Title

Adaptation and diversification of an RNA replication system under initiation- or termination-impaired translational conditions

Authors

Ryo Mizuuchi[a], Norikazu Ichihashi*[a, b], Tetsuya Yomo[a, b]

Affiliations:

[a]Graduate School of Information Science and Technology, Osaka University.

[b]Graduate School of Frontier Biosciences, Osaka University.

Corresponding to:

Norikazu Ichihashi

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

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

Keywords:

adaptation; evolution; in vitro synthetic biology; RNA replication system; translation

Abstract

Adaptation to various environments is a remarkable characteristic of life. Is it limited to extant complex living organisms, or is it possible for a simpler self-replication system? In this study, we address this question using a translation-coupled RNA replication system, which comprises a reconstituted translation system and an RNA genome encoding only a replicase gene. We performed RNA replication reactions under four conditions, wherein different processes of translation were partially inhibited. We found that the replication efficiency increased as the number of rounds of replication increased in all the conditions tested. However, the types of dominant mutations varied depending on the condition, indicating that this simple system adapted to different environments in different ways. This suggests that even a primitive self-replication system that might have been composed of a small number of genes on the early earth could have had the ability to adapt to various environments.

Body text

The ability of adaptive evolution is one of the remarkable characteristics of living organisms. This ability enables living organisms to survive in various environments and eventually differentiate into distinct species. Extant organisms are composed of a complex network of genes and proteins. Is this complex system required for the ability to adapt or can a simpler system that could have existed before the first self-reproducing cell have this ability? One method to answer this questions is to construct a simple but evolvable self-replication system using a semi-synthetic approach[1] and investigate its ability to adapt.

To date, several types of self-replication systems have been constructed *in vitro*[2][3][4]. When self-replication is repeated for many rounds, the template RNA or DNA autonomously evolves by spontaneously introducing mutations. However, the experiments assessing the evolution of these systems have only been performed under limited conditions[2, 4b, 5].

More recently, we have developed a translation-coupled RNA replication (TcRR) system by introducing translation machinery into Spiegelman's RNA self-replication system[6]. The TcRR system consists of a reconstituted

translation system derived from *Escherichia coli*[7], and an artificial RNA genome (plus RNA, approximately 2 kb) encoding a catalytic subunit of Qβ replicase (RNA-dependent RNA polymerase). In this system, the replicase subunit is translated from the plus RNA and forms the active core replicase with the other subunits (EF-Tu and Ts) included in the translation system. The plus RNA is recognized by the replicase to synthesize the complementary strand (minus RNA), which is also recognized by the replicase for plus RNA synthesis (Figure 1A). We have previously reported that the RNA genome autonomously evolved according to Darwinian principles by repeating TcRR reactions in cell-like compartments[8]. We have also shown that the RNA genome adapted to conditions of reduced ribosome concentration by increasing translation efficiency as a result of introducing several mutations around the ribosome-binding site[9]. However, these are only two case studies. Thus, it remains unknown whether the RNA genome of only 2 kb can adapt to more primitive conditions lacking other translation factors by accumulating environment-specific mutations as observed in living organisms.

In this study, we investigated whether and how the RNA genome adapts to four different conditions, in which different sets of translation factors were omitted (Figure 1B): A) initiation factors 1 (IF1) and 3 (IF3), B) methionyl-tRNA formyltransferase (MTF), IF1, and IF3, C) initiation factor 2 (IF2), D) release factors 1 (RF1), 2 (RF2), 3 (RF3), and ribosome recycling factor (RRF). The initiation of translation is inhibited in conditions A, B, and C[10], and termination is inhibited in condition D[11]. For comparison, we also show the results of the above adaptation experiment under ribosome-reduced conditions (condition E) performed in our previous study with permission from the American Chemical Society[9].

We started the replication reaction using the homogeneous population of a single RNA genome (N96), an RNA clone obtained after 128 rounds of in vitro evolution in our previous study[8]. We mixed the RNA genome with the different translation systems, omitting each set of translation factors as outlined above, and encapsulated the reaction in a water-in-oil emulsion (Figure 1A). After incubating the TcRR reaction for 4 hours at 37°C, the water droplets were collected by centrifugation and the minus RNA was

amplified by reverse-transcription followed by PCR. The plus RNA was synthesized from the cDNA via *in vitro* transcription, and the reaction was again encapsulated with the respective translation system for the next round of the TcRR reaction. We repeated this cycle for 33–35 rounds.

The average minus RNA concentration from the collected droplets after each TcRR reaction was measured by quantitative PCR after reverse transcription (Figure 2). The average minus RNA concentration increased under all the conditions, but the timing of the increase was different for each condition. For conditions A, B, and D, the minus RNA concentration did not change for approximately 20 rounds and then increased. However, for conditions C and E, minus RNA concentration increased before round 5. This difference suggests that the RNA genome adapted to each condition differently.

We chose eight clones at the final round of each condition and analyzed their sequences. All of the detected mutations are listed in Tables S1–S5. Common mutations that were detected in greater than 50% of the eight clones tested for each condition were defined as "dominant mutations". The number of dominant mutations ranged from 6 to 15 among the conditions, and was lower under the initiation-impaired conditions (A, B, and C) than that under the termination-impaired (D) or the ribosome-reduced conditions (E) (Figure 3). These numbers are smaller than expected from the mutation rate measured in our previous study (20–25 mutations)[8], indicating that some mutations were negatively selected. The number of synonymous mutations, non-synonymous mutations, and mutations in untranslated regions varied among the conditions. All the dominant mutations observed under the different conditions are listed in Figure 4A. Some dominant mutations were commonly detected under all the conditions (colored in green), but some are detected only under specific conditions. For example, mutations C184A and A1603G (colored in orange) were detected under all the initiation-impaired conditions (A, B, and C) but not under the termination-impaired condition (D). Mutations C721T, A825G, A1055G, C1612T, A1729G, and C1978A (colored in blue) were detected only under the termination-impaired condition but not under the initiation-impaired conditions. This condition-dependent pattern of dominant mutations also suggests that the RNA genome adapted differently to each condition.

Interestingly, most of the dominant mutations present under both initiation- and termination-impaired conditions were also present under condition E, in which the ribosome concentrations was reduced and therefore the entire process of translation was inhibited. For example, the common mutations present under the initiation-impaired conditions A, B, C (C184A and A1603G), and most of the mutations present under the termination-impaired condition (C721T, A825G, A1055G, and A1729G) were also found under condition E. These results are consistent with the notion that the type of the dominant mutations depends on the translation step that is impaired.

To visualize the mutual evolutionary distance between the clones, we constructed a phylogenetic tree of all of the analyzed clones (Figure 4B). Clones obtained using the same conditions are represented as circles filled with the same color. All the clones from the termination-impaired condition (D) are located in a distinct branch from those of the initiation-impaired conditions (A, B, and C), and the clones from the initiation-impaired conditions A, B, and C exist more closely, forming mixed branches. The bootstrap value[12] of the center branch (indicated by an arrowhead) was found to be 86 from 1000 replications. This high value supports the reliability of this branch. These results support the idea that the RNA genome populations specifically evolved dependent on the translation processes that were impaired.

Presently, the molecular mechanisms underlying the adaptation to these translation-impaired conditions are unknown. One of the possibilities is the RNA structure changes to facilitate translation initiation[13]. Under the initiation-impaired conditions, the C184A mutation was common. According to the secondary structure prediction, this mutation breaks a base pairing in the Shine-Dalgarno (SD) sequence, and thus could possibly facilitate the requirements of ribosome to compensate the impaired initiation[13b]. In termination, the RNA sequence following the stop codon has been reported to affect termination efficiency[14]; however, this is not the case in our study. To the best of our knowledge, the effect of RNA structure on termination has not been demonstrated. Further studies focusing on these mutations might reveal a new termination mechanism that does not depend on release factors.

Notably, the dominant mutations may contain those facilitate experimental procedures other than the TcRR reaction, such as reverse transcription, PCR, and in vitro transcription, and also may change the interaction with the detergents on the droplet surface. The existence of such mutations, if any, does not affect our conclusions or the condition-specific evolution of the RNA genome, because such mutations would be commonly observed under all conditions.

An important insight obtained in this study is that the roles of initiation and termination factors can be compensated for, at least partially, by changes in RNA genome sequence, although the mechanism is presently unknown. The final RNA genomes obtained in this study were able to replicate even with insufficient amount of translation initiation and termination factors. Such a compensatory mechanism for translation initiation and termination might be a remnant of the ancient protein translation system that existed at the transition from the RNA world to the protein world.

In this study, we demonstrated that the artificial RNA genome in the TcRR system evolved differently according to the different conditions tested. This result indicates that an artificial self-replication system, which has only one gene and is much simpler than living organisms, can possess a certain level of adaptation ability. This result implies that primitive life forms may have possessed adaptation ability, which might have played a role in surviving the severe environments on the early earth.

Experimental Section

Reconstituted translation system

Proteins were purified and mixed as described in our previous study[8] with the following exceptions: 1 μM of ribosome, 63 nM *E. coli* HrpA, 1.56 mg/mL tRNA mix (Roche) and 0 μM *E. coli* TrxC. To prepare the translation-impaired conditions described in Figure 1B, the indicated proteins were omitted from the translation system.

Cycle of TcRR reaction

The detailed method was described in our previous study[8-9]. The starting RNA (0.1 nM), the R128 clone in the previous study[1] labeled with GTP-αS,

was encapsulated with each reconstituted translation system in a water-in-oil emulsion, the water droplets of which are approximately 2 μm in diameter. The emulsion was prepared by mixing an aqueous solution containing the TcRR reaction with saturated oil and filtering the mixture through a multi-pore hydrophilic membrane (20 μm, SPG techno, Japan). To prepare the saturated oil, we mixed the oil phase (95% mineral oil, 2% Span 80, and 3% Tween 80) with saturation buffer containing all components of the TcRR system except for RNAs and proteins and 6-fold more dithiothreitol, and obtained the supernatant after centrifugation at 20,000 ×g for 5 min.. After a 4-hour incubation at 37°C for the TcRR reaction, the water droplets were collected by centrifugation. The recovered fraction was treated with nine volumes of an iodine solution containing 10 mM Tris-HCl (pH 7.4) and 1–2 mM iodine for 5 min at 37°C to degrade the initial plus RNA genome labeled with GTP-αS, and then the reaction was stopped by adding 10 mM dithiothreitol. The minus RNA genome was purified and amplified by PCR after reverse transcription. The cDNA was converted to RNA by *in vitro* transcription in the presence of 1 mM GTP-αS, except for the 1–12 cycles performed under condition D. The 0.1 nM plus RNA genome was encapsulated with each reconstituted translation system again for the next round of TcRR reactions. The concentration of minus RNA was measured in every round by quantitative PCR after reverse transcription as described previously[8].

Sequence analysis

The RNA genomes were cloned as described previously[8]. Eight clones were randomly obtained from the final round of each condition and their sequences were analyzed.

Phylogenetic analysis

The alignment of the sequences and the construction of the phylogenetic tree were conducted using MEGA 5.2 software[15]. The sequences of the selected clones were aligned by ClustalW[16]. The phylogenetic tree was constructed with the original RNA genome and all the clones obtained in this study (Table S1-S4) and our previous study (ribosome-reduced condition, Table S5)[9].

Acknowledgments

We thank N. Kamimura (Miki), H. Komai, R. Otsuki, T. Sakamoto, Y. Fujii, and E. Furushima for technical assistance. This work was partly supported by JSPS KAKENHI Grant Numbers, 15H04407 and 15KT0080.

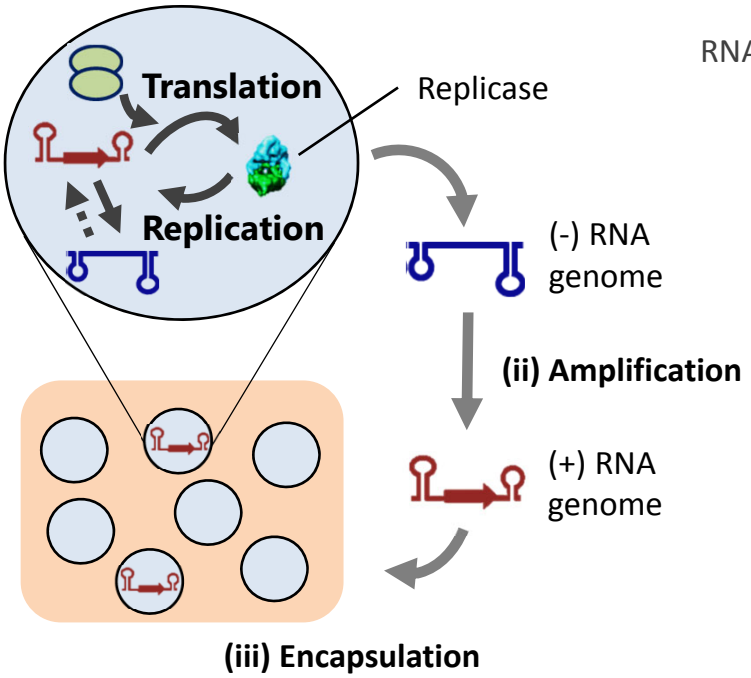
References

[1] P. L. Luisi, F. Ferri, P. Stano, *Naturwissenschaften* **2006**, *93*(1), 1-13.

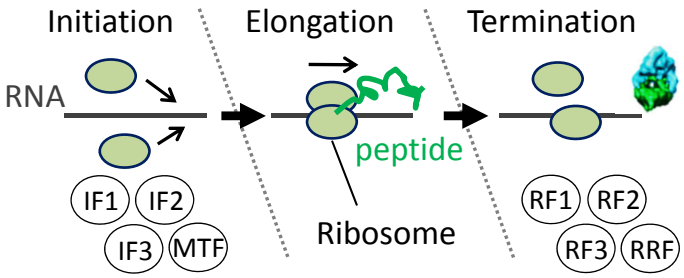
[2] D. R. Mills, R. L. Peterson, S. Spiegelman, *Proc Natl Acad Sci U S A* **1967**, *58*(1), 217-224.

A

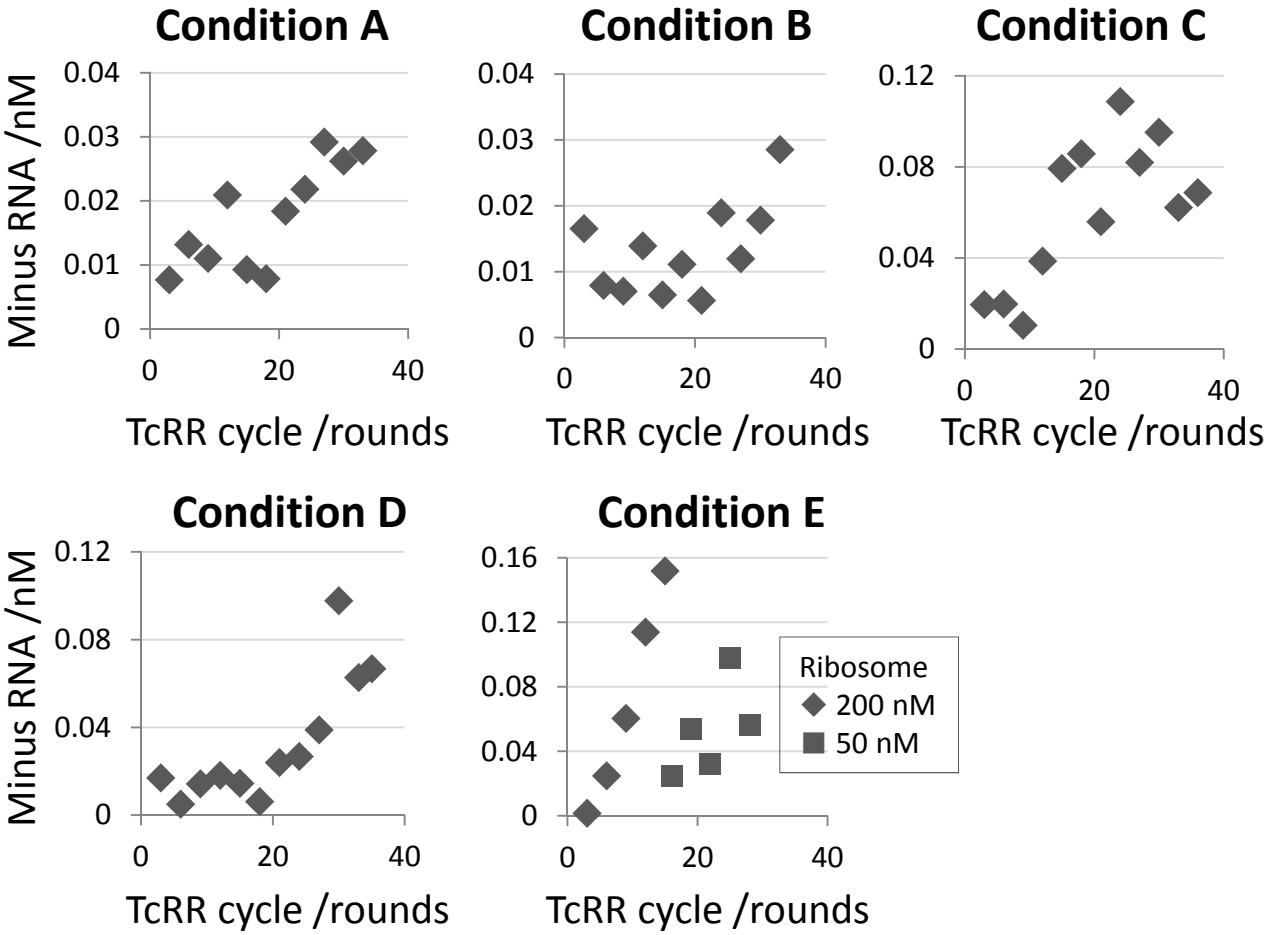
(i) TcRR reaction through impaired translation

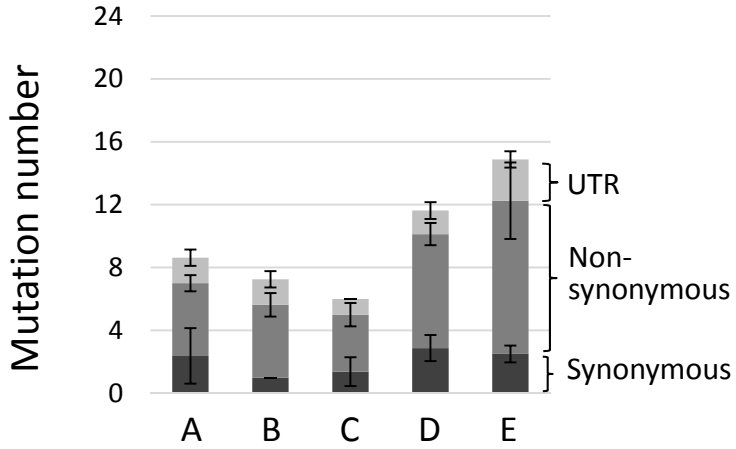


B



	Condition	Impaired step
A	no IF1, IF3	Initiation
B	no IF1, IF3, MTF	Initiation
C	no IF2	Initiation
D	no RF1, RF2, RF3, RRF	Termination
E	low ribosome	All the steps


Condition A
Condition B
Condition C
Condition D
Condition E
Minus RNA /nM
TcRR cycle /rounds
0.04
0.03
0.02
0.01
0
0.12
0.08
0.04
0.16
0.12
0.08
0.04
0
0
20
40
Ribosome
200 nM
50 nM


Mutation number
24
20
16
12
8
4
0
A
B
C
D
E
UTR
Non-
synonymous
Synonymous

A

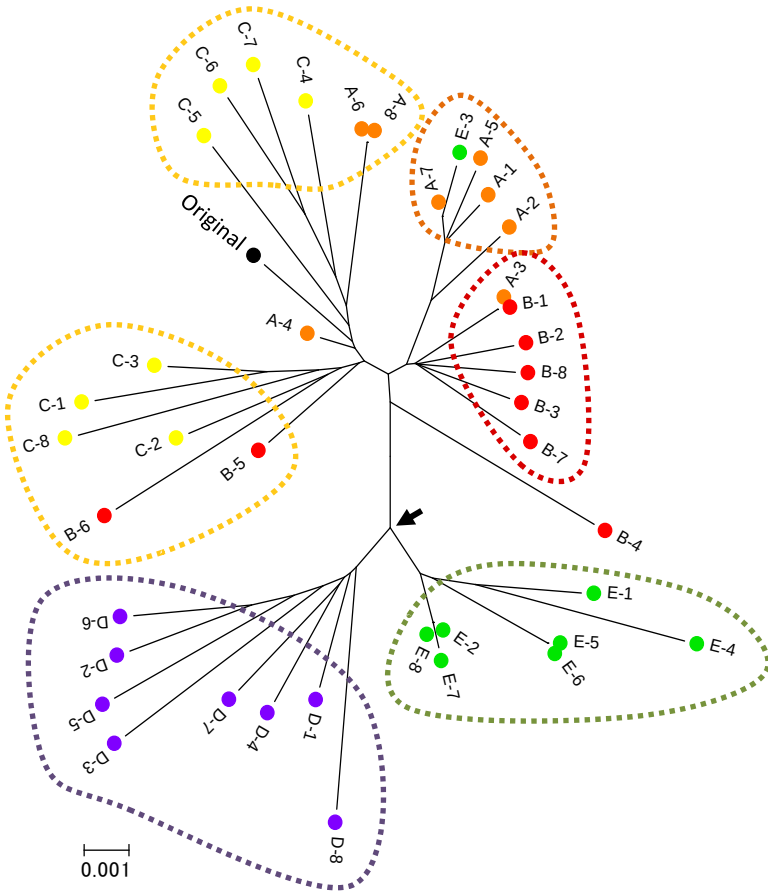
		A	B	C	D	E
C182T	5'UTR					
C184A	5'UTR					
A355T	Q43L					
T435C	Y70H					
G660A	V145I					
C721T	S165L					
C750T	L175F					
A825G	I200V					
T845C	F206F					
T1010A	A261A					
A1055G	S276S					
T1229C	L334L					
C1282A	S352Y					
A1603G	Q459R					
C1612T	T462I					
A1729G	Q501R					
C1738A	A504D					
G1794A	D523N					
T1796C	D523D					
A1809G	R528G					
G1825A	C533Y					
A1855G	Q543R					
T1871C	S548S					
C1973A	3'UTR					
C1978A	3'UTR					

Frequency

8

0

B



Supporting Information

Adaptation and diversification of an RNA replication system under initiation- or termination-impaired translational conditions

Ryo Mizuuchi, Norikazu Ichihashi, Tetsuya Yomo]

Table of Contents

Table S1. List of mutations of the selected 8 RNA clones in the condition A.

		A-							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
G171A	5'UTR			+					
C184A	5'UTR	+	+	+	+	+	+	+	+
A212G	5'UTR	+							
G356A	Q43Q	+				+		+	
G394A	S56N						+		+
T435C	Y70H			+					
A524G	A99A		+						
T566C	P113P						+		+
584-5 insertion	-	+							
T616C	M130T						+		+
G638A	K137K					+			
T662G	V145V	+				+		+	
C709T	T161M			+					
T731C	H168H				+				
T845C	F206F	+	+			+		+	
T953C	D242D					+			
T1010A	A261A	+	+			+		+	
A1015G	E263G	+							
A1040G	A271A							+	
G1100A	L291L						+		+
T1229C	L334L	+	+			+		+	
C1282A	S352Y	+	+	+	+	+	+	+	+
T1484A	P419P						+		+
A1603G	Q459R		+						
C1738A	A504D	+	+	+	+	+	+	+	+
G1772A	S515S		+						
G1774A	R516H		+						
A1788G	S521G						+		+
G1794A	D523N	+	+	+		+		+	
T1824C	C533R			+					
G1825A	C533Y	+	+	+	+	+	+	+	+
T1835C	A536A						+		+
A1855G	Q543R	+	+	+	+	+	+	+	+
T1871C	S548S	+	+	+		+	+	+	+
G1957A	3'UTR						+		+
C1961T	3'UTR					+			
C1973A	3'UTR	+	+	+		+		+	

Table S2. List of mutations of the selected 8 RNA clones in the condition B.

		B-							
		1	2	3	4	5	6	7	8
A167G	5'UTR						+		
G171A	5'UTR	+							
C182T	5'UTR				+				
C184A	5'UTR	+	+	+		+	+	+	+
G279A	A18T					+			
T302C	A25A						+		
G303A	E26K						+		
C323T	S32S								+
C339T	L38L					+			
A355T	Q43L				+				
T365C	F46F								+
C368A	N47K								+
T390C	F55L				+				
T435C	Y70H	+							
A511G	K95R						+		
G663A	E146K						+		
C709T	T161M	+							
C750T	L175F						+		
A791T	L188F						+		
A809G	T194T			+					
T848C	N207N			+					
A863G	V212V				+				
T896C	A223A				+				
C1002T	R259C							+	
T1046C	V273V				+				
1104 deletion	-								+
C1149A	R308R					+			
G1165A	S313N				+				
C1282A	S352Y	+	+	+	+	+		+	+
A1346G	E373E						+		
A1346T	E373D								+
C1402T	P392L						+		
C1439T	G404G				+				
C1460T	Y411Y					+			
G1533A	D436N			+					
A1603G	Q459R						+		
T1613A	T462T			+					

		B-							
		1	2	3	4	5	6	7	8
C1639A	A471D							+	
A1700G	V491V							+	
A1710G	T495A		+						
C1738A	A504D	+	+	+	+	+	+	+	+
G1751A	S508S				+				
G1794A	D523N	+	+	+	+			+	+
T1796C	D523D					+			
G1799A	G524G						+		
A1817G	P530P		+						
T1824C	C533R	+							
G1825A	C533Y	+	+	+	+	+	+	+	+
G1851A	D542N							+	
A1855G	Q543R	+	+	+	+	+	+	+	+
T1871C	S548S	+	+	+	+	+	+	+	+
G1923A	A566T		+						
G1942A	C572Y							+	
T1964C	3'UTR		+						
C1973A	3'UTR	+	+	+	+			+	+
T1981C	3'UTR				+				

Table S3. List of mutations of the selected 8 RNA clones in the condition C.

		C-							
		1	2	3	4	5	6	7	8
A167G	5'UTR				+				
C184A	5'UTR	+	+	+	+	+	+	+	+
201 deletion	5'UTR	+		+					
202 deletion	5'UTR	+							
C227T	5'UTR						+		
G264A	A13T			+					
G300A	A25T					+			
T302C	A25A				+				
G420T	D65Y		+						
A494G	E89E					+			
A508G	E94G						+		
T566C	P113P						+	+	+
A615G	M130V	+							
T616C	M130T						+	+	
630 deletion	-			+					
G660A	V145I					+			
T662G	V145V	+		+					
C680T	H151H					+			
T728C	S167S	+		+					
A749G	A174A	+							
G767A	T180T							+	
A798C	R191R								+
T845C	F206F								+
A849G	K208E						+		
A851G	K208K		+						
A863G	V212V							+	
T962G	R245R	+							
C1039T	A271V	+							
T1046C	V273V								+
T1049C	D274D		+			+			
T1076C	S283S		+						
C1092T	L289F							+	
T1095C	L290L								+
G1100A	L291L				+		+	+	
T1124C	L299L								+

		C-							
		1	2	3	4	5	6	7	8
G1167A	V314I				+		+	+	
C1178T	Y317Y					+			
T1238C	A337A		+						
A1270G	D348G				+		+	+	
C1277T	D350D								+
C1282A	S352Y	+		+					
1458-9 (insertion)	insertion					+			
T1501G	V425G			+					
T1521C	W432R				+				
C1538T	G437G						+		
A1558G	H444R						+		
A1603G	Q459R	+	+	+	+			+	+
C1617T	P464S		+						
C1622T	D465D	+							
G1652A	S475S	+							+
G1706T	V493V			+					
C1738A	A504D	+	+	+	+	+	+	+	+
G1751A	S508S					+	+	+	
T1796C	D523D	+		+				+	+
C1815A	P530T								+
G1825A	C533Y	+	+	+	+	+	+	+	+
A1855G	Q543R	+	+	+	+	+	+	+	+
G1870A	S548N				+	+			
T1871C	S548S	+	+	+			+	+	+
G1957A	3'UTR				+	+			
G1960A	3'UTR								+
A1962G	3'UTR	+							
T1964A	3'UTR			+				+	

Table S4. List of mutations of the selected 8 RNA clones in the condition D.

		D-							
		1	2	3	4	5	6	7	8
T175C	5'UTR				+				
C182T	5'UTR	+	+	+	+	+	+	+	+
A211G	5'UTR								+
A212G	5'UTR		+						
C256T	S10F			+					
A304G	E26G						+		
G344A	L39L					+			
A355T	Q43L	+	+	+	+	+	+	+	+
T377C	A50A								+
A609G	I128V			+					
C614T	H129H						+		
T652C	V142A	+							
G660A	V145I	+	+	+	+	+	+	+	+
G663A	E146K	+							
C721T	S165L	+	+	+	+	+	+	+	+
C725T	Y166Y					+			
C750T	L175F	+	+	+		+			+
T770G	P181P								+
814-5 insertion	-							+	
A825G	I200V	+	+	+	+	+	+	+	+
T845C	F206F	+		+	+	+	+	+	+
A854G	A209A				+	+		+	
C886A	R220H						+		
T957C	L244L						+		
A1012G	H262R				+				
C1013T	H262H			+					+
G1017T	G264C				+				
C1052T	L275L					+			
A1055G	S276S	+	+		+	+	+	+	
G1073A	M282I					+			
G1100A	L291L		+						
A1151G	R308R					+			
A1211C	T328T			+					
C1282A	S352Y					+			
A1301G	G358G			+					
G1319A	P364P		+						
A1384G	K386R							+	

		D-							
		1	2	3	4	5	6	7	8
C1427T	H400H		+						
A1461G	I412V								+
G1537A	G437D							+	
A1582G	K452R							+	
C1612T	T462I		+	+	+	+	+		
A1649G	G474G		+			+	+		
A1729G	Q501R								+
C1738A	A504D	+		+	+			+	+
T1801C	L525S			+					
G1825A	C533Y	+	+	+	+	+	+	+	+
G1845A	A540T	+							
A1855G	Q543L	+	+	+	+	+	+	+	+
C1862T	I545M			+					
T1871C	S548S	+			+	+		+	+
C1875T	P550S					+			
A1881G	K552E							+	
G1957A	3'UTR								+
C1973A	3'UTR	+			+				
C1978A	3'UTR		+	+			+	+	
T1981C	3'UTR			+					

Table S5. List of mutations of the selected 8 RNA clones in the condition E.

		E-							
		1	2	3	4	5	6	7	8
C182T	5'UTR	+	+		+			+	+
C184A	5'UTR	+	+	+	+	+	+	+	+
C225T	5'UTR					+	+		
C263T	S12S				+				
A355T	Q43L	+	+		+	+	+	+	+
G356A	Q43Q			+					
T371C	S48S	+							
T435C	Y70H	+	+		+	+	+	+	+
G660A	V145I	+	+		+	+	+	+	+
T662G	V145V			+					
C721T	S165L	+	+		+	+	+	+	+
C821T	I198I				+				
A825G	I200V	+	+		+	+	+	+	+
T845C	F206F	+	+	+	+	+	+	+	+
T1010A	A261A			+					
G1023A	V266I					+	+		
A1040G	A271A			+					
A1055G	S276S	+			+	+	+		
G1158A	D311N				+				
C1174T	T316I				+				
C1220T	L331L							+	
T1229C	L334L			+	+				
C1282A	S352Y	+	+	+	+	+	+	+	+
C1298T	Y357Y					+	+		
G1539A	V438I	+							
C1578T	R451C				+	+	+		
A1603G	Q459R		+					+	+
A1719G	T498A	+							
A1729G	Q501R		+		+	+	+	+	+
C1738A	A504D	+	+	+	+	+	+	+	+
G1761T	D512Y			+					
C1794A	S519S	+							
A1791G	N522D				+				
G1794A	D523N			+					
A1809G	R528G		+			+	+	+	+
G1825A	C533Y	+	+	+	+	+	+	+	+
A1855G	Q543R	+	+	+	+	+	+	+	+
T1871C	S548S	+	+	+	+	+	+	+	+

		E-							
		1	2	3	4	5	6	7	8
A1890C	R555R			+					
C1907T	F560F	+			+				
G1960A	3'UTR	+			+				
C1973A	3'UTR	+	+	+	+	+	+	+	+

1. Ichihashi, N., Usui, K., Kazuta, Y., et al. (2013) Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment. *Nat. Commun.*, **4**, 1–7.

2. Mizuuchi, R., Ichihashi, N., Usui, K., et al. (2015) Adaptive Evolution of an Artificial RNA Genome to a Reduced Ribosome Environment. *ACS Synth. Biol.*, **4**, 292–298.