Supplementary Information for

Evolutionary transition from a single RNA replicator to a multiple replicator network

Ryo Mizuuchi*, Taro Furubayashi, Norikazu Ichihashi*

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

This file includes:

Supplementary Text

1. Two additional long-term replication experiments

We performed two additional long-term replication experiments (E2 and E3). Initiated with the droplet mixture at round 76 of the main experiment (Fig. 1c), we performed independent 164 rounds (total 240 rounds) of serial transfer in each experiment. Host RNA concentrations showed similar yet distinct oscillation patterns (Supplementary Figs. S5a and b). The size of dominant parasitic RNAs for the additional experiments was similar to that for the main long-term replication experiment. We then performed PacBio sequencing for host and ~500 nt parasitic RNAs (if detected) at 12 points between rounds 92 and 239 of E2 and 9 points between rounds 114 and 239 of E3. From 4270–10000 reads of the host and parasitic RNAs (Supplementary Table S1), we identified 79 and 37, and 77 and 30 dominant mutations in host and parasitic RNAs for E2 and E3, respectively. Among these mutations, 30 and 20 mutations in host RNAs of E2 and E3, respectively, and 17 and 10 mutations in parasitic RNAs of E2 and E3, respectively, were not found in the dominant mutations of the main experiment, although 42 and 20 mutations were common in host and parasitic RNAs in all long-term replication experiments (Supplementary Figs. S6a and b). These data show that different host and parasitic RNAs evolved in the different experiments.

Next, we created consensus genotypes and phylogenetic trees based on the dominant mutations identified in both experiments (Supplementary Figs. S6c and d). We first defined the ancestral host RNA lineages (HL0) so that these lineages include all genotypes that have the same sets of mutations found in the ancestral host RNA lineage in the main experiment. In addition, we defined two host RNA lineages (HL4 and HL5) and two parasitic RNA lineages (PL4 and PL5) in E2, and three host RNA lineages (HL6, HL7, and HL8) and two parasitic RNA lineages (PL6 and PL7) in E3, which accumulated different sets of mutations. Although the phylogenetic trees indicate relatedness between PL4 and PL5 and between PL6 and PL7, these parasitic RNA lineages had different

deletion regions and may have originated from distinct host RNAs, as described in Supplementary Text 3.

We then analyzed the population dynamics of each lineage using the 100 most frequent host and parasitic RNA genotypes at each sequenced round. The genotypes in different lineages show distinct patterns of mutation accumulation (Supplementary Figs. S8 and S9). The frequency of each lineage for both E2 and E3 (Supplementary Fig. S10) throughout the rounds revealed similar trends to the main long-term replication experiment (Figs. 3b and 3c), i.e., gradual diversification toward relatively stable coexistence of the evolved lineages. In E2, the host RNA lineages, HL5 and HL6, were first detected at round 33. Their frequencies varied from less than 0.1% to nearly 100% up to round 200 and afterwards were persistently maintained at more than 15% of the population. In parasitic RNAs, PL4 was detected throughout the sequenced rounds with more than 46% of the population, but PL5 coexisted as a dominant lineage from round 200 with similar frequencies. In E3, the host RNA lineages, HL8, HL9, and HL10 successively appeared. Their frequencies varied from less than 0.1% to nearly 100% up to round 198, and thereafter, all three lineages were consistently maintained at more than 0.5% of the population. In parasitic RNAs, PL6 was detected throughout the sequenced rounds with more than 83% of the population, but PL5 coexisted as 0.1–17% of the population from round 198. Overall, the four and five RNA lineages in E2 and E3, respectively, coexisted in the last ~40 rounds. These results support the possibility that a replicating RNA complexifies toward replicator communities through Darwinian evolution.

2. Investigation of higher-order interactions between the selected RNA clones

The interactions between RNA clones (Fig. 3d) were determined by examining each RNA replication by the replicase of a specific RNA. However, in the long-term replication experiment (Fig. 1c), more than two types of RNAs were co-replicated, which may show

more complex interactions. To examine the existence of such higher-order interactions, we incubated each RNA clone, a pair of two RNA clones, or a combination of three RNA clones for all possible RNA sets of selected rounds at 37 °C for 5 h and determined the extent of RNA (co-) replications through the translation of replicases (Supplementary Fig. S17). Next, we estimated the contribution of higher-order interactions arising only in the presence of three RNA clones by subjecting the replication data to Bahadur expansion analysis[1,2]. In Bahadur expansion analysis, the measured replication amounts of each RNA were converted into an orthogonal system consisting of interaction terms. For example, considering interactions among RNA_i, RNA_j, and RNA_k, the contributions of RNA_j, RNA_k, and a set of RNA_j and RNA_k to the replication of RNA_i were quantified as Bahadur coefficients w_j, w_k, and w_{jk}, respectively, in a comparable form (lower coefficients indicate smaller contributions, and vice versa). Supplementary Fig. S18 shows the calculated Bahadur coefficients in all examined combinations of three RNA clones and the sum of coefficients of determination (R^2) for first-order Bahadur coefficients (e.g., w_j and w_k), which represent the contribution of direct interactions between two RNAs. R^2 values in 40 out of 45 cases exceeded 0.7; in the other five cases, the effect of measurement errors became significant because all Bahadur coefficients were extremely low. These results indicated that higher-order interactions among three RNA clones did not significantly contribute to our results; the interdependent replication of RNA clones can essentially be understood from interactions between two RNAs, as represented in Fig. 3d.

3. Deletion sites of dominant parasitic RNAs

For parasitic RNA lineages that appeared in the main long-term replication experiment, PL1 deleted 224–536 and 744–1961 nt based on the ancestral host RNA sequence, as observed for parasite-γ in the previous study[3], whereas PL2 and PL3 deleted 178–526 and 766–1943 nt, which was not previously identified. For parasitic RNA lineages in the two

additional experiments, PL5 and PL7 deleted similar RNA regions to PL2 and PL3, whereas PL4 and PL6 deleted previously unidentified regions, 173–293 and 673–1956 nt, and 224–536 and 745–1955 nt, respectively.

4. Extension of the theoretical model

The theoretical model was extended to describe RNA concentration as a function of both RNA and replicase concentrations in the same compartment by assuming the translation of replicases. The differential equations are as follows:

$$\frac{d[RNA_i]}{dt} = [RNA_i] \sum_{j=1}^{5} k_{ij}[Rep_j](1 - \frac{\sum_{i=1}^{5}[RNA_i]}{C}) \quad (5)$$

and

$$\frac{d[Rep_i]}{dt} = k_i^{\mathrm{t}}[RNA_i](1 - \frac{\sum_{i=1}^{5}[Rep_i]}{C^{\mathrm{t}}}), \quad (6)$$

where $[Rep_i]$, k^{t}_i, and C^{t} are the concentration of the replicase translated from RNA$_i$ in each compartment, the rate constant of replicase translation for RNA$_i$, and the carrying capacity for translation, respectively. Other parameters were not modified or added. k^{t}_i was set to 1 ($1 \leq i \leq 3$, host RNAs) or 0 ($4 \leq i \leq 5$, parasitic RNAs) as the original rate constant (k_{ij}, determined based on experiments) encompasses the translation activity. C^{t} was set to 30. Using the extended model, we simulated the continuous replication of the five RNAs (based on HL1-, HL2-, HL3-, PL2-, and PL3-228) and obtained similar concentration dynamics to those based on the original simpler model (Fig. 5b and Supplementary Fig. S24). We note that we did not explicitly model the association and dissociation of an RNA and a replicase due to the lack of experimental data for the properties of replicases. However, we believe that such modification of the model has a minor effect on the dynamics because the continuous replication of the four RNAs was reproduced by the simple model described above.

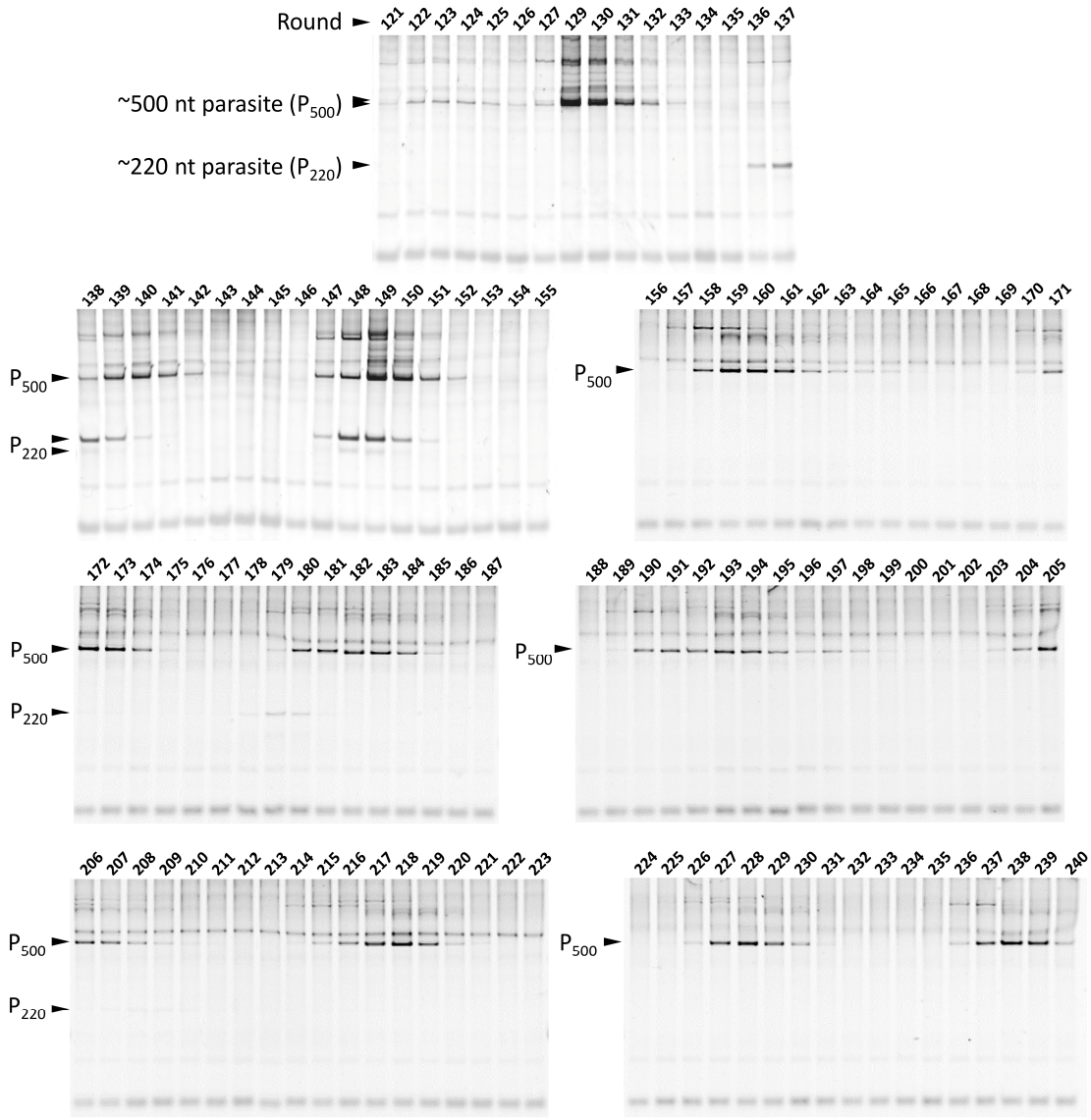


Fig. S1 | Native polyacrylamide gel analysis of parasitic RNAs during the long-term replication experiment. RNA mixtures at rounds 121–240 were subjected to native polyacrylamide gel electrophoresis. Band intensities of ~220 nt and ~500 nt parasitic RNAs were quantified and plotted in Fig. 1c. Multiple bands were sometimes detected in each class of the parasitic RNAs due to structural or size heterogeneity. Source data are provided as a Source Data file.

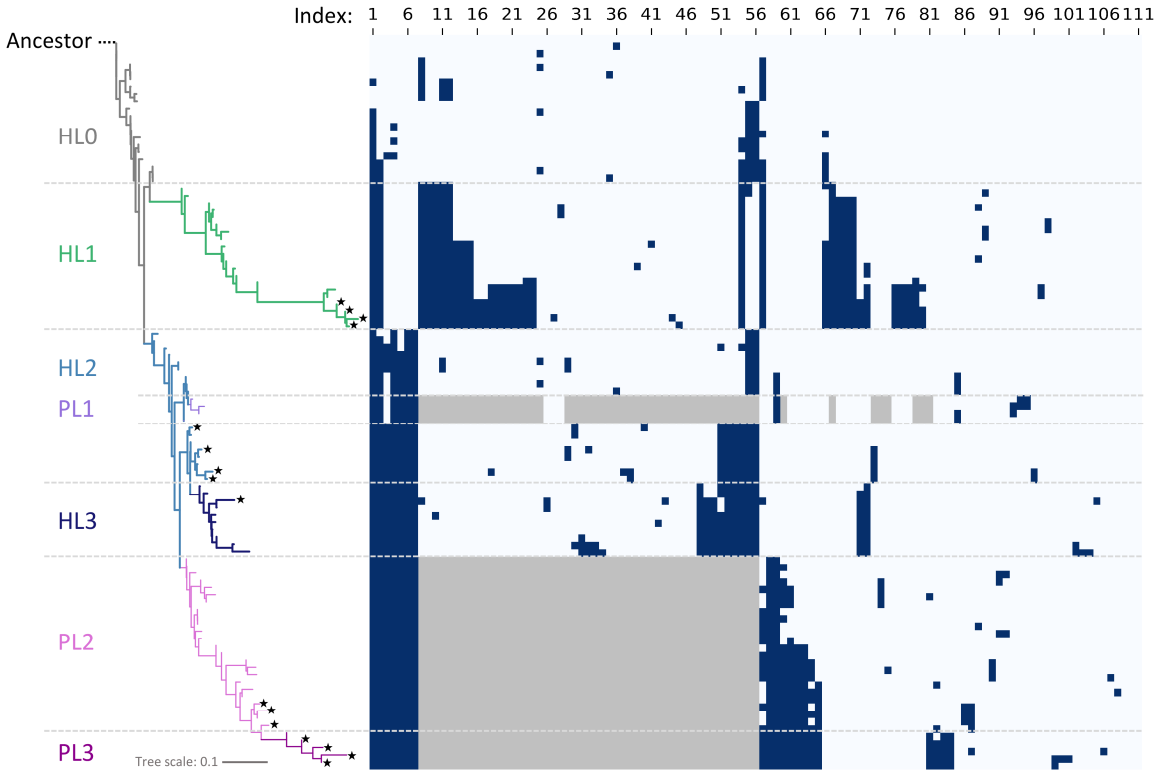
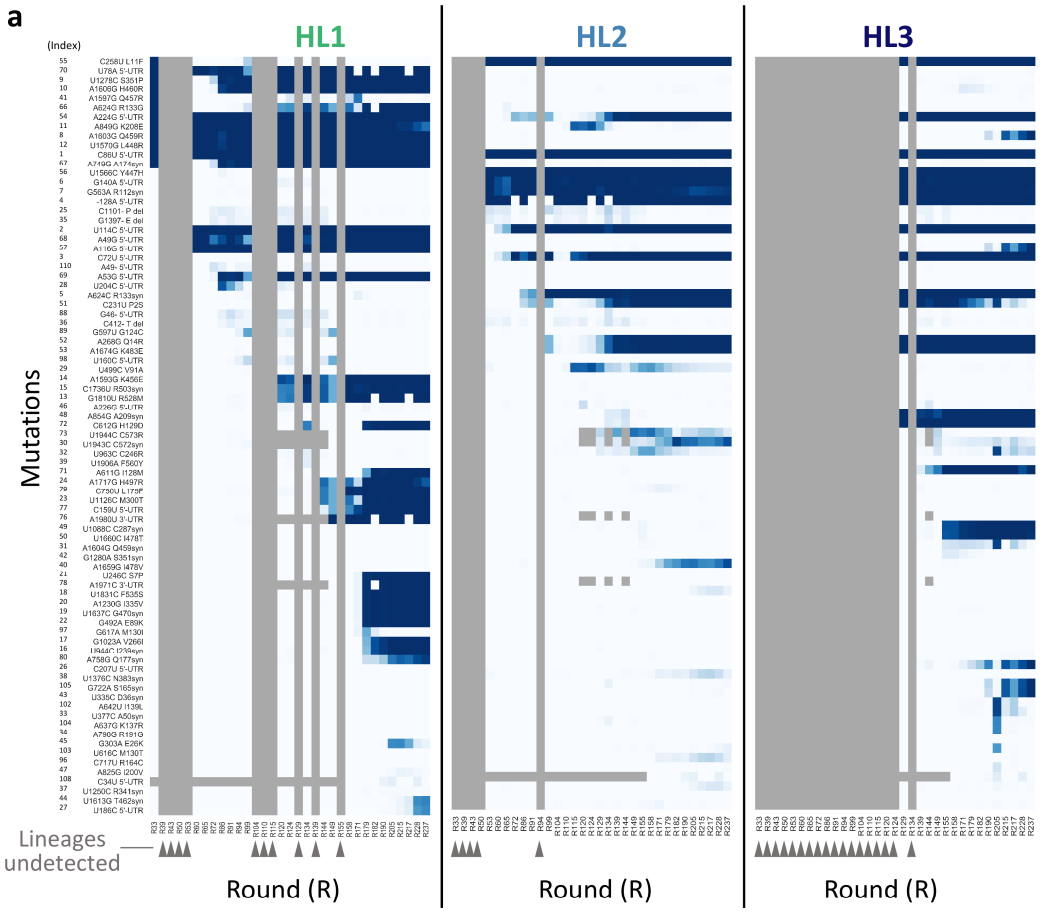


Fig. S2 | An enlarged view of the dominant mutation map in Fig. 2. Navy and grey colors indicate the presence of a point mutation and deletion, respectively. Mutation indices at the top correspond to ones in Supplementary Figs. S3 and S4.



only HL1	28
only HL2	3
only HL3	6
only HL1 & 2	0
only HL1 & 3	4
only HL2 & 3	9
All three lineages	3

Fig. S3 | Dominant mutations and fixation dynamics in the host RNA lineages. a, Dominant mutations accumulated in HL1, HL2, and HL3 over rounds. The base numbers are based on the original host RNA. "syn" and "del" in mutation names stand for synonymous and deletion, respectively. Numbers to the left of the mutation names correspond to mutation indices shown in Supplementary Fig. 2. The intensity of the blue color indicates fixation frequency. Lineages were not detected at rounds marked with grey

arrowheads. Some mutations near 5′ and 3′ ends (e.g., index 73, U1994C, C573R) were not determined at several rounds because we used different primers for efficient cDNA library preparation. **b**, Number of fixed mutations (accumulated in more than 50% sequences) in specific lineages at the last sequenced round (237). **c**, Number of fixed mutations in each lineage over rounds.

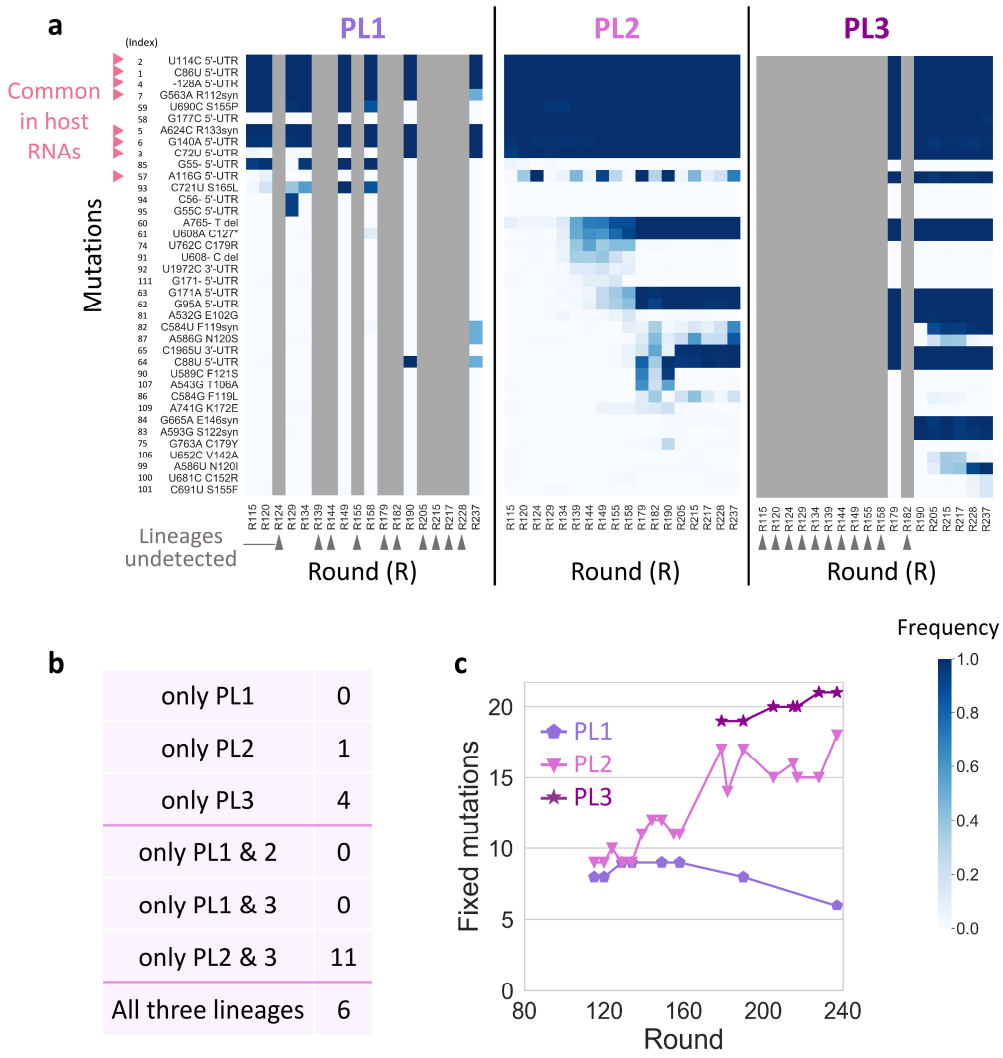


Fig. S4 | Dominant mutations and fixation dynamics in the parasitic RNA lineages. **a**, Dominant mutations accumulated in PL1, PL2, and PL3 over rounds. The base numbers are based on the original host RNA. "syn" and "del" in mutation names stand for synonymous and deletion, respectively. Numbers to the left of the mutation names correspond to mutation indices shown in Supplementary Fig. S2. The intensity of the blue color indicates fixation frequency. Grey regions indicate that mutations were not observed because lineages were not detected (indicated with grey arrowheads at rounds). Pink arrowheads indicate mutations commonly observed in host RNAs (Supplementary Fig. S3). Deletions at recombination sites were not shown. **b**, The number of fixed mutations (accumulated in more than 50% sequences) in specific lineages at the last sequenced round (237). Deletions due to different recombination sites were not counted. **c**, Number of fixed mutations in each lineage over rounds.

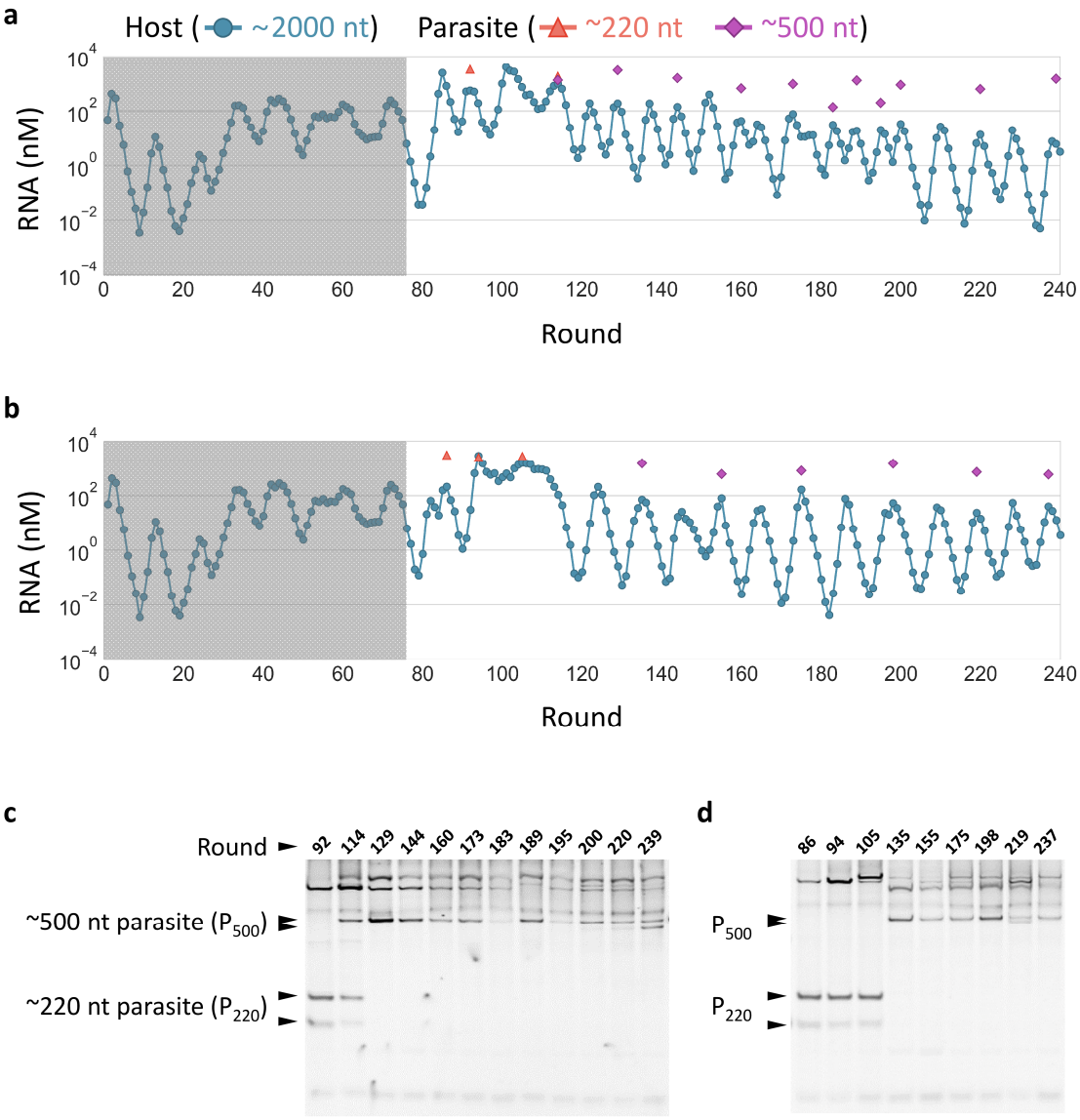


Fig. S5 | Two additional long-term replication experiments. a–b, Concentration changes of host and parasitic RNAs of different lengths in E2 (a) and E3 (b), where we newly performed 164 cycles of replications started with the droplet mixture at round 76 of the main long-term replication experiment. Parasitic RNA concentrations were determined only at sequenced rounds. The plot of host RNA concentrations in the shaded regions (up to 76 round) is the same as that of Fig. 1c. **c–d**, Native polyacrylamide gel electrophoresis of RNA mixtures in E2 (c) and E3 (d). Band intensities of ~220 nt and ~500 nt parasitic RNAs were quantified and plotted in panels a and b. Multiple bands were sometimes detected in each class of the parasitic RNAs due to structural or size heterogeneity. Source data are provided as a Source Data file.

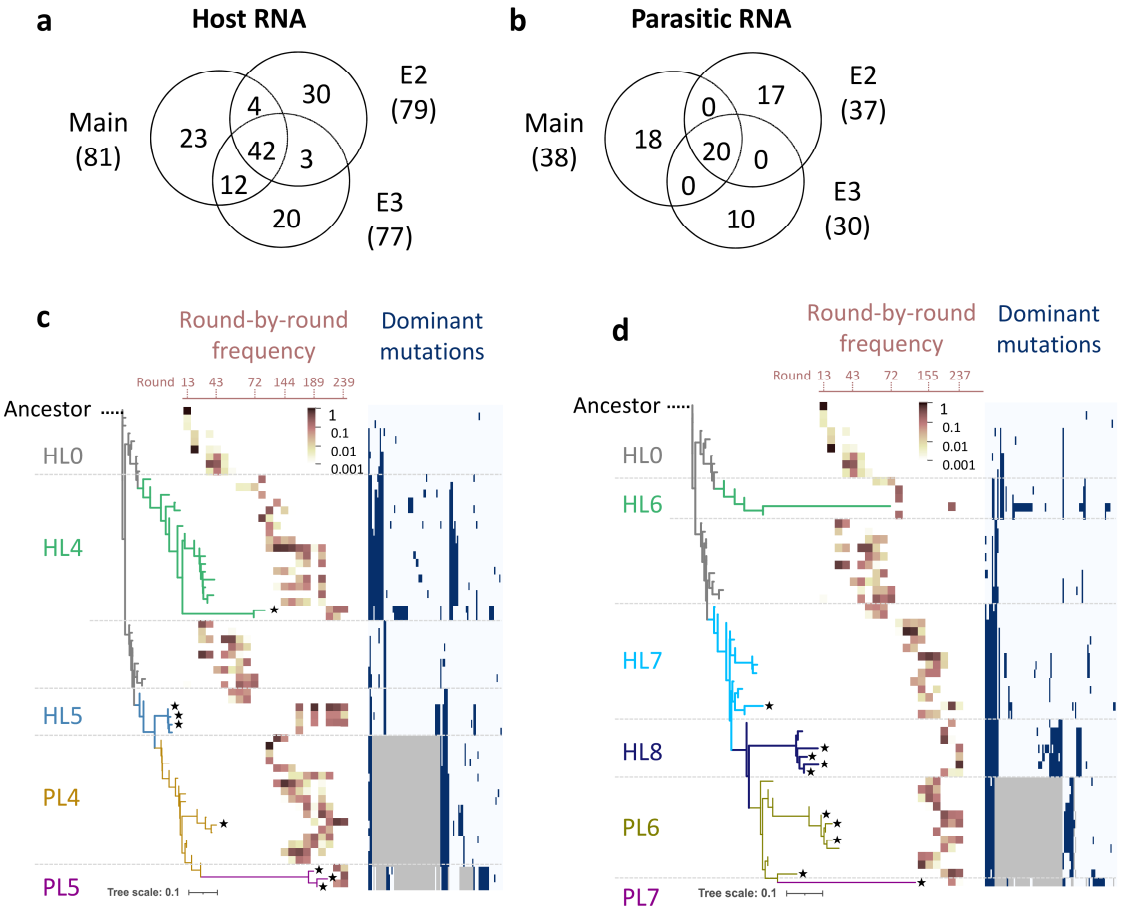


Fig. S6 | Sequence and phylogenetic analyses of the two additional long-term replication experiments. a–b, Venn diagrams showing the number of dominant mutations in host RNAs (a) and parasitic RNAs (b) identified in each of the three long-term replication experiments. Numbers in parenthesis at the name of experiments indicate the total mutation numbers. **c–d**, Phylogenetic trees were constructed based on the three most frequent host and parasitic RNA genotypes in all sequenced rounds for E2 (c) and E3 (d). The ancestral host RNA ("Ancestor") was designated as the root of the trees. Branches comprising defined lineages are colored differently. Host and parasitic RNA lineages are shown as thick and thin lines, respectively. The heatmaps superimposed on the trees show the frequencies of each genotype over all sequenced rounds (from left to right). Black star shapes at the tips of branches mark genotypes that remained to the last sequenced round. The lists of dominant mutations are shown on the right; navy and grey colors indicate the presence of a point mutation and deletion, respectively. An enlarged view of the list for each experiment is available in Supplementary Fig. S7.

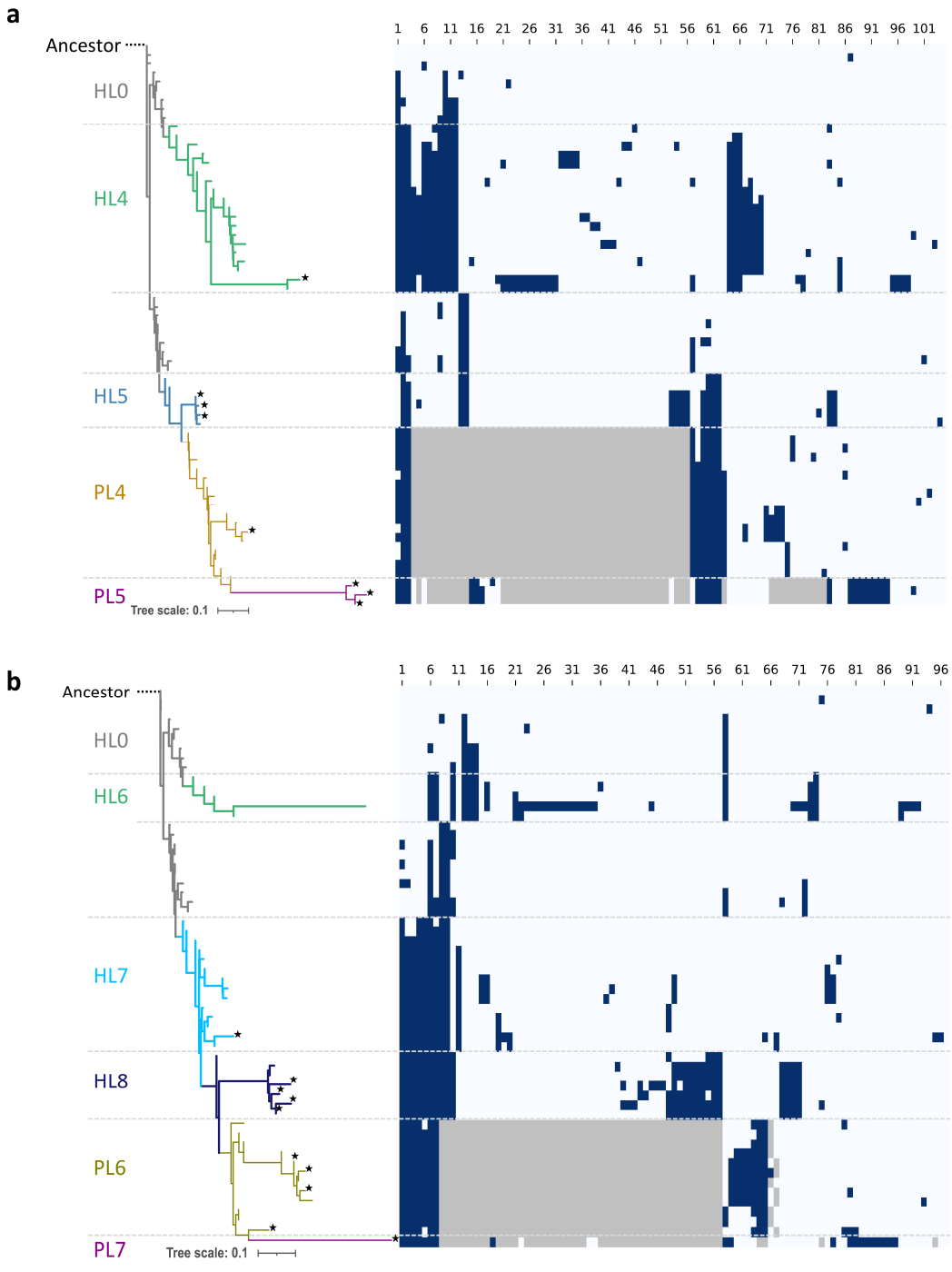


Fig. S7. Enlarged views of the dominant mutation maps in Supplementary Figs. S6c and d. a–b, The maps correspond to those in Supplementary Figs. S6c (a) and d (b). Navy and grey colors indicate the presence of a point mutation and deletion, respectively. Mutation indices at the top correspond to ones in Supplementary Figs. S8 (a) and S9 (b).

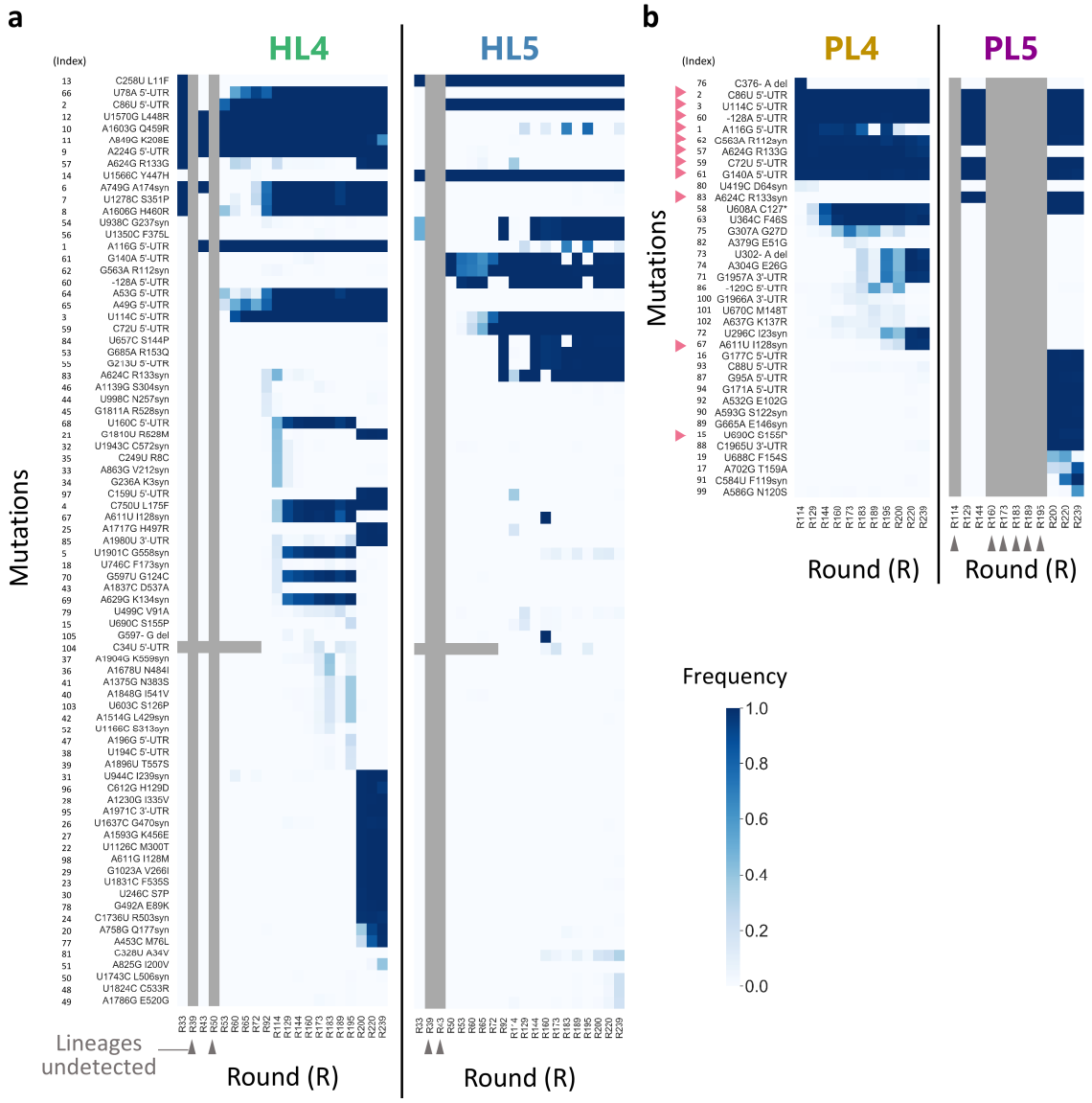


Fig. S8 | Dominant mutations and fixation dynamics in the host and parasitic RNA lineages in E2. a–b, Dominant mutations accumulated in HL4 and HL5 (a), or PL4 and PL5 (b) over rounds. The base numbers are based on the original host RNA. "syn" and "del" in mutation names stand for synonymous and deletion, respectively. Numbers to the left of the mutation names correspond to mutation indices shown in Supplementary Fig. S7. The intensity of the blue color indicates fixation frequency. Grey regions indicate that mutations were not observed because lineages were not detected (indicated with grey arrowheads at rounds) or different primers were used for cDNA library preparation. Pink arrowheads at mutations for parasitic RNA lineages indicate those commonly observed in host RNAs

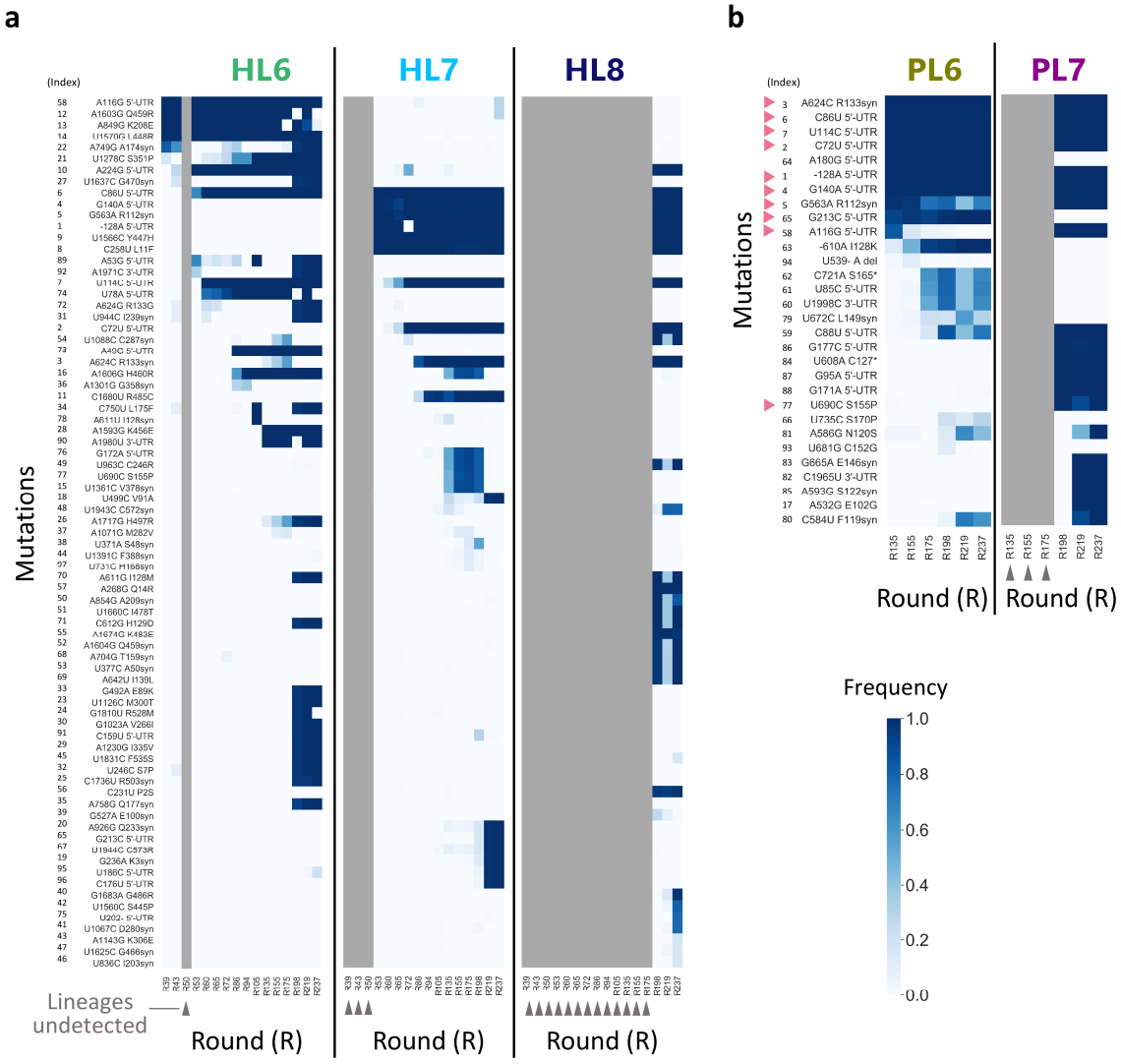


Fig. S9. Dominant mutations and fixation dynamics in the host and parasitic RNA lineages in E3. a–b, Dominant mutations accumulated in HL6, HL7, and HL8 (a), or PL6 and PL7 (b) over rounds. The base numbers are based on the original host RNA. "syn" and "del" in mutation names stand for synonymous and deletion, respectively. Numbers to the left of the mutation names correspond to mutation indices shown in Supplementary Fig. S7. The intensity of the blue color indicates fixation frequency. Grey regions indicate that mutations were not observed because lineages were not detected (indicated with grey arrowheads at rounds). Pink arrowheads at mutations for parasitic RNA lineages indicate those commonly observed in host RNAs.

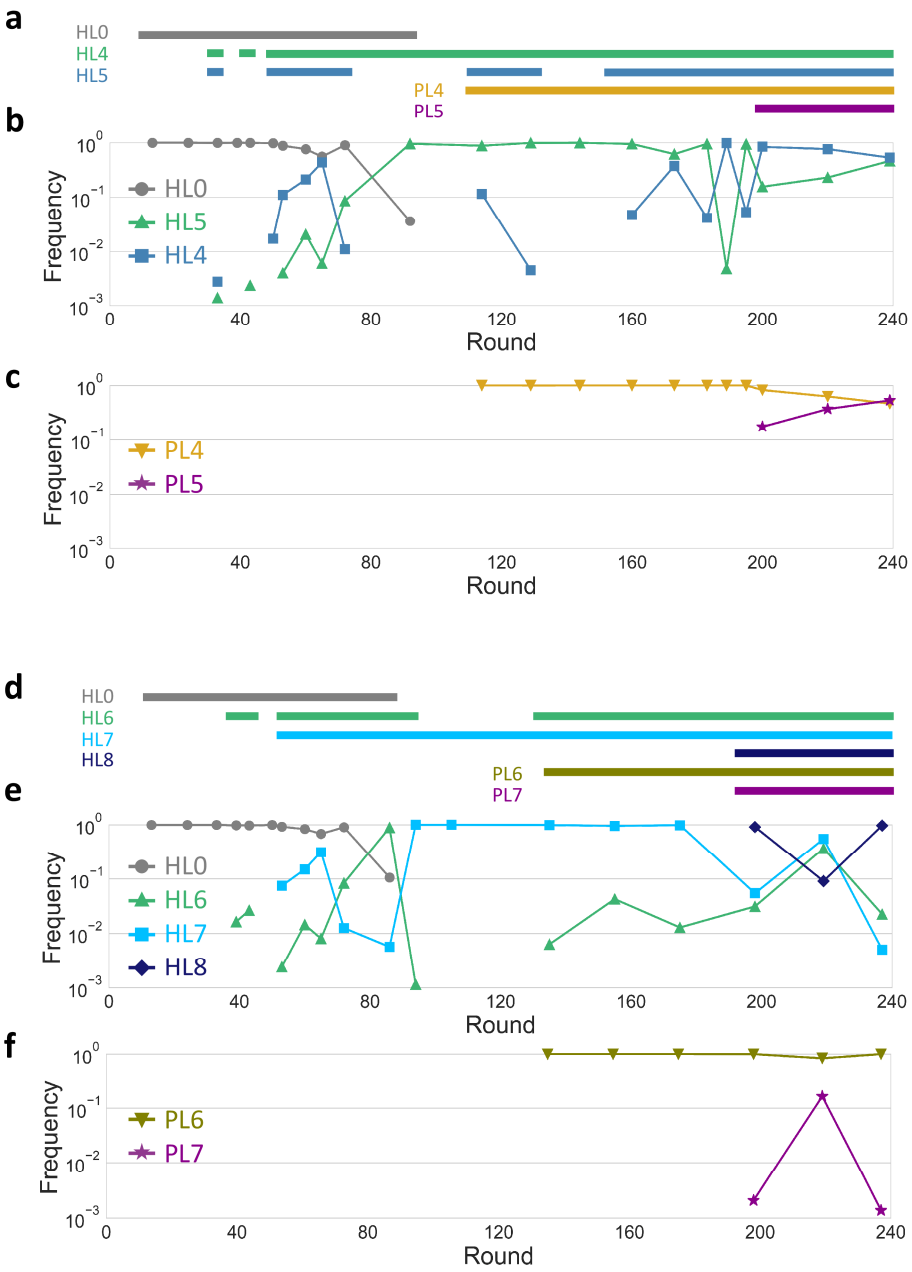


Fig. S10 | Population dynamics of the lineages in the additional long-term replication experiments. a–f, Frequencies of the lineages in total sequence reads of the analyzed genotypes for host (b for E2 and e for E3) and parasitic (c for E2 and f for E3) RNAs. Horizontal lines above the graphs (a, d) indicate rounds where the frequency of each lineage in the same color was plotted (above 0.1 %).

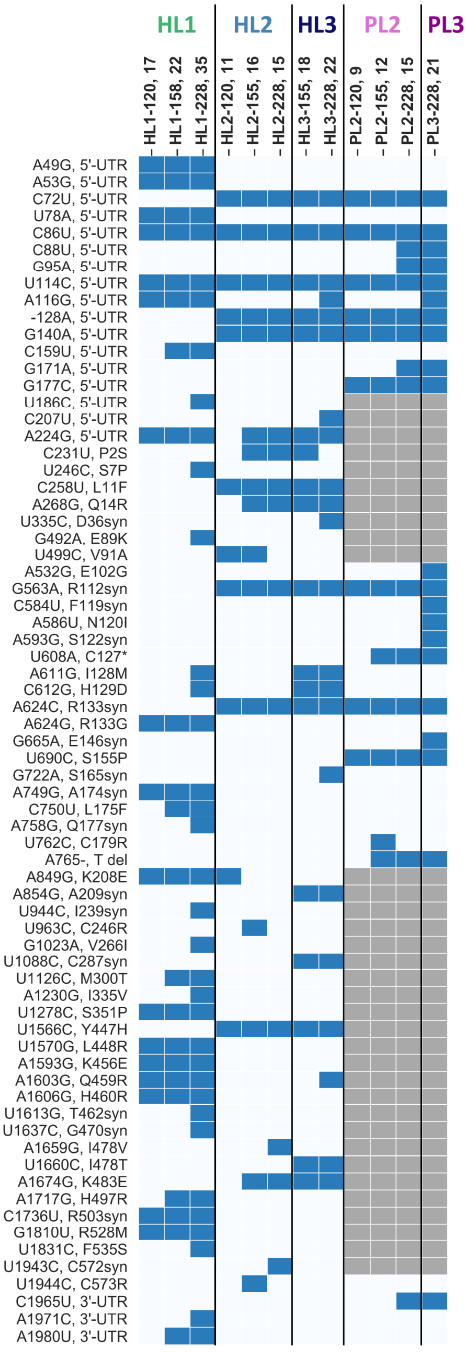


Fig. S11 | List of mutations in the selected RNA clones. The base numbers are based on the original host RNA. “syn” and “del” in mutation names stand for synonymous and deletion, respectively. Grey regions indicate deleted sites. The numbers after the names of each clone show mutation numbers.

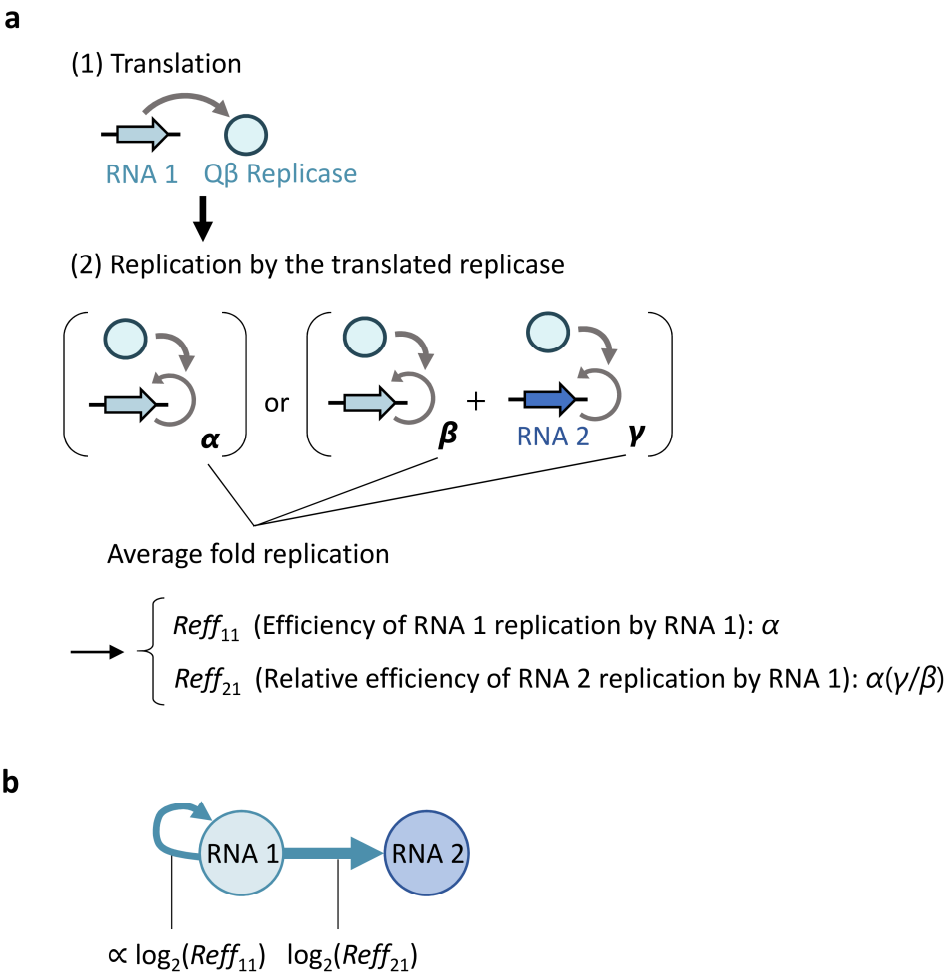


Fig. S12 | Procedure for the depiction of replication relationships between the RNA clones by directed graphs. a, In a translation-uncoupled replication experiment, where the translation of the replicase gene from RNA 1 was followed by RNA replication in the presence or absence of RNA 2 (Fig. 4a), three types of fold replications, RNA 1 replication in the absence of RNA 2 (α), RNA 1 replication in the presence of RNAs 1 and 2 (β), and RNA 2 replication in the presence of RNAs 1 and 2 (γ), were determined. Using the average of these replications above a background level (>1.5-fold), the efficiencies of RNA 1 and RNA 2 replications by the replicase translated from RNA1, $Reff_{11}$ and $Reff_{21}$, were determined. **b**, Directed graphs were depicted by setting the widths of arrows proportional to the binary logarithm of $Reff_{11}$ and $Reff_{21}$.

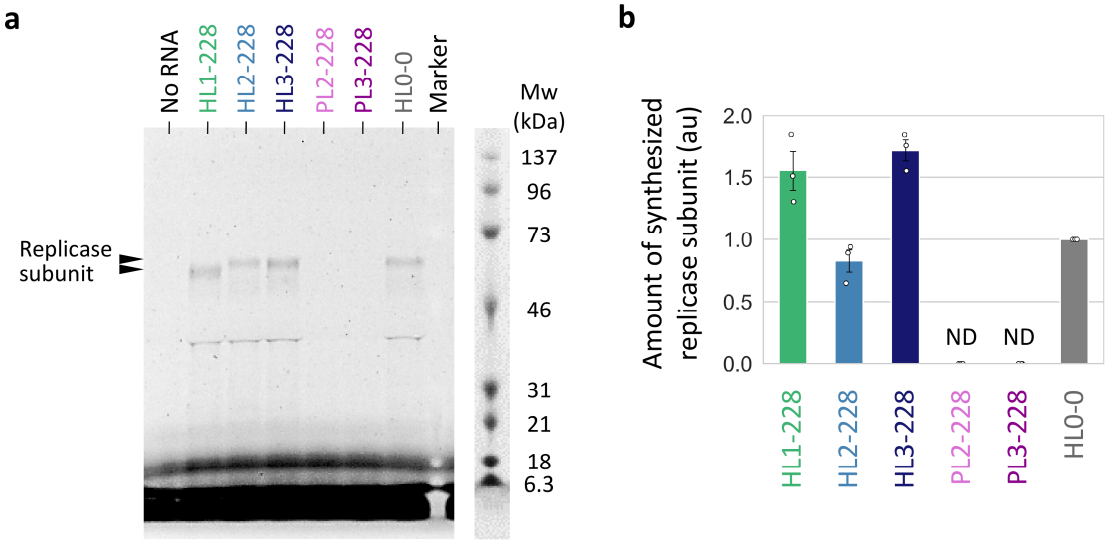


Fig. S13 | Translation activity of RNA clones at round 228. a, Protein translation was analyzed by SDS-PAGE after incubation of each RNA clone (300 nM) at 37 °C for 2 h with a fluorescently labeled lysyl-tRNA. An example of an analyzed fluorescent gel image is displayed, sided with a trimmed white-light image of the same gel (Lane "Marker") to visualize the pre-stained molecular weight (Mw) marker (right). The expected bands of the replicase subunit (~64 kDa) are indicated by the black arrow. Translated proteins from the parasitic RNAs were possibly undetectable with this experimental setup. Source data are provided as a Source Data file. **b**, Amount of synthesized replicase subunit, normalized to that of the ancestral host RNA (HL0-0). Error bars indicate mean ± SEM (n = 3). Measurements were taken from distinct samples. ND, not detected.

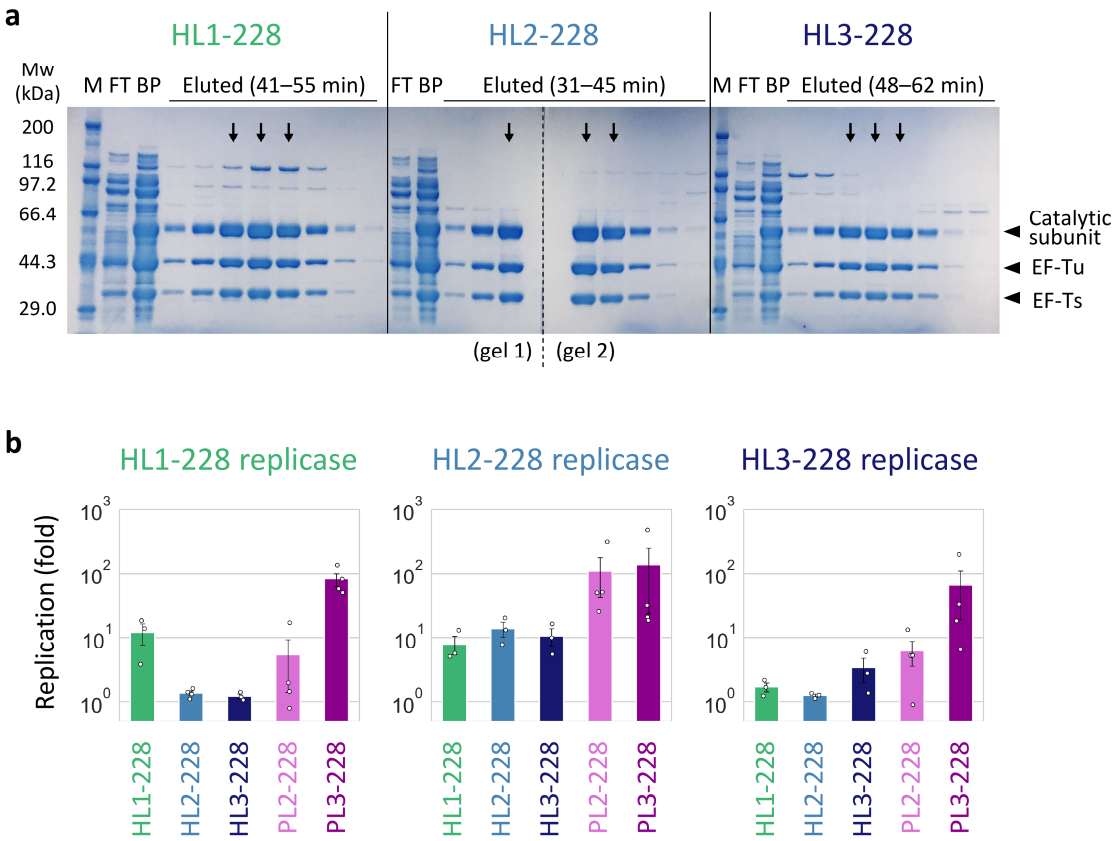


Fig. S14 | Replication of RNA clones at round 228 by their encoded replicases. a, Purification of mutant Qβ replicases composed of EF-Tu, EF-Ts, and each of the catalytic subunits encoded by HL1-, HL2-, and HL3-228. The purified replicases after cation exchange chromatography were analyzed by 10% SDS-PAGE. M, molecular weight (Mw) marker; FT, flow-through fraction; BP, samples before purification; Eluted, samples eluted at the indicated times. The expected bands of the catalytic subunit (~64 kDa), EF-Tu (~43 kDa), and EF-Ts (~30 kDa) are indicated by the black arrowheads. Two separate gels were displayed as indicated. Eluted fractions including ones indicated by the black arrows were collected as purified Qβ replicases. Source data are provided as a Source Data file. **b**, Replication of the RNA clones (10 nM) by each of the purified Qβ replicases (10 nM) at 37 °C for 2 h, measured by RT-qPCR. Error bars indicate mean ± SEM (n = 3 or 4 as shown as individual data points). Measurements were taken from distinct samples.

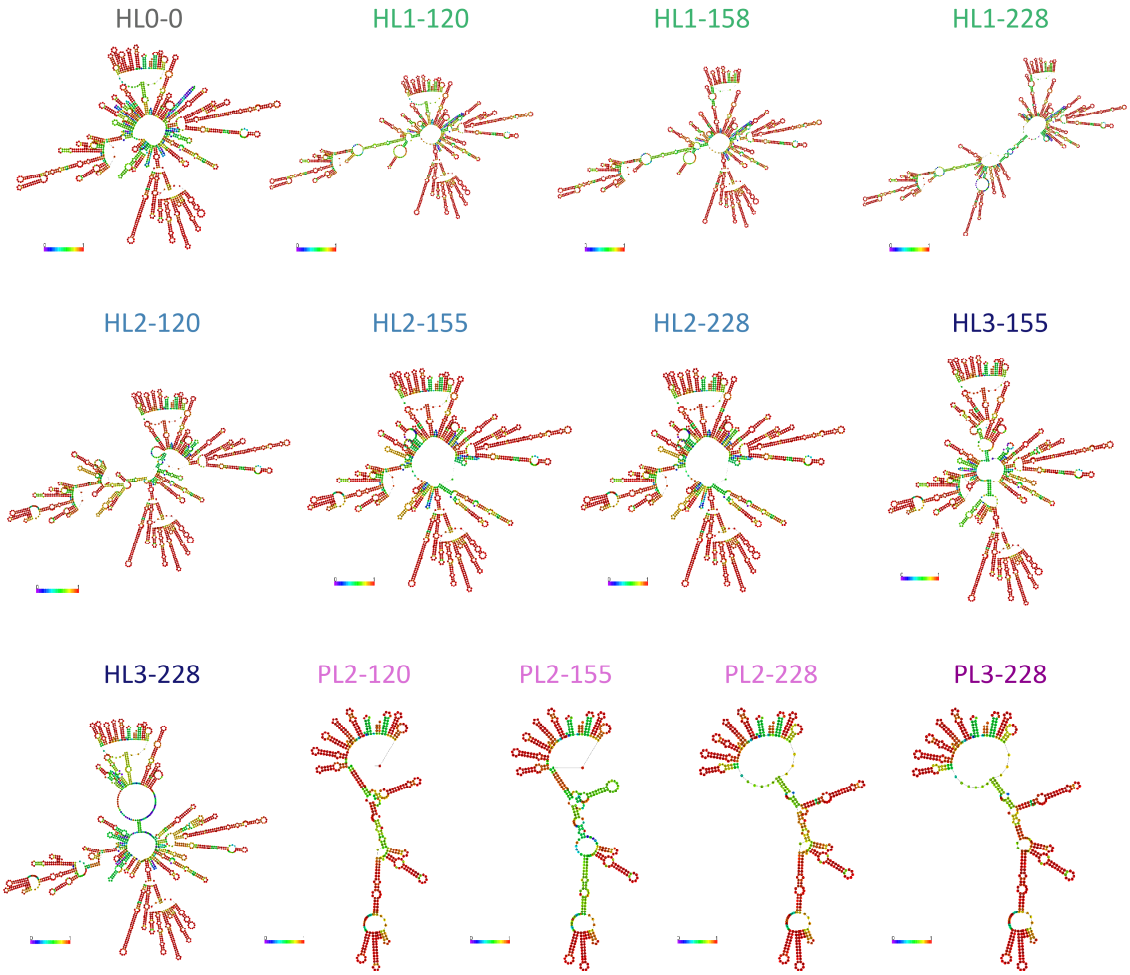


Fig. S15 | Typical secondary structures of the RNA clones (plus strands). Centroid structures predicted by ViennaRNA[4] are shown. Colors indicate the probability of base pairing, from purple to red (more probable). RNA sequences are available in Supplementary Data 1.

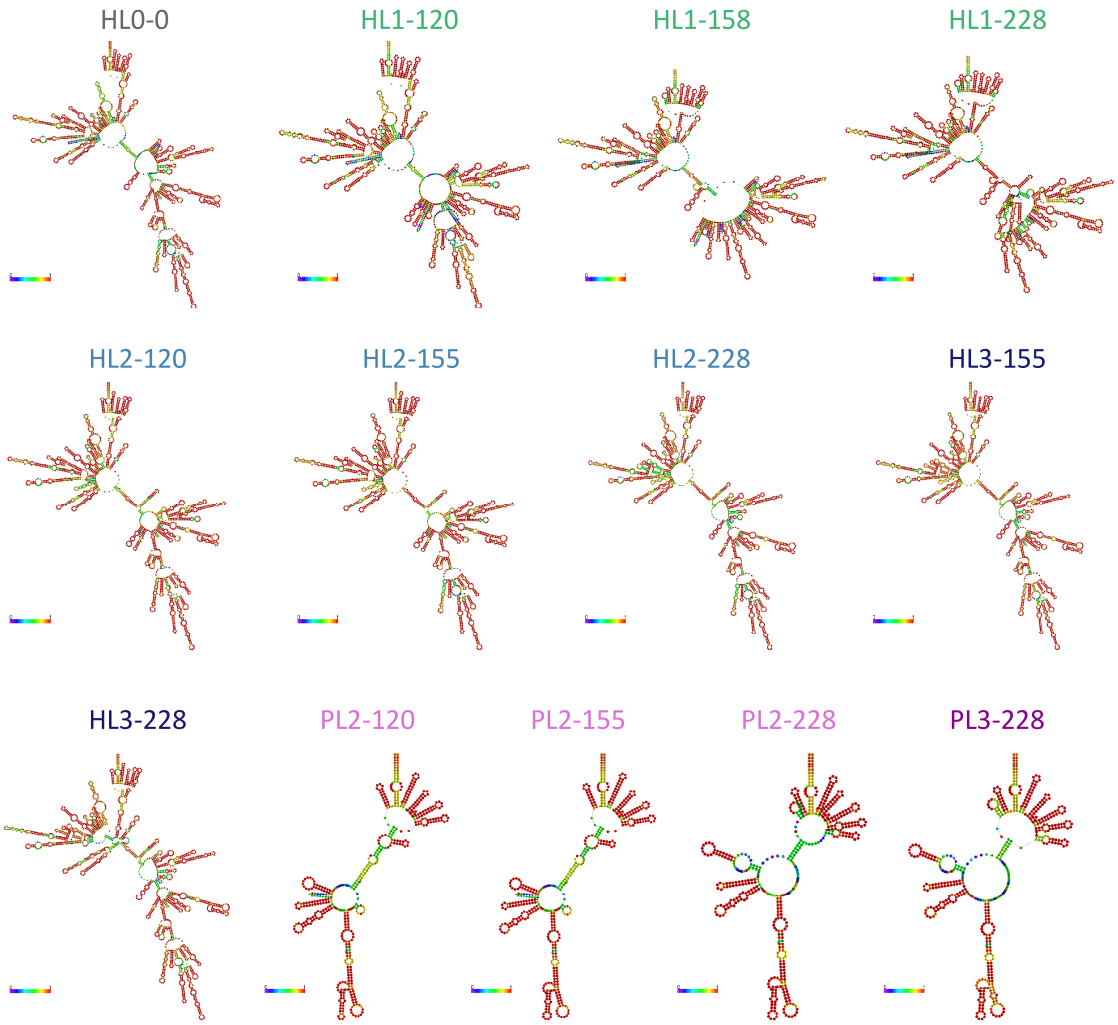


Fig. S16 | Typical secondary structures of the RNA clones (minus strands). Centroid structures predicted by ViennaRNA[4] are shown. Colors indicate the probability of base pairing, from purple to red (more probable). RNA sequences are available in Supplementary Data 1.

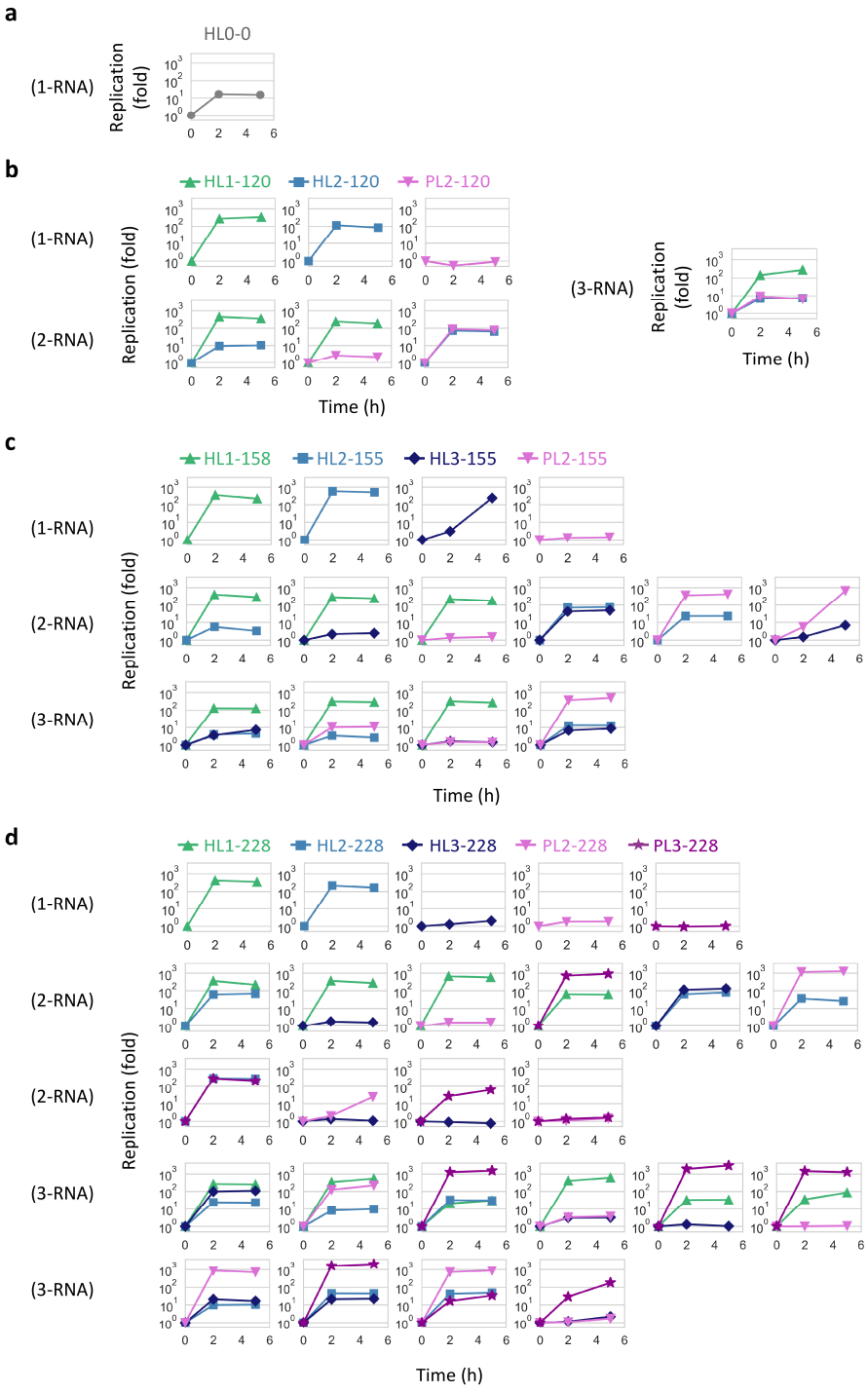


Fig. S17 | Translation-coupled replication experiments. a–d, One, two, or three RNA clones (10 nM each) at rounds 0 (a), 120 (b), 155–158 (c), and 228 (d) were incubated at 37 °C for 5 h in the translation system, and replications of each RNA at 2 and 5 h were measured by sequence-specific RT-qPCR.

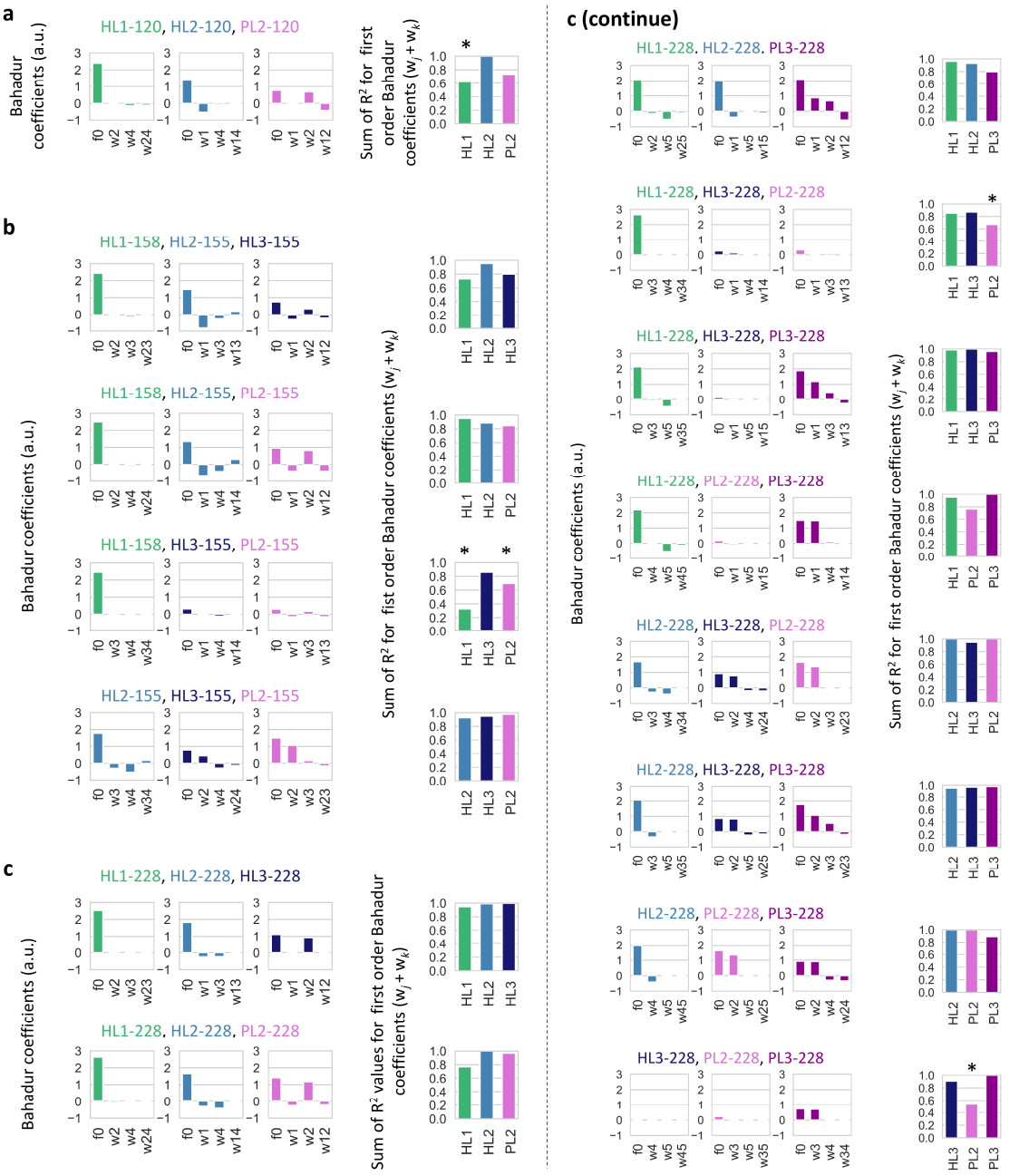


Fig. S18 | Bahadur expansion analysis. a–c, Bahadur coefficients (left) and the sum of coefficients of determination (R^2) for first order Bahadur coefficients (i.e., $w_j + w_k$) (right) for each combination of three RNA clones at rounds 120 (a), 155–158 (b), and 228 (c), calculated from fold replications (at 2 h) in the translation-coupled replication experiments (Supplementary Fig. S17). Number j in w_j is 1, 2, 3, 4, and 5 for RNA clones in HL1, HL2, HL3, PL2, and PL3, respectively. 5 out of 45 cases for which calculated R^2 are low (<0.7) are indicated with asterisks.

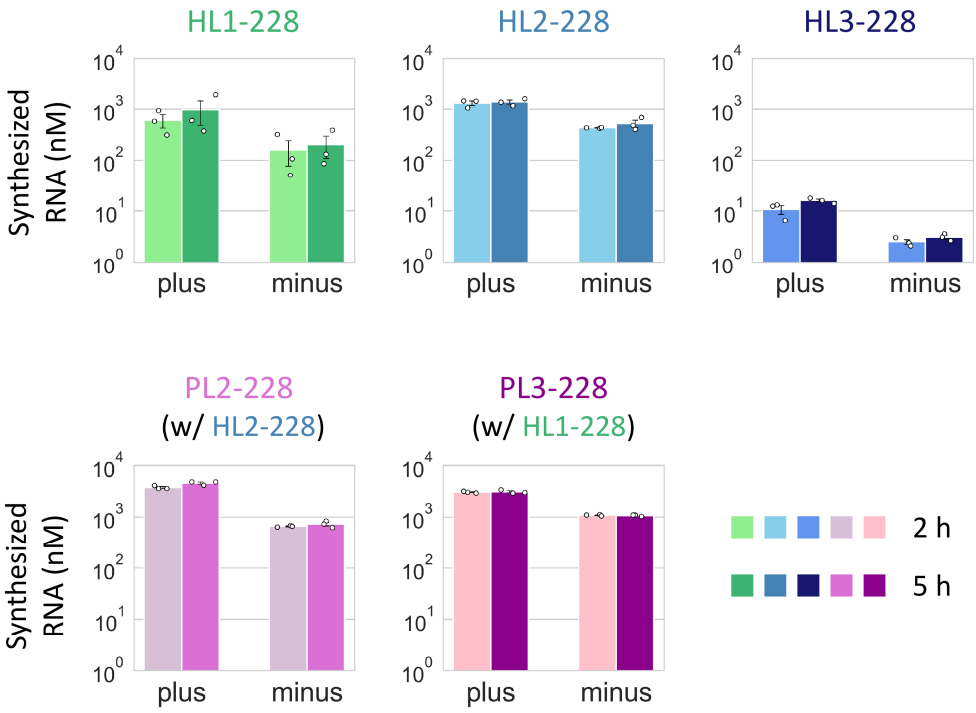


Fig. S19 | Synthesis of plus and minus strands during translation-coupled RNA replication. 10 nM of each RNA clone at round 228 was incubated at 37 °C for 5 h in the translation system. PL2- and PL3-228 were incubated in the presence of host RNA clones that replicated each RNA most efficiently (Fig. 3d). The amounts of synthesized plus and minus strand RNAs at 2 and 5 h were measured by each strand-specific RT followed by qPCR. Error bars indicate mean ± SEM (n = 3). Measurements were taken from distinct samples. Source data are provided as a Source Data file.

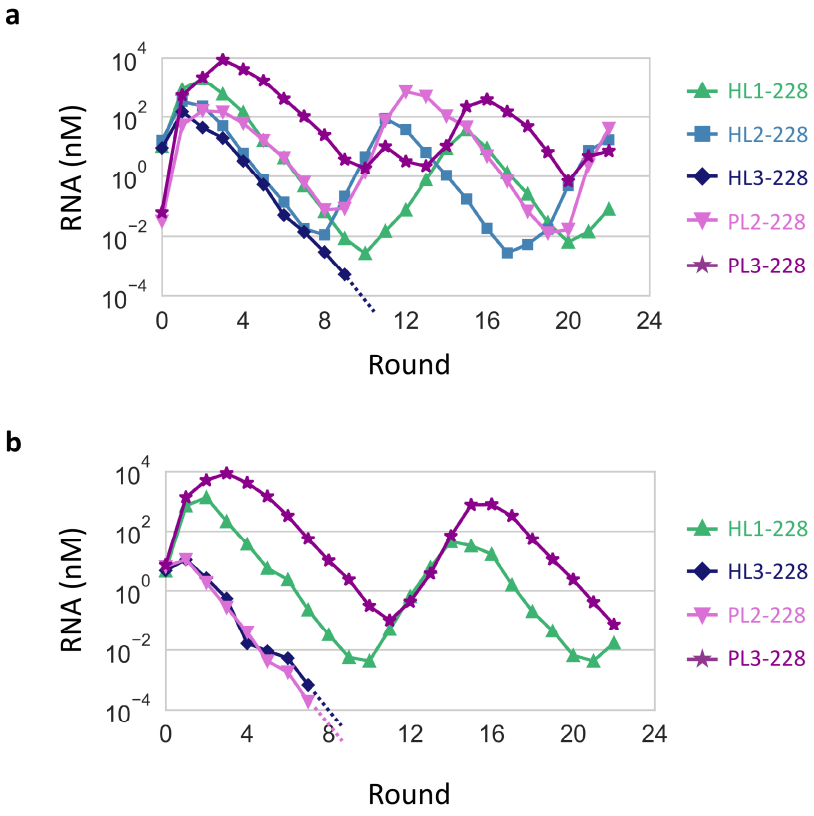


Fig. S20 | Additional long-term replication experiments started with a mixture of the RNA clones at round 228. a–b, RNA concentration changes in long-term replication experiments initiated with 10 nM each of HL1-, HL2-, and HL3-228, and 0.1 nM each of PL2- and PL3-228 (a) or 10 nM each of HL1-, HL3-, PL2-, and PL3-228 (b). The concentrations were measured by sequence-specific RT-qPCR.

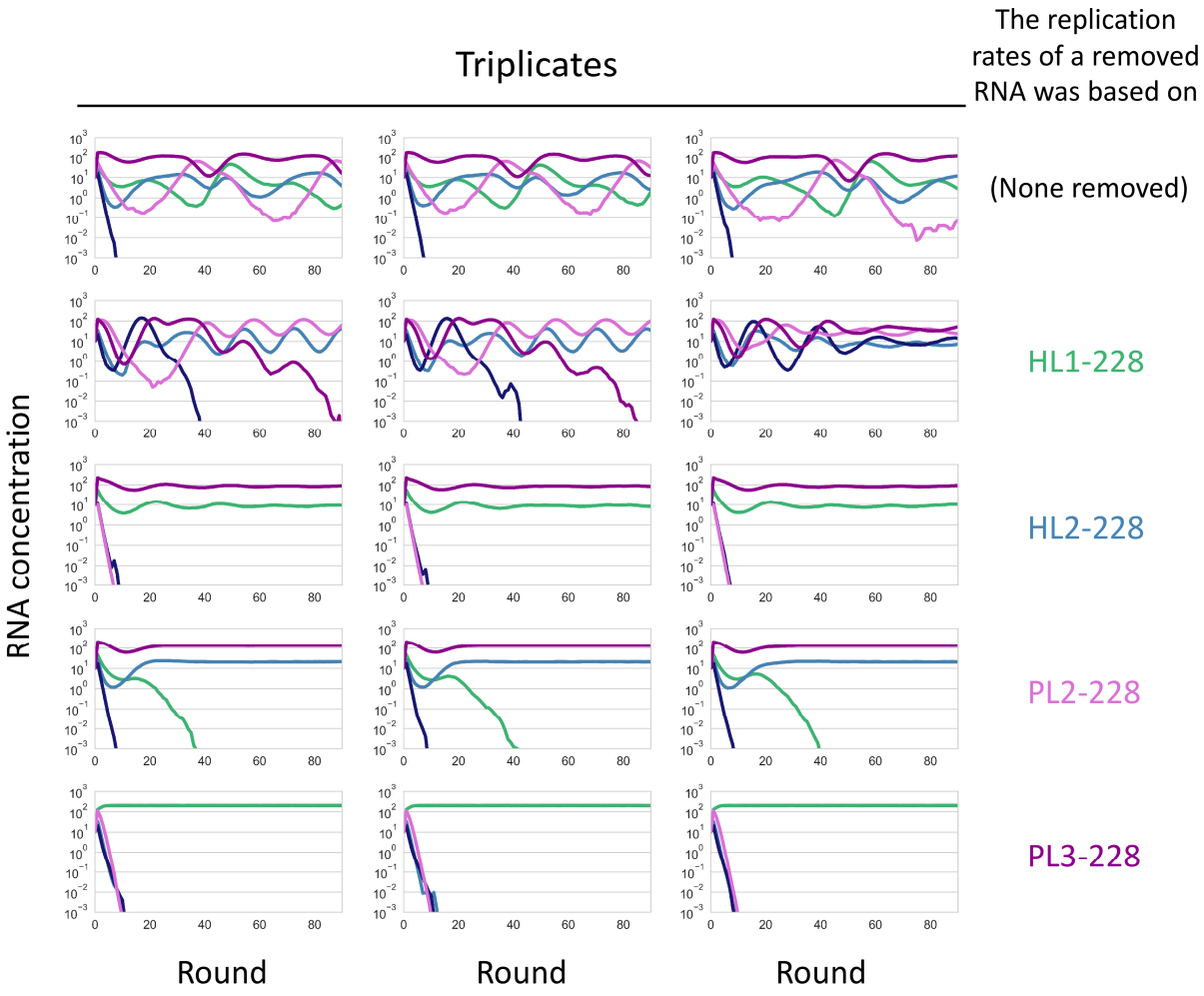


Fig. S21 | Dynamics of the RNA replicator network in the absence of one of the RNAs. Simulations were performed as that presented in Fig. 5b, in the absence of one of the four RNAs that sustainably replicated. Each simulation was performed three times independently. The upper leftmost panel is the same as Fig. 5b.

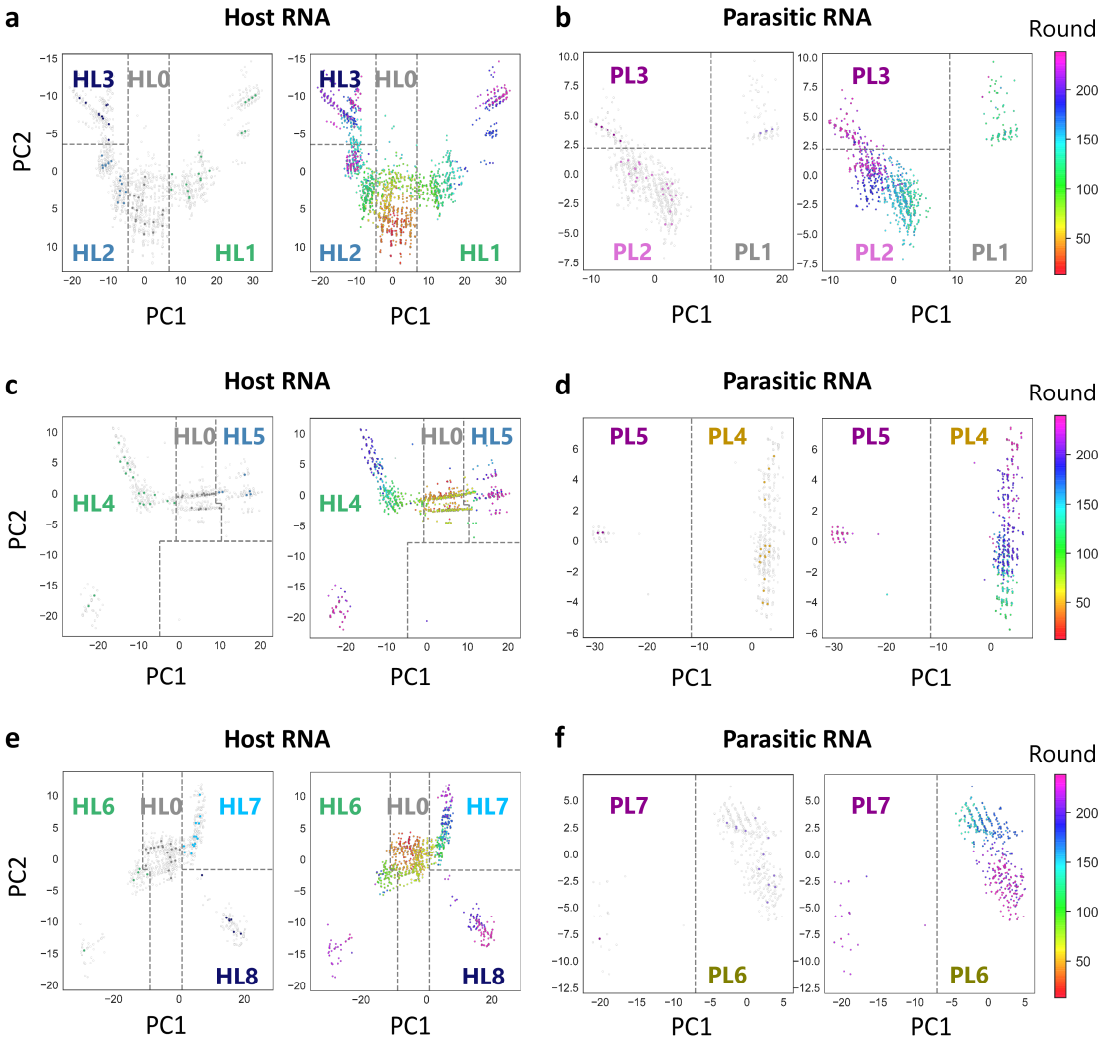


Fig. S22 | Mapping of consensus host and parasitic RNA genotypes in sequence spaces. a–f, Two-dimensional (2D) maps were created based on Hamming distances between all top 100 consensus host RNA genotypes or ~500 nt parasitic RNA genotypes obtained throughout the main long-term replication experiment (a, b), the additional long-term replication experiment E2 (c, d), and that of E3 (e, f). The Hamming distance matrices were plotted on the maps using Principal Coordinate Analysis for dimension reduction. In the 2D maps, each genotype is shown as a point; genotypes located in closer areas are expectedly more related. Left plots in each panel highlight genotypes represented in the phylogenetic trees (Fig. 2 and Supplementary Fig. S6), colored to indicate lineages defined based on the trees. All displayed genotypes were then classified in each lineage as indicated. Right plots show the same maps with genotypes colored to indicate the appearance of rounds.

		RNA_i				
		HL1-228	HL2-228	HL3-228	PL2-228	PL3-228
RNA_j	HL1-228	2.615577	0	0	0	3.372959
	HL2-228	1.419954	2.79044	1.422017	5.711838	3.040918
	HL3-228	0	0	0.438937	0	3.058117
	PL2-228	0	0	0	0	0
	PL3-228	0	0	0	0	0

Fig. S23 | RNA replication rate constants in the theoretical model (k_{ij}). RNA_i replication is catalyzed by RNA_j in each compartment. The detail was described in Methods.

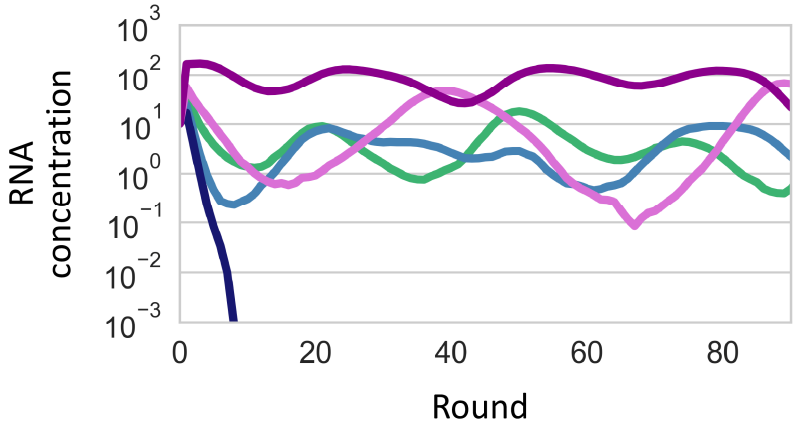


Fig. S24 | Dynamics of the RNA replicator network using the extended model. The simulation was performed and displayed as that presented in Fig. 5b.

Table S1 | Number of analyzed reads obtained by PacBio sequencing.

Experiment	Spieces	Round	Read number
Main	Host	13	4143
Main	Host	24	605
Main	Host	33	718
Main	Host	39	365
Main	Host	43	484
Main	Host	50	1020
Main	Host	53	1358
Main	Host	60	1409
Main	Host	65	3097
Main	Host	72	1364
Main	Host	86	2135
Main	Host	91	637
Main	Host	94	2058
Main	Host	99	855
Main	Host	104	2535
Main	Host	110	1758
Main	Host	115	4003
Main	Host	120	1202
Main	Host	124	1253
Main	Host	129	1879
Main	Host	134	841
Main	Host	139	2395
Main	Host	144	2329
Main	Host	149	1091
Main	Host	155	4998
Main	Host	158	10000
Main	Host	171	10000
Main	Host	179	10000
Main	Host	182	10000
Main	Host	190	10000
Main	Host	205	10000
Main	Host	215	10000
Main	Host	217	10000
Main	Host	228	9999
Main	Host	237	10000
E2	Host	92	5882
E2	Host	114	7840
E2	Host	129	7259
E2	Host	144	9056
E2	Host	160	8477
E2	Host	173	8415
E2	Host	183	10000
E2	Host	189	4270
E2	Host	195	10000
E2	Host	200	9288
E2	Host	220	7209
E2	Host	239	10000

Experiment	Spieces	Round	Read number
E3	Host	86	5393
E3	Host	94	7149
E3	Host	105	10000
E3	Host	135	10000
E3	Host	155	10000
E3	Host	175	10000
E3	Host	198	8910
E3	Host	219	5187
E3	Host	237	10000
Main	Parasite	115	1753
Main	Parasite	120	680
Main	Parasite	124	2073
Main	Parasite	129	1979
Main	Parasite	134	1102
Main	Parasite	139	1254
Main	Parasite	144	1099
Main	Parasite	149	758
Main	Parasite	155	840
Main	Parasite	158	10000
Main	Parasite	179	10000
Main	Parasite	182	10000
Main	Parasite	190	10000
Main	Parasite	205	9999
Main	Parasite	215	10000
Main	Parasite	217	10000
Main	Parasite	228	10000
Main	Parasite	237	10000
E2	Parasite	114	5435
E2	Parasite	129	8077
E2	Parasite	144	9533
E2	Parasite	160	7292
E2	Parasite	173	8999
E2	Parasite	183	7403
E2	Parasite	189	4711
E2	Parasite	195	4483
E2	Parasite	200	5548
E2	Parasite	220	7232
E2	Parasite	239	8118
E3	Parasite	135	10000
E3	Parasite	155	10000
E3	Parasite	175	10000
E3	Parasite	198	9856
E3	Parasite	219	4587
E3	Parasite	237	10000

Grey regions indicate reads obtained in the previous study[3].

Table S2 | Methods for the construction of each plasmid.

Evolved RNA clones encoded in plasmids	Methods of plasmid construction
HL1-120	Site-specific mutagenesis of the plasmid encoding Host-99 in the previous study[3]
HL2-120	Obtained in the previous study[3]
PL2-120	Gene synthesis service of Eurofins Genomics
HL1-158	Site-specific mutagenesis of the plasmid encoding Host-99 in the previous study[3]
HL2-155	Site-specific mutagenesis of the plasmid encoding HL3-155
HL3-155	Gene synthesis service of Eurofins Genomics
PL2-155	Site-specific mutagenesis of the plasmid encoding PL2-120
HL1-228	Site-specific mutagenesis of the plasmid encoding a randomly cloned RNA at round 190*
HL2-228	Site-specific mutagenesis of the plasmid encoding HL3-155
HL3-228	Site-specific mutagenesis of the plasmid encoding HL3-155
PL2-228	Site-specific mutagenesis of the plasmid encoding PL2-155
PL3-228	Site-specific mutagenesis of the plasmid encoding PL2-155

*Cloning was performed by using SMARTer® RACE 5'/3' Kit (Takara) according to the manufacture's protocol.

*Cloning was performed by using SMARTer® RACE 5'/3' Kit (Takara) according to the manufacture's protocol.

Table S3 | The list of primers (from 5' end to 3' end).

Primers for the measurement of host RNA concentrations in the long-term replication experiments by quantitative RT-PCR

Names	Sequences	Description
Primer 1	CAAGTATCGTAAGTTGCTGCC	Used in the main long-term replication experiment
Primer 2	CCGTAATCACCGGTACGTAC	Used in the main long-term replication experiment
Primer 3	GCTGCCTAAACAGCTGCAAC	Used in the additional long-term replication experiments (E2, E3)
Primer 4	CGCTCTTGGTCCCTTGTATG	Used in the additional long-term replication experiments (E2, E3)

Primers for RT-PCR to prepare cDNA libraries for sequence analysis

Names	Sequences	Description (if any)
Primer 5	CCCGGAAGGGGGGGACGAGG	
Primer 6	GGGGGGTCACCTCGCGCAG	
Primer 7	ACAACCCGAACAACAGCAC	Used instead of Primer 5 for host RNAs at round 120, 124, 130, 134 in the main long-term replication experiment
Primer 8	GGGTCACCTCGCGCAGC	Used instead of Primer 6 for host and parasitic RNAs at round 120–155 samples in the main long-term replication experiment

Sequence specific primers for the detection of each RNA clone by quantitative RT-PCR

Names	Sequences	Target RNA clones
Primer 9	ATACACATGGCTCGTAGAAAA	HL0-0
Primer 10	GGCGTACACGCTTGCGGAAGT	HL0-0
Primer 11	CGAACGCTCGTCTCTATAGG	HL1-120
Primer 12	GTACACGCTTGCGGAAGC	HL1-120
Primer 13	AAGGTCGCGCCTCTCCA	HL2-120
Primer 14	ATGCTGTCTTAGGCATGTGT	HL2-120
Primer 15	AGGTCTCCGGCTGAATGTG	PL2-120 and PL2-155
Primer 16	CAAGCCTAACATACACGCTTG	PL2-120
Primer 17	TCCGTCCTTCAAGTTTGCGT	HL1-158 and HL1-228
Primer 18	CTTAGGTACGGTAACTGCTTC	HL1-158 and HL1-228
Primer 19	AAGGTCGCGCCTCTCCA	HL2-155 (with HL1-155) and HL2-228
Primer 20	CGCGAAGATGCTGTCTTAGA	HL2-155 (with HL1-155) and HL2-228
Primer 21	TGGGCGAGTCATGTATAC	HL2-155 (in the presence of both HL1-155 and HL3-155)
Primer 22	CACGCTTGCGGAAGT	HL2-155 (in the presence of both HL1-155 and HL3-155)
Primer 23	TACGCGATCGGTTGCGTC	HL2-155 (unless otherwise noted)
Primer 24	TCGGGGCAATAAGAGCTCA	HL2-155 (unless otherwise noted)
Primer 25	TGATATTAGCCCTTTTAATAAAGCG	HL3-155
Primer 26	TCGGGGCAATAAGAGCTCG	HL3-155
Primer 27	CAAGCCTAACAACGCGCTTG	PL2-155
Primer 28	AACAACGAATAACCGTTCA	HL3-228
Primer 29	TGTTCTTAGGTACGGTAACC	HL3-228
Primer 30	TCTAGAAAGTCTCCGGCTGA	PL2-228
Primer 31	GCAGTGACGCAACATATCC	PL2-228
Primer 32	TCTAGAAAGTCTCCGGCTGG	PL3-228
Primer 33	GCAGTGACGCAACATATCT	PL3-228

Primers for the detection of plus and minus RNA strands by quantitative RT-PCR

Names	Sequences	Step, species, plus or minus strand
Primer 34	GCAAGTGACTCAGGATTCGTACCCGTAATCACCGGTACGTAC	RT, HL1-, HL2-, and HL3-228, plus strand
Primer 35	TAAGCGAATGTTGCGAGCACCAAGTATCGTAAGTTGCTGCC	RT, HL1-, HL2-, and HL3-228, minus strand
Primer 1	CAAGTATCGTAAGTTGCTGCC	PCR, HL1-, HL2-, and HL3-228, plus strand
Primer 2	CCGTAATCACCGGTACGTAC	PCR, HL1-, HL2-, and HL3-228, minus strand
Primer 36	GCAAGTGACTCAGGATTCGTAC	PCR, HL1-, HL2-, HL3-, PL2, and PL3-228, plus strand
Primer 37	TAAGCGAATGTTGCGAGCAC	PCR, HL1-, HL2-, HL3-, PL2, and PL3-228, minus strand
Primer 38	GCAAGTGACTCAGGATTCGTACGCAGTGACGCAACATATCC	RT, PL2-228, plus strand
Primer 39	TAAGCGAATGTTGCGAGCACTCTAGAAAGTCTCCGGCTGA	RT, PL2-228, minus strand
Primer 40	GCAAGTGACTCAGGATTCGTACGCAGTGACGCAACATATCT	RT, PL3-228, plus strand
Primer 41	TAAGCGAATGTTGCGAGCACTCTAGAAAGTCTCCGGCTGG	RT, PL3-228, minus strand
Primer 30	TCTAGAAAGTCTCCGGCTGA	PCR, PL2-228, plus strand
Primer 31	GCAGTGACGCAACATATCC	PCR, PL2-228, minus strand
Primer 32	TCTAGAAAGTCTCCGGCTGG	PCR, PL3-228, plus strand
Primer 33	GCAGTGACGCAACATATCT	PCR, PL3-228, minus strand

References

1. Bahadur, R. R. A representation of the joint distribution of responses to n dichotomous items. in *Studies in Item Analysis and Prediction.* (ed. Solomon, H.) 158–168 (1961).
2. Matsuura, T., Kazuta, Y., Aita, T., Adachi, J. & Yomo, T. Quantifying epistatic interactions among the components constituting the protein translation system. *Mol. Syst. Biol.* **5**, 297 (2009).
3. Furubayashi, T. *et al.* Emergence and diversification of a host-parasite RNA ecosystem through Darwinian evolution. *Elife* **9**, e56038 (2020).
4. Gruber, A. R., Lorenz, R., Bernhart, S. H. & Neubo, R. The Vienna RNA Websuite. *Nucleic Acids Res.* **36**, W70–W74 (2008).