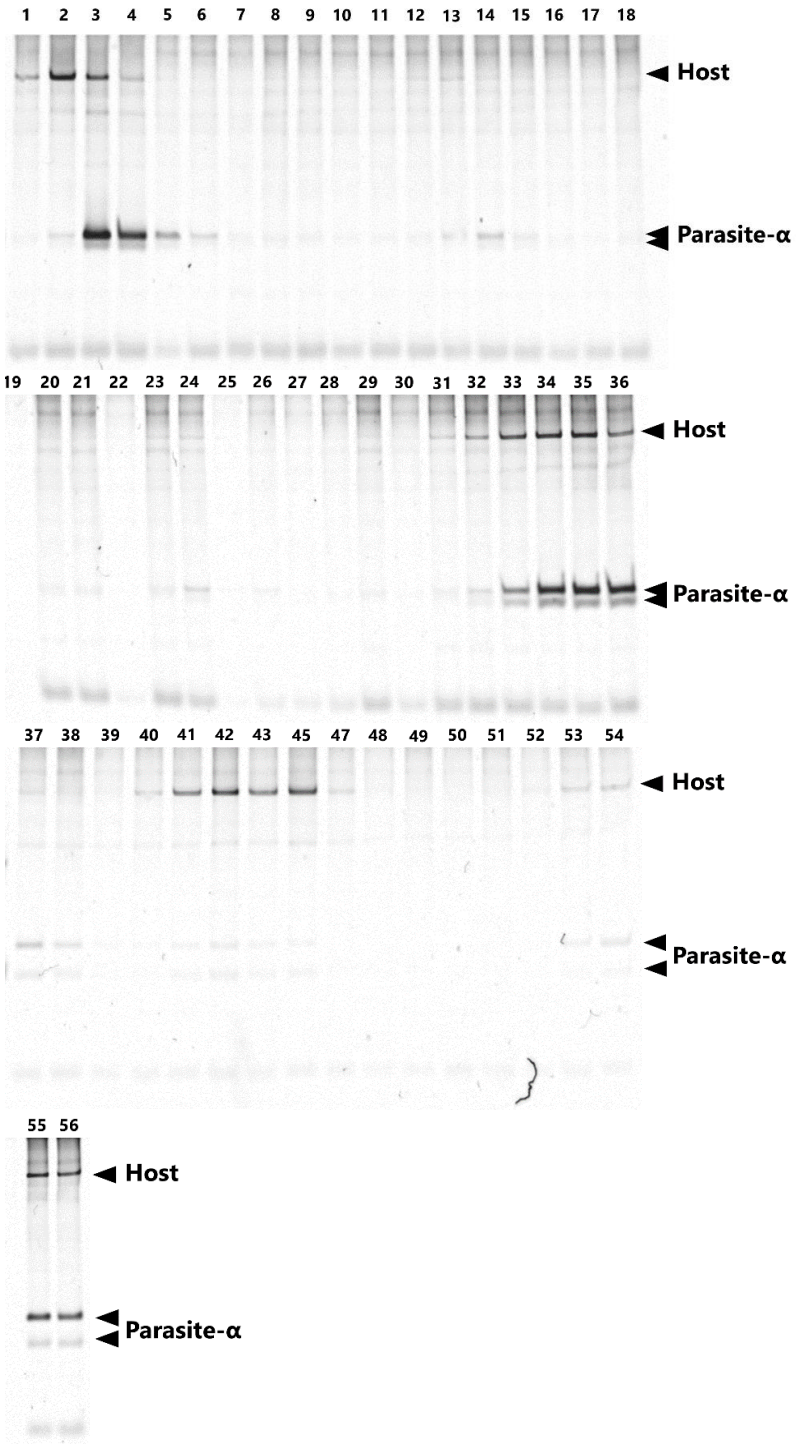
We would like to thank Nobuto Takeuchi, Kunihiko Kaneko, Ryo Mizuuchi, Yoshihiro Sakatani, and Tetsuya Yomo for useful discussions and comments. This work was supported by JSPS KAKENHI grant numbers JP15KT0080, JP15H04407 and JP17J01023.

Author contributions

T.F. and N.I. designed the research. T.F., K.U., Y.B., D.M., S.N. conducted the experiments. T.F. analyzed the data. T.F and N.I. wrote the manuscript.

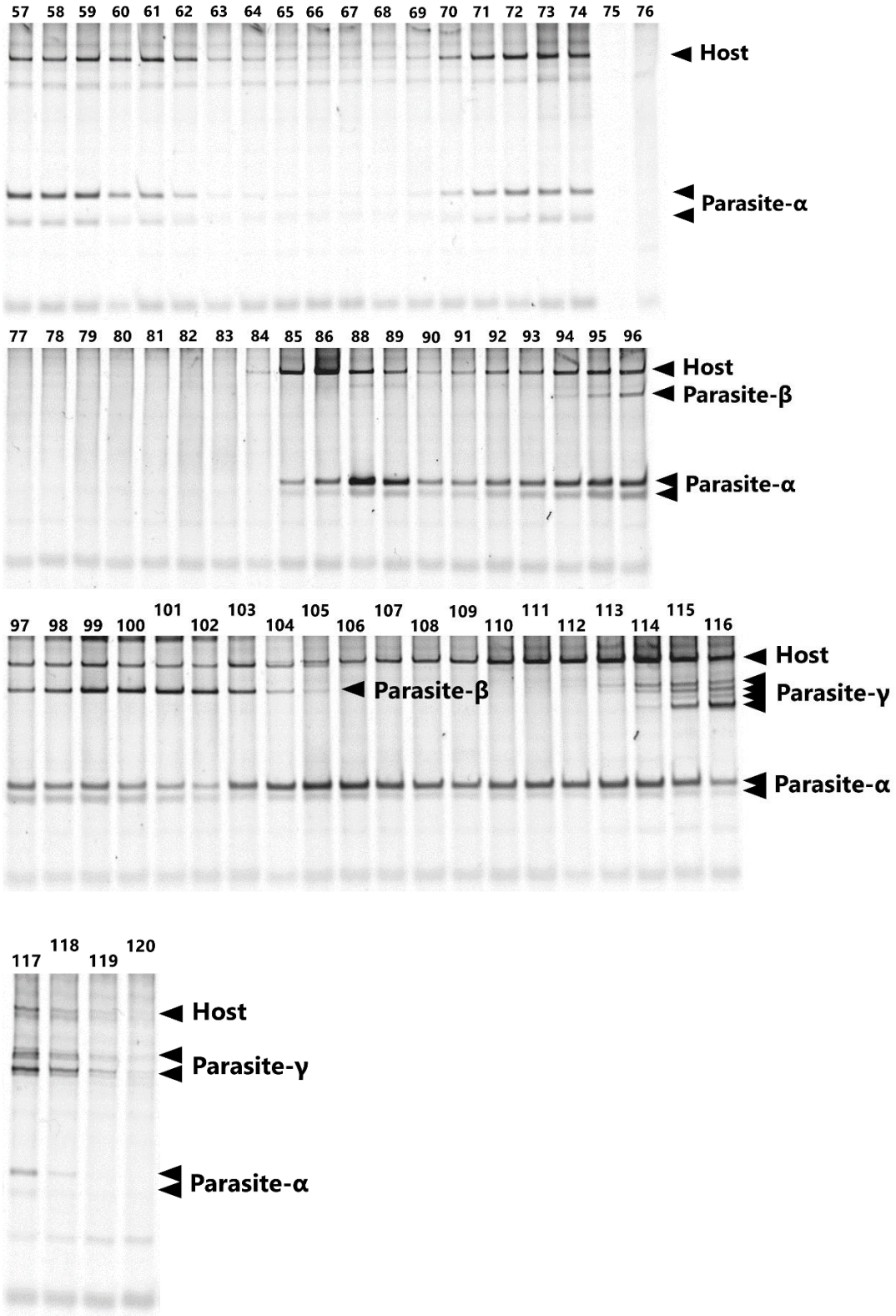
Competing interests

The authors declare no competing interests.

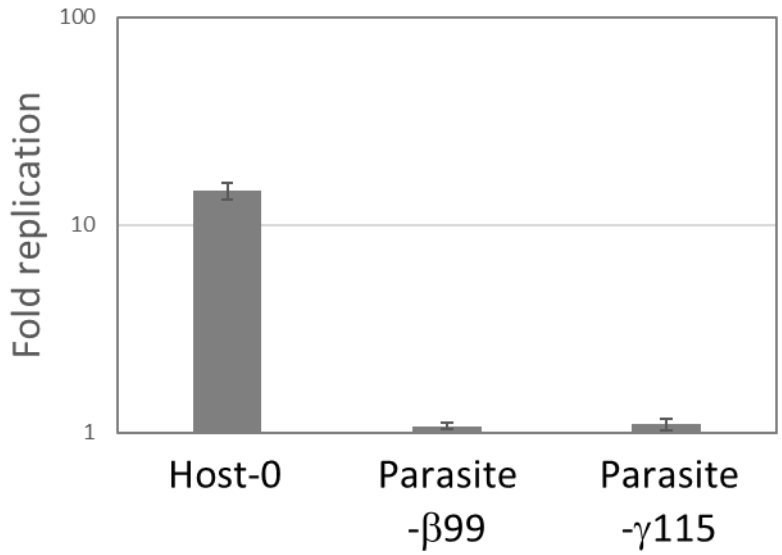


Supplementary Figure 1. The native polyacrylamide gel electrophoresis of the RNA mixture during the long-term replication experiment.

The numbers above the gels indicate the sampled rounds. The parasitic RNAs exhibit multiple bands due to structural heterogeneity.



Supplementary Figure 1 (continued).



Supplementary Figure 2. Replication of Parasite-β99 and Parasite-γ115 without host species.
The RNA replication reactions were performed with 10 nM of each RNA for 5 h, and RNA concentration was measured by sequence-specific RT-qPCR. Error bars represent standard errors of three independent competition assays. Host-0 was also replicated for comparison.

Host

(PacBio RS Ⅱ)

Round	Read number
13	4143
24	605
33	718
39	365
43	484
50	1020
53	1358
60	1409
65	3097
72	1364
86	2135
91	637
94	2058
99	855
104	2535
110	1758
115	4003

Parasite-α

(MiSeq)

Round	Read number
13	5000
24	4990
33	4999
43	4986
53	4987
60	4963
72	4994
86	4974
91	4998
94	4995
99	4984
104	4926
110	4929
115	4959

Parasite-β

(PacBio RS Ⅱ)

Round	Read number
99	4999

Parasite-γ

(PacBio RS Ⅱ)

Round	Read number
115	1753

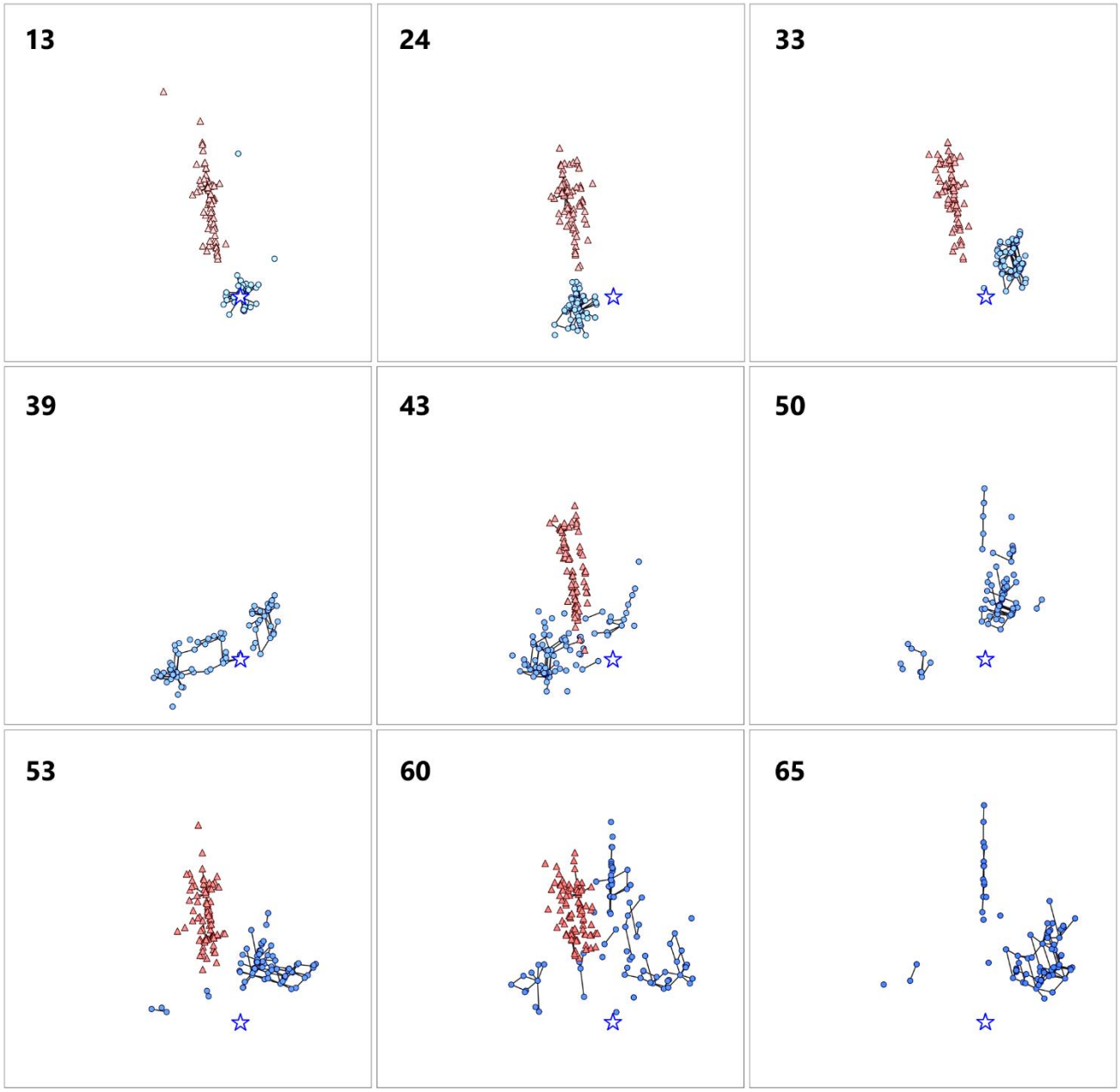
Supplementary Table 1. Read numbers of deep sequencing.
Note that the coverage is 100% for all the reads.

Mutation index	Base change	Amino acid change (host)	Appearance in different species
1	A40G		α
2-1 2-2	G46A G46deletion		α Host
3	A49G		Host, β
4	A53G		Host, β
5	G55deletion		Host, α, β, γ
6	C67U		α
7	C72U		Host, α, γ
8	U78A		Host, α, β
9	C86U		Host, α, β, γ
10	U114C		Host, α, β, γ
11	C115A		α
12	A116G		Host, α, β
13	insertion129A		Host, α, β, γ
14	G141A		Host, α, β, γ
15	C152deletion		α
16	U156C		α
17	C158A		α
18	U159deletion		α
19	C160U C160A		α
20	U161C		Host, β
21	G172U		Host
22	G178C		γ
23	C188U		Host, β
24	U198C		Host
25	U205C		Host
26	A215G		Host
27	A225G		Host, β
28	C232U	Pro2Ser	Host
29	C259U	Leu11Phe	Host, β
30	A294G	synonymous	Host
31	insertion296U	frameshift	Host
32	U321deletion	frameshift	Host
33	C399U	Pro57Leu	Host
34	C414deletion	frameshift	Host, β
35	A438G	Tyr70Cys	Host, β
36	A452G	Ile75Val	Host, β

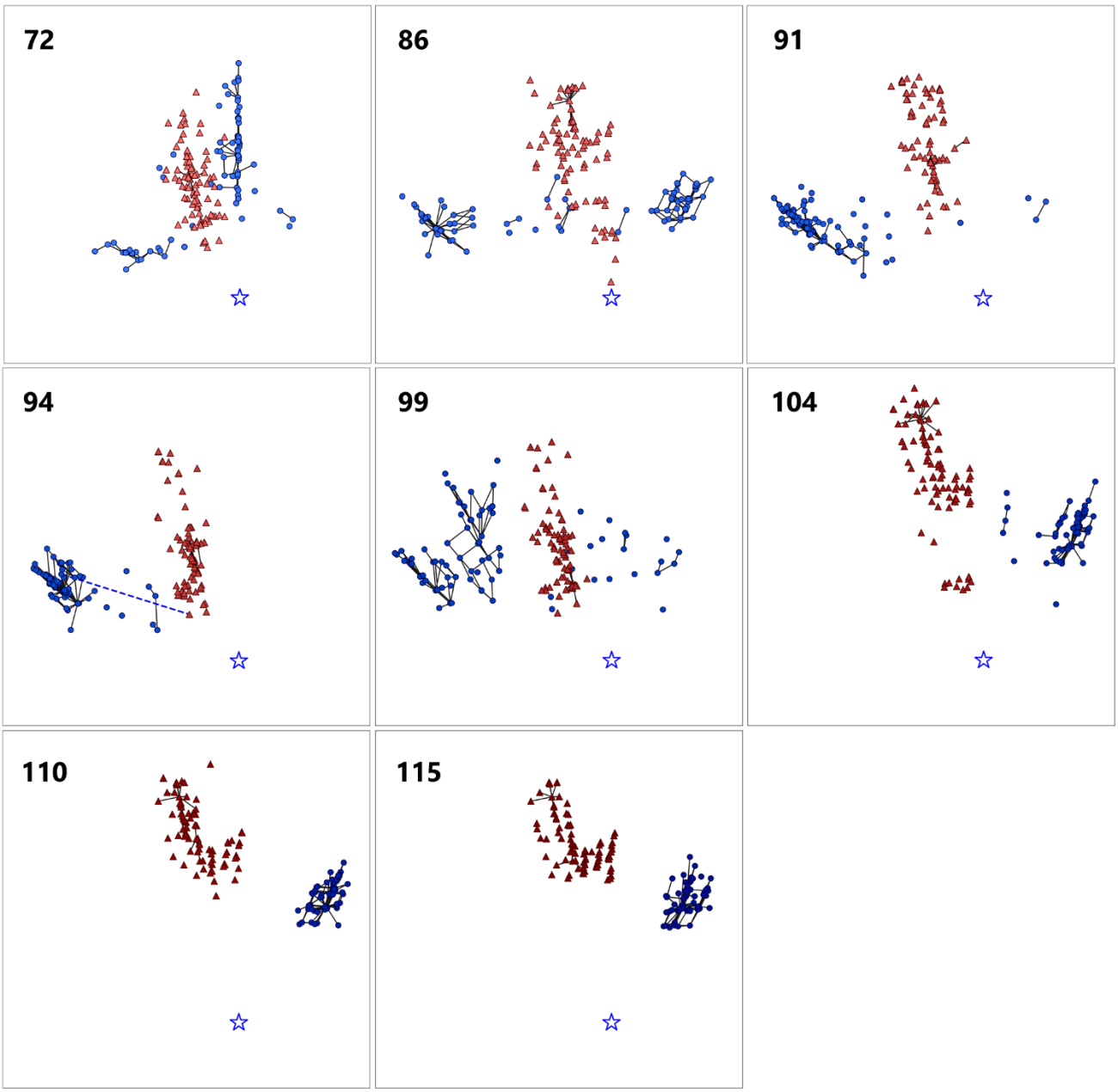
Mutation index	Base change	Amino acid change (host)	Appearance in different species
37	U501C	Val91Ala	Host
38	G565A	synonymous	Host, β, γ
39	U568C	synonymous	Host
40	G599U	Gly124Cys	Host
41-1 41-2	A626G A626C	Arg133Gly synonymous	Host, β Host, γ
42	A631G	synonymous	Host
43	U672C	Met148Thr	Host
44	U683C	Cys152Arg	β
45	U692C	Ser155Pro	Host, γ
46	A706G	synonymous	Host
47	A751G	synonymous	Host
48	G761A	Ala178Thr	β
49	A851G	Lys208Asp	Host
50	U920deletion	frameshift	Host
51	C1103deletion	frameshift	Host
52	U1116C	Phe296Ser	Host
53	U1117C	synonymous	Host
54	U1190C	Ser321Pro	Host
55	C1246A	synonymous	Host
56	U1280C	Ser351Pro	Host
57	G1399deletion	frameshift	Host
58	U1568C	Tyr447His	Host
59	U1572G	Leu448Arg	Host
60	A1605G	Gln459Arg	Host
61	A1608G	His460Arg	Host
62	U1633C	synonymous	Host
63	U1720A	His497Gln	Host
64	C1982A		α
65	A1983G		α
66	insertion1985U		α
67	C1986U		α
68	A1987G		α
69	G1989deletion		α
70	G2001C		α
71	insertion2002U		α
72	U2010C		α

Supplementary Table 2. List of dominant mutations.

The untranslated regions and the coding region are indicated with red and blue backgrounds respectively. The base numbers are determined based on the single reference sequence (shown in the Supplemental text), which contains all dominant insertions (129A, 296U, 1985T, and 2002T) in the original host sequence (Host-0). α, β, and γ in the "Appearance in different species" are shorthand expressions for the parasite-α, -β, and -γ.



Supplementary Figure 3. Round-by-round top-90 host and parasite-α genotypes on the 2D space. A star in each figure represents the coordinate of the original host genotype (Host-0). Black lines connect pairs of points one Hamming distance apart in the same RNA species. A broken blue line connects a pair of points zero Hamming distance apart in the different RNA species, ignoring the large deletion of the parasite-α. Parasite-α is not shown in the round-39, 50, and 65 because they could not have been recovered and sequenced.



Supplementary Figure 3 (continued).

Mutation index	Host-99	Host-115	Amino acid change
3	+	−	5'-UTR
4	+	−	
7	−	+	
8	+	−	
9	+	+	
10	+	+	
12	+	−	
13	−	+	
14	−	+	
27	+	−	
29	−	+	Leu11Phe
37	−	+	Val91Ala
38	−	+	synonymous
40	+	−	Gly124Cys
41	−	+	synonymous
42	+	−	synonymous
47	+	−	synonymous
49	+	+	Lys208Asp
56	+	−	Ser351Pro
58	−	+	Tyr447His
59	+	−	Leu448Arg
60	+	−	Gln459Arg
61	+	−	His460Arg

Supplementary Table 3. Dominant mutations in Host-99 and Host-115.
Mutation indexs correspond to those in Supplementary Table 2.

Supplementary Text

The reference sequence used to determine the position of bases in Supplementary Table 2.

The sequence consists of the original host sequence with four dominant insertion cites (2045 bases in total). The potential insertion positions are represented with asterisks.

GGGAACCCCCCUUCGGGGGGUCACCUCGCGCAGCGGGCUACGCGAGGGAGCCACGCUGCGAAG
CAGCGUGGCGGUUCUCGUGCGUCACCGAAACGCACGAAGGUCGCGCCUCUUCACGAGGCGUC
ACC*UGGGAGAGCGCGAAAGCGCUAGCCCGUGCUCUAGCUCUAGAAGGUCUCGAGAUCUCCUC
UAGAGAUAAUUUUGUUUAACUCUAAGAAGGAGAUAUACACAUGCCUAAGACAGCAUCUUCGC
GUAACUCUCUCAGCGCACAAUUGCGCCGAGCCGCGAACACAAGAA*UUGAGGCUGAAGGUAA
CCUCGCACUUUCCAUUGCCAACGAUUUACUGUUGGCCUAUGGUCAGUCGCCAUUUAACUCUG
AGGCUGAGUGUAUUUCAUUCAGCCCGAGAUUCGACGGGACCCCGGAUGACUUUAGGAUAAAU
UAUCUUAAAGCCGAGAUCAUGUCGAAGUAUGACGACUUCAGCCUAGGUAUUGAUACCGAAGC
UGUUGCCUGGGAGAAGUUCCUGGCAGCAGAGGCUGAAUGUGCUUUAACGAACGCUCGUCUCU
AUAGGCCUGACUACAGUGAGGAUUUCAAUUUCUCACUGGGCGAGUCAUGUAUACACAUGGCU
CGUAGAAAAAUAGCCAAGCUAAUAGGAGAUGUUCCGUCCGUUGAGGAUAUGUUGCGUCACUG
CCGAUUUUCUGGCGGUGCUACAACAACGAAUAACCGUUCGUACAGUCAUCCGUCCUUCAAGU
UUGCACUUCCGCAAGCGUGUACGCCUCGGGCUUUGAAGUAUGUUUUAGCUCUCAGAGCUUCU
ACACAUUUCGAUAUCAGAAUUUCUGAUAUUAGCCCUUUUAAUAAAGCAGUUACCGUACCUAA
GAACAGUAAGACAGAUCGUUGUAUUGCUAUCGAACCUGGUUGGAAUAUGUUUUUCCAACUGG
GUAUCGGUGGCAUUCUACGCGAUCGGUUGCGUUGCUGGGGUAUCGAUCUGAAUGAUCAGACG
AUAAAUCAGCGCCGCGCUCACGAAGGCUCCGUUACUAAUAACUUAGCAACGGUUGAUCUCUC
AGCGGCAAGCGAUUCUAUGUCUCUUGCCCUCUGUGAGCUCUUAUUGCCCCGAGGCUGGUUUG
AGGUUCUUAUGGACCUCAGAUCACCUAAGGGGCGAUUGCCUGACGGUAGUGUUGUUACCUAC
GAGAAGAUUUCUUCUAUGGGUAACGGUUACACAUUCGAGCUCGAGUCGCUUAUUUUUGCUUC
UCUCGCUCGUUCCGUUUGUGAGAUACUGGACUUAGACUCGUCUGAGGUCACUGUUUACGGAG
ACGAUAUUAUUUUACCGUCCCGUGCAGUCCCUGCCCUCCGGGAAGUUUUUAAGUAUGUUGGU
UUUACGACCAAUACUAAAAAGACUUUUUCCGAGGGGCCGUUCAGAGAGUCGUGCGGCAAGCA
CUACUAUUCUGGCGUAGAUGUUACUCCCUUUUACAUACGUCACCGUAUAGUGAGUCCUGCCG
AUUUAAUACUGGUUUUGAAUAACCUAUAUCGGUGGGCCACUAUUGACGGCGUAUGGGAUCCU
AGGGCCCAUUCUGUGUACCUCAAGUAUCGUAAGUUGCUGCCUAAACAGCUGCAACAUAAUAC
UAUACCUGACGGUUACGGUGAUGGUGCCCUCGUCGGAUCGGUCCUAAUCAAUCCUUUCGCGA
AAAACCGCGGGUGGAUCCGGUACGUACCGGUGAUUACGGACCAUACAAGGGACCAAGAGCGC
GCUGAGUUGGGGUCGUAUCUCUACGACCUCUUCUCGCGUUGUCUCUCGGAAAGUAACGAUGG
GUUGCCUCUUAGGGGUCCAUCGGGUUGCGAUUUUGCUGAUCUAUUUGCCAUCGAUCAGCUUA
UCUGUAGGAGUGAUCCUACGAAGAUAAGCAGGCCUACCGGUAAAUUCGAUAUACAGUACAUC
GCGUGCUGUUGUUCGGGUUGUUGUUAGGCUUGCGGCCGCACUCGAGAGAUCUAGAGCAU*CA
CGGUCGAACUCCCG*UACGAGGUGCCCGCACCUCGUCCCCCCCUUCCGGGGGGGUCCCC

Primer list and sequences

Primers used for RT-qPCR for measurement of host RNA concentrations in the long-term replication experiment

Primer 1 : GCTGCCTAAACAGCTGCAAC

Primer 2 : CGCTCTTGGTCCCTTGTATG

Primers for reverse transcription and PCR to recover cDNAs for sequence analysis

Primer 3 : CCGGAAGGGGGGGACGAGG

Primer 4 : GGGTCACCTCGCGCAGC

Primers for RT-qPCR for measurement of RNA concentrations in the competitive replication assay

Host-0 forward : ATACACATGGCTCGTAGAAAA
Host-0 reverse : GGCGTACACGCTTGCGGAAGT
Host-99 forward : TTTGTGAGATACTGGACTTAGACC
Host-99 reverse : GCAGCAACTTACGATACTTGC
Host-115 forward : CCTAGGGCCCATTCTGTGC
Host-115 reverse : GGTTTTTCGCGAAAGGATTGA
Parasite-α13 forward : TACCGAAACGCACGAAGG
Parasite-α13 reverse : CGTACGGGAGTTCGACCG
Parasite-β99 forward : GATAAATTGTCTTAAAGCCGAGG
Parasite-β99 reverse : CTCCTATTAGCTTGGCTATTTTTCC
Parasite-γ115 forward : TAAGAAGGAGATATGCTTTAACGAA
Parasite-γ115 reverse : TAGATCTCTCGAGTCTTGAAGGAC

Sequences of the host and parasitic RNAs used in the competitive replication assay

Host-0 (original host RNA)

GGGAACCCCCCUUCGGGGGGUCACCUCGCGCAGCGGGCUACGCGAGGGAGCCACGCUGCGAAG
CAGCGUGGCGGUUCUCGUGCGUCACCGAAACGCACGAAGGUCGCGCCUCUUCACGAGGCGUC
ACCUGGGAGAGCGCGAAAGCGCUAGCCCGUGCUCUAGCUCUAGAAGGUCUCGAGAUCUCCUC
UAGAGAUAAUUUUGUUUAACUCUAAGAAGGAGAUAUACACAUGCCUAAGACAGCAUCUUCGC
GUAACUCUCUCAGCGCACAAUUGCGCCGAGCCGCGAACACAAGAAUUGAGGCUGAAGGUAAC
CUCGCACUUUCCAUUGCCAACGAUUUACUGUUGGCCUAUGGUCAGUCGCCAUUUAACUCUGA
GGCUGAGUGUAUUUCAUUCAGCCCGAGAUUCGACGGGACCCCGGAUGACUUUAGGAUAAAUU
AUCUUAAAGCCGAGAUCAUGUCGAAGUAUGACGACUUCAGCCUAGGUAUUGAUACCGAAGCU
GUUGCCUGGGAGAAGUUCCUGGCAGCAGAGGCUGAAUGUGCUUUAACGAACGCUCGUCUCUA
UAGGCCUGACUACAGUGAGGAUUUCAAUUUCUCACUGGGCGAGUCAUGUAUACACAUGGCUC
GUAGAAAAAUAGCCAAGCUAAUAGGAGAUGUUCCGUCCGUUGAGGAUAUGUUGCGUCACUGC
CGAUUUUCUGGCGGUGCUACAACAACGAAUAACCGUUCGUACAGUCAUCCGUCCUUCAAGUU
UGCACUUCCGCAAGCGUGUACGCCUCGGGCUUUGAAGUAUGUUUUAGCUCUCAGAGCUUCUA
CACAUUUCGAUAUCAGAAUUUCUGAUAUUAGCCCUUUUAAUAAAGCAGUUACCGUACCUAAG
AACAGUAAGACAGAUCGUUGUAUUGCUAUCGAACCUGGUUGGAAUAUGUUUUUCCAACUGGG
UAUCGGUGGCAUUCUACGCGAUCGGUUGCGUUGCUGGGGUAUCGAUCUGAAUGAUCAGACGA
UAAAUCAGCGCCGCGCUCACGAAGGCUCCGUUACUAAUAACUUAGCAACGGUUGAUCUCUCA
GCGGCAAGCGAUUCUAUGUCUCUUGCCCUCUGUGAGCUCUUAUUGCCCCGAGGCUGGUUUGA
GGUUCUUAUGGACCUCAGAUCACCUAAGGGGCGAUUGCCUGACGGUAGUGUUGUUACCUACG
AGAAGAUUUCUUCUAUGGGUAACGGUUACACAUUCGAGCUCGAGUCGCUUAUUUUUGCUUCU
CUCGCUCGUUCCGUUUGUGAGAUACUGGACUUAGACUCGUCUGAGGUCACUGUUUACGGAGA
CGAUAUUAUUUUACCGUCCCGUGCAGUCCCUGCCCUCCGGGAAGUUUUUAAGUAUGUUGGUU
UUACGACCAAUACUAAAAAGACUUUUUCCGAGGGGCCGUUCAGAGAGUCGUGCGGCAAGCAC
UACUAUUCUGGCGUAGAUGUUACUCCCUUUUACAUACGUCACCGUAUAGUGAGUCCUGCCGA
UUUAAUACUGGUUUUGAAUAACCUAUAUCGGUGGGCCACUAUUGACGGCGUAUGGGAUCCUA
GGGCCCAUUCUGUGUACCUCAAGUAUCGUAAGUUGCUGCCUAAACAGCUGCAACAUAAUACU
AUACCUGACGGUUACGGUGAUGGUGCCCUCGUCGGAUCGGUCCUAAUCAAUCCUUUCGCGAA
AAACCGCGGGUGGAUCCGGUACGUACCGGUGAUUACGGACCAUACAAGGGACCAAGAGCGCG
CUGAGUUGGGGUCGUAUCUCUACGACCUCUUCUCGCGUUGUCUCUCGGAAAGUAACGAUGGG
UUGCCUCUUAGGGGUCCAUCGGGUUGCGAUUUUGCUGAUCUAUUUGCCAUCGAUCAGCUUAU
CUGUAGGAGUGAUCCUACGAAGAUAAGCAGGCCUACCGGUAAAUUCGAUAUACAGUACAUCG

CGUGCUGUUGUUCGGGUUGUUGUUAGGCUUGCGGCCGCACUCGAGAGAUCUAGAGCAUCACG
GUCGAACUCCCGUACGAGGUGCCCGCACCUCGUCCCCCCCUUCCGGGGGGGUCCCC

Host-99

GGGAACCCCCCUUCGGGGGGUCACCUCGCGCAGCGGGCUACGCGAGGGGGCCGCGCUGCGAAG
CAGCGUGGCGGUUCACGUGCGUUACCGAAACGCACGAAGGUCGCGCCUCUCCGCGAGGCGUC
ACCUGGGAGAGCGCGAAAGCGCUAGCCCGUGCUCUAGCUCUAGAAGGUCUCGAGAUCUCCUC
UAGAGAUAAUUUUGUUUAACUCUAAGAAGGAGAUAUGCACAUGCCUAAGACAGCAUCUUCGC
GUAACUCUCUCAGCGCACAAUUGCGCCGAGCCGCGAACACAAGAAUUGAGGCUGAAGGUAAC
CUCGCACUUUCCAUUGCCAACGAUUUACUGUUGGCCUAUGGUCAGUCGCCAUUUAACUCUGA
GGCUGAGUGUAUUUCAUUCAGCCCGAGAUUCGACGGGACCCCGGAUGACUUUAGGAUAAAUU
AUCUUAAAGCCGAGAUCAUGUCGAAGUAUGACGACUUCAGCCUAGGUAUUGAUACCGAAGCU
GUUGCCUGGGAGAAGUUCCUGGCAGCAGAGGCUGAAUGUGCUUUAACGAACGCUCGUCUCUA
UAGGCCUGACUACAGUGAGGAUUUCAAUUUCUCACUGUGCGAGUCAUGUAUACACAUGGCUC
GUAGAAAGAUAGCCAAGCUAAUAGGAGAUGUUCCGUCCGUUGAGGAUAUGUUGCGUCACUGC
CGAUUUUCUGGCGGUGCUACAACAACGAAUAACCGUUCGUACAGUCAUCCGUCCUUCAAGUU
UGCGCUUCCGCAAGCGUGUACGCCUCGGGCUUUGAAGUAUGUUUUAGCUCUCAGAGCUUCUA
CACAUUUCGAUAUCAGAAUUUCUGAUAUUAGCCCUUUUAAUGAAGCAGUUACCGUACCUAAG
AACAGUAAGACAGAUCGUUGUAUUGCUAUCGAACCUGGUUGGAAUAUGUUUUUCCAACUGGG
UAUCGGUGGCAUUCUACGCGAUCGGUUGCGUUGCUGGGGUAUCGAUCUGAAUGAUCAGACGA
UAAAUCAGCGCCGCGCUCACGAAGGCUCCGUUACUAAUAACUUAGCAACGGUUGAUCUCUCA
GCGGCAAGCGAUUCUAUGUCUCUUGCCCUCUGUGAGCUCUUAUUGCCCCGAGGCUGGUUUGA
GGUUCUUAUGGACCUCAGAUCACCUAAGGGGCGAUUGCCUGACGGUAGUGUUGUUACCUACG
AGAAGAUUUCUUCUAUGGGUAACGGUUACACAUUCGAGCUCGAGUCGCUUAUUUUUGCUUCU
CUCGCUCGUUCCGUUUGUGAGAUACUGGACUUAGACCCGUCUGAGGUCACUGUUUACGGAGA
CGAUAUUAUUUUACCGUCCCGUGCAGUCCCUGCCCUCCGGGAAGUUUUUAAGUAUGUUGGUU
UUACGACCAAUACUAAAAAGACUUUUUCCGAGGGGCCGUUCAGAGAGUCGUGCGGCAAGCAC
UACUAUUCUGGCGUAGAUGUUACUCCCUUUUACAUACGUCACCGUAUAGUGAGUCCUGCCGA
UUUAAUACUGGUUUUGAAUAACCUAUAUCGGUGGGCCACUAUUGACGGCGUAUGGGAUCCUA
GGGCCCAUUCUGUGUACCGCAAGUAUCGUAAGUUGCUGCCUAAACAGCUGCGACGUAAUACU
AUACCUGACGGUUACGGUGAUGGUGCCCUCGUCGGAUCGGUCCUAAUCAAUCCUUUCGCGAA
AAACCGCGGGUGGAUCCGGUACGUACCGGUGAUUACGGACCAUACAAGGGACCAAGAGCGCG
CUGAGUUGGGGUCGUAUCUCUACGACCUCUUCUCGCGUUGUCUCUCGGAAAGUAACGAUGGG
UUGCCUCUUAGGGGUCCAUCGGGUUGCGAUUUUGCUGAUCUAUUUGCCAUCGAUCAGCUUAU
CUGUAGGAGUGAUCCUACGAAGAUAAGCAGGCCUACCGGUAAAUUCGAUAUACAGUACAUCG
CGUGCUGUUGUUCGGGUUGUUGUUAGGCUUGCGGCCGCACUCGAGAGAUCUAGAGCAUCACG
GUCGAACUCCCGUACGAGGUGCCCGCACCUCGUCCCCCCCUUCCGGGGGGGUCCCC

Host-115

GGGAACCCCCCUUCGGGGGGUCACCUCGCGCAGCGGGCUACGCGAGGGAGCCACGCUGCGAAG
CAGCGUGGUGGUUCUCGUGCGUUACCGAAACGCACGAAGGUCGCGCCUCUCCACGAGGCGUC
ACCAUGGGAGAGCGCAAAAGCGCUAGCCCGUGCUCUAGCUCUAGAAGGUCUCGAGAUCUCCU
CUAGAGAUAAUUUUGUUUAACUCUAAGAAGGAGAUAUACACAUGCCUAAGACAGCAUCUUCG
CGUAACUCUUUCAGCGCACAAUUGCGCCGAGCCGCGAACACAAGAAUUGAGGCUGAAGGUAA
CCUCGCACUUUCCAUUGCCAACGAUUUACUGUUGGCCUAUGGUCAGUCGCCAUUUAACUCUG
AGGCUGAGUGUAUUUCAUUCAGCCCGAGAUUCGACGGGACCCCGGAUGACUUUAGGAUAAAU
UAUCUUAAAGCCGAGAUCAUGUCGAAGUAUGACGACUUCAGCCUAGGUAUUGAUACCGAAGC
UGCUGCCUGGGAGAAGUUCCUGGCAGCAGAGGCUGAAUGUGCUUUAACGAACGCUCGUCUCU

AUAGACCUGACUACAGUGAGGAUUUCAAUUUCUCACUGGGCGAGUCAUGUAUACACAUGGCU
CGUCGAAAAAUAGCCAAGCUAAUAGGAGAUGUUCCGUCCGUUGAGGAUAUGUUGCGUCACUG
CCGAUUUUCUGGCGGUGCUACAACAACGAAUAACCGUUCGUACAGUCAUCCGUCCUUCAAGU
UUGCACUUCCGCAAGCGUGUACGCCUCGGGCUUUGAAGUAUGUUUUAGCUCUCAGAGCUUCU
ACACAUUUCGAUAUCAGAAUUUCUGAUAUUAGCCCUUUUAAUGAAGCAGUUACCGUACCUAA
GAACAGUAAGACAGAUCGUUGUAUUGCUAUCGAACCUGGUUGGAAUAUGUUUUUCCAACUGG
GUAUCGGUGGCAUUCUACGCGAUCGGUUGCGUUGCUGGGGUAUCGAUCUGAAUGAUCAGACG
AUAAAUCAGCGCCGCGCUCACGAAGGCUCCGUUACUAAUAACUUAGCAACGGUUGAUCUCUC
AGCGGCAAGCGAUUCUAUGUCUCUUGCCCUCUGUGAGCUCUUAUUGCCCCGAGGCUGGUUUG
AGGUUCUUAUGGACCUCAGAUCACCUAAGGGGCGAUUGCCUGACGGUAGUGUUGUUACCUAC
GAGAAGAUUUCUUCUAUGGGUAACGGUUACACAUUCGAGCUCGAGUCGCUUAUUUUUGCUUC
UCUCGCUCGUUCCGUUUGUGAGAUACUGGACUUAGACUCGUCUGAGGUCACUGUUUACGGAG
ACGAUAUUAUUUUACCGUCCCGUGCAGUCCCUGCCCUCCGGGAAGUUUUUAAGUAUGUUGGU
UUUACGACCAAUACUAAAAAGACUUUUUCCGAGGGGCCGUUCAGAGAGUCGUGCGGCAAGCA
CUACUAUUCUGGCGUAGAUGUUACUCCCUUUUACAUACGUCACCGUAUAGUGAGUCCUGCCG
AUUUAAUACUGGUUUUGAAUAACCUAUAUCGGUGGGCCACUAUUGACGGCGUAUGGGAUCCU
AGGGCCCAUUCUGUGCACCUCAAGUAUCGUAAGUUGCUGCCUAAACAGCUGCAACAUAAUAC
UAUACCUGACGGUUACGGUGAUGGUGCCCUCGUCGGAUCGGUCCUAAUCAAUCCUUUCGCGA
AAAACCGCGGGUGGAUCCGGUACGUACCGGUGAUUACGGACCAUACAAGGGACCAAGAGCGC
GCUGAGUUGGGGUCGUAUCUCUACGACCUCUUCUCGCGUUGUCUCUCGGAAAGUAACGAUGG
GUUGCCUCUUAGGGGUCCAUCGGGUUGCGAUUUUGCUGAUCUAUUUGCCAUCGAUCAGCUUA
UCUGUAGGAGUGAUCCUACGAAGAUAAGCAGGCCUACCGGUAAAUUCGAUAUACAGUACAUC
GCGUGCUGUUGUUCGGGUUGUUGUUAGGCUUGCGGCCGCACUCGAGAGAUCUAGAGCAUCAC
GGUCGAACUCCCGUACGAGGUGCCCGCACCUCGUCCCCCCCUUCCGGGGGGGUCCCC

Parasite-α13

GGGAACCCCCCUUCGGGGGGUCACCUCGCGCAGCGGGCUGCGCGAAGGAGCCACGCUGCGAAG
CAGUGUGGCGGUUCUCGUGCGUUACCGAAACGCACGAAGGUCGCGCCUCUUCACGAGGCGUC
ACCUGGGAGAGCGCGAAAGCGCUAGCCCGUGAUUCGUCACGGUCGAACUCCCGUACGAGGUG
CCCGCACCUCGUCCCCCCCUUCCGGGGGGGUCCCC

Parasite-β99

GGGAACCCCCCUUCGGGGGGUCACCUCGCGCAGCGGGCUACGCGAGGGAGCCACGCUGCGAAG
CAGCGUGGCGGUUCUCGUGCGUUACCGAAACGCACGAAGGUCGCGCCUCUCCGCGAGGCGUC
ACCUGGGAGAGCGCGAAAGCGCUAGCCCGUGCUCUAGCUCUAGAAGGUCUCGAGAUCUCCUU
UAGAGAUAAUUUUGUUUAACUCUAAGAAGGAGAUAUGCACAUGCCUAAGACAGCAUCUUCGC
GUAACUCUUUCAGCGCACAAUUGCGCCGAGCCGCGAACACAAGAAUUGAGGCUGAAGGUAAC
CUCGCACUUUCCAUUGCCAACGAUUUACUGUUGGCCUAUGGUCAGUCGCCAUUUAACUCUGA
GGCUGAGUGUAUUUCAUUCAGCCCGAGAUUCGACGGGACCCCGGAUGACUUUAGGAUAAAUU
GUCUUAAAGCCGAGGUCAUGUCGAAGUAUGACGACUUCAGCCUAGGUAUUGAUACCGAAGCU
GUUGCCUGGGAGAAGUUCCUGGCAGCAGAGGCUGAAUGUGCUUUAACGAACGCUCGUCUCUA
UAGGCCUGACUACAGUGAGGAUUUCAAUUUCUCACUGGGCGAGUCAUGUAUACACAUGGCUC
GUGGAAAAAUAGCCAAGCUAAUAGGAGAUGUUCCGUCCGUUGAGGAUAUGUUGCGUCACUGC
CGAUUUUCUGGCGGUGCUACAACAACGAAUAACCGUUCGUACAGUCAUCCGUCCUUCAAGUU
UGCACUUCCGCAAACGUGUACGCCUCGGGCUUUGAAGUAUGUUUUAGCUCUCAGAGCUUCUA
CACAUUUCGAUAUCAGAAUUUCUGAUAUUAGCCCUUUUAAUGGGUUGCGAUUUUGCUGAUCU
AUUUGCCAUCGAUCAGCUUAUCUGUAGGAGUGAUCCUACGAAGAUAAGCAGGCCUACCGGUA
AAUUCGAUAUACAGUACAUCGCGUGCUGUUGUUCGGGUUGUUGUUAGGCUUGCGGCCGCACU

CGAGAGAUCUAGAGCAUCACGGUCGAACUCCCGUACGAGGUGCCCGCACCUCGUCCCCCCCUU
CCGGGGGGGUCCCC

Parasite-γ115

GGGAACCCCCCUUCGGGGGGUCACCUCGCGCAGCGGGCUACGCGAGGGAGCCACCUGCGAAGC
AGCGUGGCGGUUCUCGUGCGUUACCGAAACGCACGAAGGUCGCGCCUCUCCACGAGGCGUCA
CCAUGGGAGAGCGCAAAAGCGCUAGCCCGUGCUCUAGCUCUAGAAGGUCUCGAGAUCUCCUC
UAGAGAUAAUUUUGUUUAACUCUAAGAAGGAGAUAUGCUUUAACGAACGCUCGUCUCUAUAG
ACCUGACUACAGUGAGGAUUUCAAUUUCUCACUGGGCGAGUCAUGUAUACACAUGGCUCGUC
GAAAAAUAGCCAAGCUAAUAGGAGAUGUUCCGUCCGUUGAGGAUAUGUUGCGUCACUGCCGA
UUUCCUGGCGGUGCUACAACAACGAAUAACCGUUCGUACAGUCAUCCGUCCUUCAAGACUCG
AGAGAUCUAGAGCAUCACGGUCGAACUCCCGUACGAGGUGCCCGCACCUCGUCCCCCCCUUCC
GGGGGGGUCCCC