PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY OPEN ACCESS

Pervasive RNA folding is crucial for narnavirus genome maintenance

Makiha Fukuda[a,b], Jitong Cai[c], Joel S. Bader[c], and Jef D. Boeke[a,b,1]

Contributed by Jef D. Boeke; received March 10, 2023; accepted May 18, 2023; reviewed by Nobuhiro Suzuki and Reed B. Wickner

A synthetic biology approach toward constructing an RNA-based genome expands our understanding of living things and opens avenues for technological advancement. For the precise design of an artificial RNA replicon either from scratch or based on a natural RNA replicon, understanding structure–function relationships of RNA sequences is critical. However, our knowledge remains limited to a few particular structural elements intensively studied so far. Here, we conducted a series of site-directed mutagenesis studies of yeast narnaviruses ScNV20S and ScNV23S, perhaps the simplest natural autonomous RNA replicons, to identify RNA elements required for maintenance and replication. RNA structure disruption corresponding to various portions of the entire narnavirus genome suggests that pervasive RNA folding, in addition to the precise secondary structure of genome termini, is essential for maintenance of the RNA replicon in vivo. Computational RNA structure analyses suggest that this scenario likely applies to other "narna-like" viruses. This finding implies selective pressure on these simplest autonomous natural RNA replicons to fold into a unique structure that acquires both thermodynamic and biological stability. We propose the importance of pervasive RNA folding for the design of RNA replicons that could serve as a platform for in vivo continuous evolution as well as an interesting model to study the origin of life.

RNA genome | RNA virus | narnavirus | RNA structure | RNA maintenance

Significance

Artificial RNA replicons could serve as a platform to study the origin of life as well as to promote new technological advancement. To engineer an RNA replicon, understanding the structure–function relationship of the sequence is critical. To gain a deeper insight into it, we focused on the virus family *Narnaviridae* that have a simple single-stranded RNA genome. By using a combined approach of computational RNA structure analysis and in vivo examination of each structural element, we found that not only typical terminal RNA structures but also pervasive internal RNA folding of the replicon is critical for the maintenance of the RNA in vivo. This finding implies that taking account of pervasive folding status is important for RNA replicon design.

Most organisms evolved to use DNA as the carrier of genetic information, while only RNA viruses use RNA. RNA viruses comprise a large and extremely diverse group. Among them, *Narnavirus* is a genus of positive-strand RNA viruses in the family *Narnaviridae*. Their genome represents one of the simplest structures among the viruses, consisting of only a single open-reading frame (ORF) encoding RNA-dependent RNA polymerase (RdRP) along with exceedingly short 5′ and 3′ untranslated regions (UTR) (1). In addition, some of them have extremely diverse genome structures and coding capacities. For example, multipartite narnaviruses encode the RdRP motifs in two separate RNA segments (2–6), while ambisense narnaviruses contain an additional uninterrupted ORF in the reverse direction overlapping the RdRP ORF, which spans almost the entire length of the virus genome (7–9).

Saccharomyces 20S RNA narnavirus (ScNV20S) and Saccharomyces 23S RNA narnavirus (ScNV23S) represent the two best-studied narnaviruses. ScNV20S was originally discovered as an RNA species accumulating under conditions promoting sporulation (10, 11) and both accumulate also under heat shock (12, 13), and were subsequently characterized by cloning and sequencing (14–16). Both industrial and wild isolates of *Saccharomyces cerevisiae* often carry ScNV20S, while fewer strains also contain ScNV23S (17–19). The molecular and biochemical characterization of those two viruses showed that the viral RNA is not encapsidated into a conventional virion (20, 21) and lacks an extracellular transmission pathway, but rather, it is transmitted along with host cell division from mother to daughter cells, or horizontally through mating. The naked viral RNA forms a ribonucleoprotein complex with its own RdRP in a 1:1 stoichiometry that resides and replicates stably in the host cytoplasm (22). Although the two viruses are evolutionarily close among narnaviruses, they share only 25% identity in amino acid sequence (Fig. 1*A* and *SI Appendix*, Fig. S1). Despite their ability to cohabit a single *S. cerevisiae* cell, they are independent and compatible, in that the RdRP translated from ScNV20S RNA only interacts with and replicates its own RNA but not ScNV23S RNA, and vice versa (21, 23).

A launching system for narnaviruses using yeast DNA expression plasmids has been established (24, 25) (Fig. 2*E*). In this system, the vector contains the full-length cDNA of ScNV20S or ScNV23S under the constitutive *PGK1* promoter, while the ribozyme sequence from hepatitis delta virus (HDV) is fused directly to the 3′ end of the cDNA so that the resultant viral RNA has the precise 3′-end in vivo through transcription and subsequent self-cleavage, which is essential for the replication activity of the viral RNA. In this launch system, the precise 5′ terminus is generated by a host exoribonuclease

Author contributions: M.F., J.S.B., and J.D.B. designed research; M.F. and J.C. performed research; M.F. and J.D.B. contributed new reagents/analytic tools; M.F., J.C., J.S.B., and J.D.B. analyzed data; and M.F. and J.D.B. wrote the paper.

Reviewers: N.S., IPSR and Okayama University; and R.B.W., NIH.

Competing interest statement: J.D.B. is a Founder and Director of CDI Labs, Inc., a Founder of and consultant to Neochromosome, Inc, a Founder, SAB member of, and consultant to ReOpen Diagnostics, LLC, and serves or served on the Scientific Advisory Board of the following: Logomix, Inc., Modern Meadow, Inc., Rome Therapeutics, Inc., Sample6, Inc., Sangamo, Inc., Tessera Therapeutics, Inc. and the Wyss Institute. J.S.B. is a Founder of Neochromosome, Inc, and a consultant to Opentrons Labworks, Inc, and Reflexion Pharmaceuticals, Inc.

[1]To whom correspondence may be addressed. Email: jef.boeke@nyulangone.org.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2304082120/-/DCSupplemental.

Published June 20, 2023.

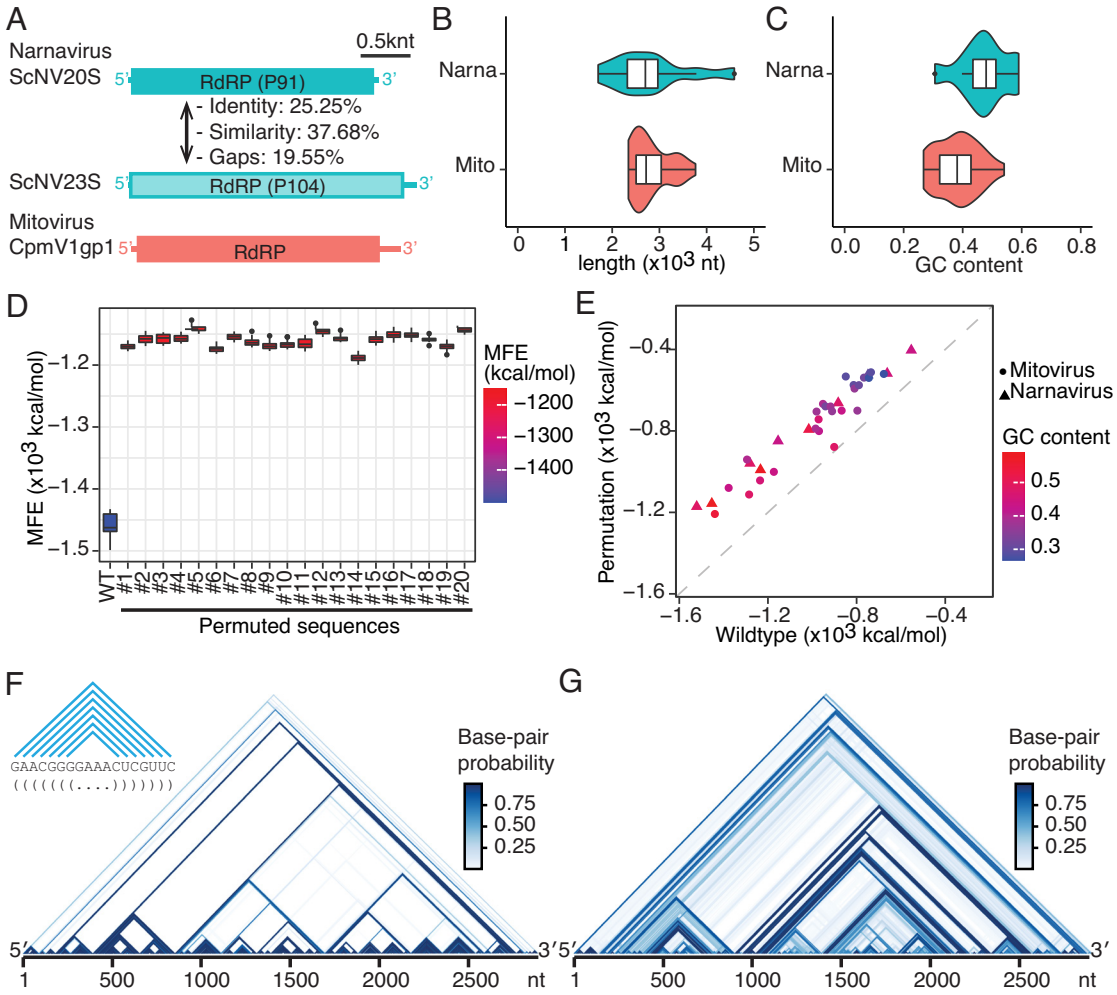

Fig. 1. In silico analysis of RNA secondary structure. (*A*) Representative genome structure of narnavirus (ScNV20S, NC_004051.1 and ScNV23S, NC_004050.1) and mitovirus (CpmV1gp1, NC_004046.1). Boxes represent RdRP coding sequence; lines represent UTRs. Protein similarity between the two narnavirus species was calculated by the Needleman–Wunsch algorithm. (*B* and *C*) Distribution of genome size (*B*) and GC content (*C*) of narnaviruses and mitoviruses registered in the NCBI database. (*D*) Distribution of minimum folding energies of ScNV23S RNA and the corresponding 20 different random permutations as calculated by mfold. Each random sequence has the same GC content and nucleotide length as ScNV23S. (*E*) Correlation between minimum folding energy of wild-type viral genome sequences (*x* axis) and the corresponding random permutations (*y* axis). For the energy of random permutations, the mean values of 20 different permuted sequences were plotted for each narnavirus or mitovirus species. The dotted diagonal represents y = x. (*F* and *G*) Contact map of predicted ScNV23S RNA structure (*F*) and one of the 20 permutated sequences that was chosen at random (*G*). Each base pair predicted by mfold is represented by a line as shown in an enlarged diagram at the left. Line color represents base-pair probability.

activity such as SKI1/XRN1 (26). Both ScNV20S and ScNV23S can be launched efficiently from such vectors and continue replicating independently of the launch vector which may then be readily eliminated through launcher plasmid loss. The launching system enabled identification of some crucial *cis*-acting sequences and structures in the viral RNA such as the 5 nt inverted repeats at their termini (5′-GGGGC…GCCCC-OH) and the first stem structure from the 3′ end; however, the exploration of functional elements has been limited to only a small part of the genome since there is no high-throughput narnavirus launching and detection pipeline.

To expand the search for functional sequence elements for maintenance and replication of the viral RNA, we improved the throughput of narnavirus launching system and performed site-directed mutagenesis against the entire narnavirus genome. Combining the results obtained with the above systems with in silico RNA structure analysis, we demonstrate the importance of pervasive RNA folding for capsidless RNA genomes to maintain their persistency. Given that the activity of RNA self-replication is the most essential requirement for the RNA-based genome, and RdRP is the protein responsible for the process, one can consider narnavirus genomes as molecular fossils from an early RNA–protein world where proteins emerged in the RNA world to replace RNA as the main biological catalysts (28). Our observations may provide insights into the properties of functional genomic RNAs and could help understand how the RNA–protein world initially emerged.

Results

In Silico Analysis of RNA Secondary Structure of *Mitovirus* and *Narnavirus* Genomic RNAs. *Mitovirus* and *Narnavirus* are two genera that belong to the positive-strand RNA virus families *Mitoviridae* and *Narnaviridae,* respectively (https://ictv.global/taxonomy/). The complete genomes for 25 mitoviruses and 10 narnaviruses are currently registered in the NCBI Virus database (Dataset S1). Those two genera share a similar genomic structure, that is, the viral RNA contains only a single ORF that covers most of the genome and encodes an RdRP which catalyzes virus replication (Fig. 1 *A* and *B*). Therefore, they are considered the simplest genomic RNA. There are two major differences between those two families. First, mitoviruses replicate in the mitochondria, whereas narnaviruses replicate in the cytosol in the host cell.

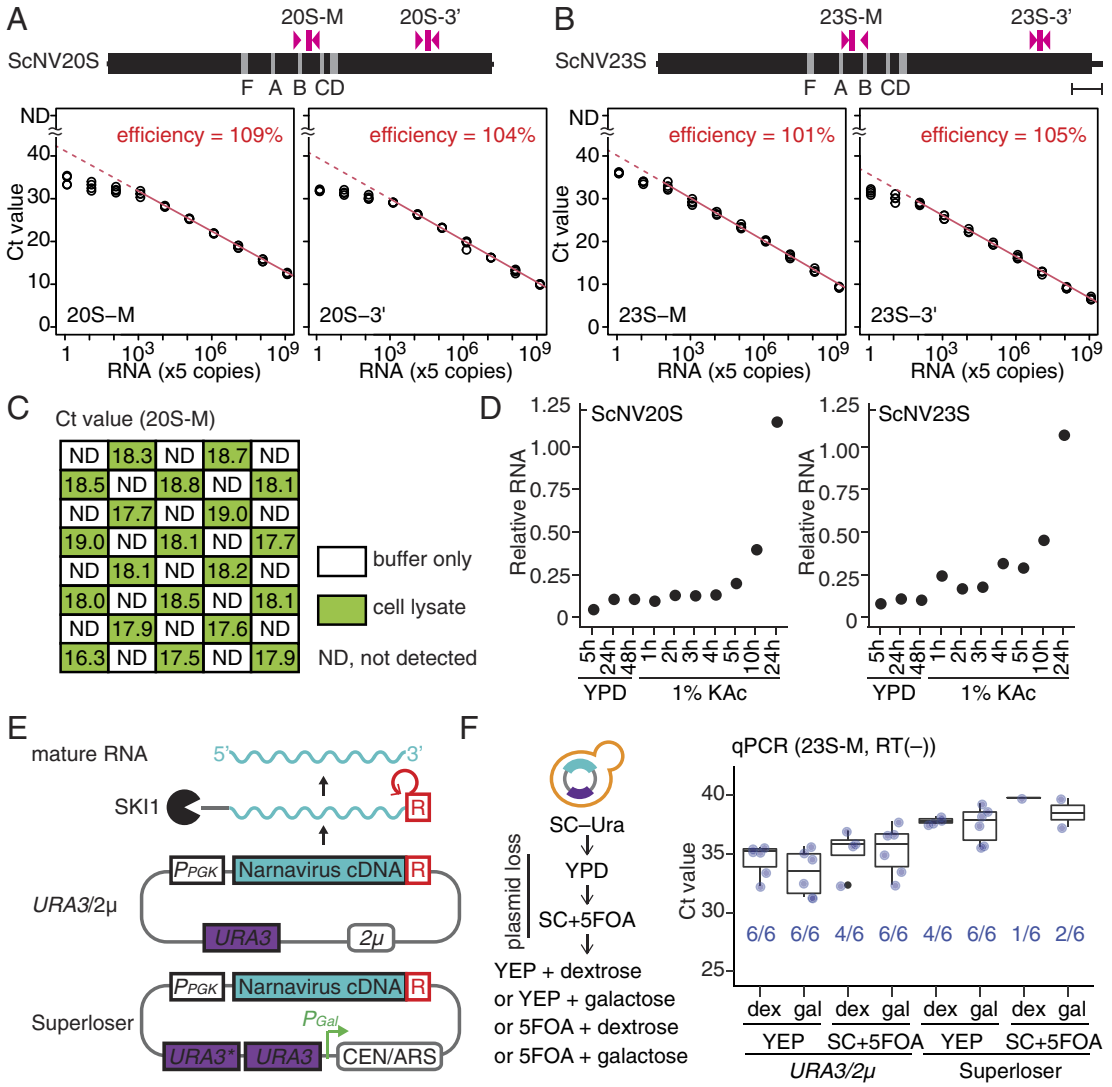

Fig. 2. Narnavirus launching and detection platform. (*A* and *B*) Quantification of ScNV20S (*A*) and ScNV23S (*B*) RNA by RT-qPCR. *Upper*, a diagram of the genome structure. Black line, black box, and gray boxes indicate UTRs, ORFs, and conserved protein motifs (*A*–*F*), respectively (2). Scale, 200 nt. The positions of primers and probes are indicated by magenta arrowheads and squares, respectively. *Lower*, standard curves of RT-qPCR. A serial dilution of in vitro transcribed DNase-I-treated narnavirus RNA was used as a template. Each dot represents the Ct value of one of four technical replicates. The amplification efficiency was calculated based on the linear regression analysis (red solid line). ND, not detected in the water control. (*C*) Minimal cross-contamination during 96-well nucleic acid extraction. The checkerboard layout of samples for nucleic acid extraction in a 96-well plate is shown, while the Ct result of RT-qPCR using the 20S-M primer/probe set is shown in each box. ND, not detected. (*D*) Expression profiles of ScNV20S (*Left*) and ScNV23S (*Right*) before and after induction of replication in potassium acetate (KAc) medium. 25S rRNA was used as the internal control transcript. (*E*) A diagram of three different designs of the narnavirus launching vector. R, the sequence of 82-bp hepatitis delta virus antigenomic ribozyme. *URA3**, codon-modified orthogonal *URA3* gene (27). (*F*) qPCR detection of the remaining ScNV23S launching vector after 5-FOA selection. A diagram of the cell culture condition is shown on the *Left*. Cell extract was prepared from transformed yeast cells (BY4741 Δ*ski2*) with *URA3*/2μ or Superloser-based ScNV23S launching vector. The DNA fraction of the resultant nucleotide samples was detected by qPCR without adding reverse transcriptase. Blue dots represent six independent colonies; blue letters show detection rate.

Second, the GC content of mitovirus genomes is substantially lower than that of narnavirus genomes (Fig. 1*C*).

To address the importance of the thermodynamic stability of viral RNA, we performed in silico structure analyses of *Narnaviridae*. We calculated the minimum folding energy (MFE) of their genomic RNA sequences using mfold software (29), while as a comparison, 20 random permutations with the same length and GC content were generated for each genomic sequence and then subjected to the same calculation (Fig. 1*D*). This analysis revealed that all narnavirus and mitovirus RNAs, regardless of GC content, exhibited significantly lower folding energy than that of the corresponding random permutations, implying a strong evolutionary bias toward folding into a thermodynamically stable conformation (Fig. 1*E*). These properties are not observed in mRNAs or noncoding RNAs (*SI Appendix*, Fig. S4); thus, the observed sequence property seems unique to this class of viral RNAs. Furthermore, the wild-type sequence shows many fewer varieties of structural conformation with higher base-pair probability at each nucleotide interaction position than that of the corresponding random permutation (Fig. 1 *F* and *G*). These observations are consistent with evolutionary pressure on the viral RNAs to fold into a unique and thermodynamically stable conformation.

A High-Throughput Narnavirus Launching System. To address relationships between RNA structure and function in viral RNA in vivo, we attempted to conduct various sequence modifications of ScNV20S and ScNV23S. The existing launching system for those two viruses (24, 25) (Fig. 2*E*) enabled us to produce replicative viral RNAs in vivo; however, there remain limitations

in scalability that prevent a systematic exploration of sequence–structure–functional relationships. To improve the throughput of the launching system, we first established a one-step RT-qPCR method for narnavirus detection as an alternative to gel-based methods such as northern blotting. Two different primer–probe sets were designed for each of ScNV20S and ScNV23S RNA, one of which targets the middle of the RNA, while the other targets the 3′ region (Fig. 2 *A* and *B*). The amplification of narnavirus cDNA by each primer set was evaluated by using 5 to 5×10^9 copies of in vitro transcribed and DNaseI-treated viral RNA as a template. As a result, at least five copies of template RNA were readily detected by those primer–probe sets, while a linear correlation between RNA copy number and signal intensity was observed in a range between $5{\cdot}10^3$ and $5{\cdot}10^9$ copies for ScNV20S, and between $5{\cdot}10^2$ and $5{\cdot}10^9$ copies for ScNV23S (Fig. 2 *A* and *B*). Second, in order to improve the throughput of nucleotide extraction from the yeast cells, we employed an automated liquid handler. Briefly, yeast cells are mechanically disrupted using glass beads in a 96-well format, and then the resultant supernatant is subjected to the combined use of a DNA/RNA extraction kit and CyBio Felix liquid handler. To test the possibility of cross-contamination during this process, ScNV20S-positive and -negative samples were applied in a checkerboard layout and subjected to nucleic acid extraction. Subsequent RT-qPCR analysis showed that no amplification from the ScNV20S negative wells was observed, confirming that the automated sample preparation does not cause false-positive signals due to cross-contamination (Fig. 2*C*). Using the improved narnavirus detection pipeline, we analyzed the induction of narnavirus replication. Yeast strains carrying one of the two narnaviruses were inoculated in YPD for 48 h followed by 1% potassium acetate (KAc), known to induce narnavirus replication. Cells were collected at each time point indicated in Fig. 2*D* and then relative RNA levels were measured. The result showed that narnavirus RNA accumulated significantly only after 10h or longer incubation in acetate induction medium, consistent with previous reports (10, 11).

Next, to further improve the scalability of the procedure, we modified the original launching vector. We first switched the *TRP1* gene, which catalyzes tryptophan biosynthesis, used as a selective marker in the original launching vector (25) with the counterselectable *URA3* gene (Fig. 2*E*). This modification allows us to use 5-fluoroorotic acid (5-FOA) to efficiently select narnavirus-launched but plasmid-cured cells which can be identified by a sequential replica plating (Fig. 2*F*). To assess the efficacy of 5-FOA counterselection in plasmid removal, we performed qPCR analysis to detect the residual launching vector by using yeast cells transformed with *URA3*/2μ-based ScNV23S launching vector. Transformed cells were selected by SC–Ura solid medium, and then sequential replica plating was performed using YPD medium to induce plasmid loss followed by SC+5FOA medium to select cells that lost the launching vector (Fig. 2*F*). After the 5-FOA selection, cells derived from a single colony were grown in YEP-based liquid media for qPCR analysis. Unexpectedly, qPCR detected residual plasmid DNA from all six independent papillae, suggesting that replica-plating onto 5-FOA was insufficient to completely eliminate the vector (Fig. 2*F*, 1st plot). The insufficient removal of the launching vector is a critical issue for narnavirus detection because it could increase the false-positive rate. To resolve this problem, an extra round of selection by 5-FOA selection was added, that is, SC+5FOA liquid medium instead of YEP was used for cell culture. This modification reduced the plasmid detection rate to 4/6 (Fig. 2*F*, 3rd plot). To further induce plasmid loss, we tested the Superloser vector, which is designed to reduce spontaneous 5-FOA resistance by introducing two copies of the *URA3* gene. This vector also contains the CEN/ARS sequence fused to the *GAL1-10* promoter to promote plasmid segregation in the presence of the sugar galactose (27). When launching ScNV23S from the Superloser vector, relatively higher Ct values around 37 were detected, suggesting that the copy number of the residual launching vector was reduced by ~fourfold compared with using the *URA3*/2μ vector (Fig. 2*F*). Furthermore, use of SC+5FOA medium significantly reduced the background plasmid detection rate (Fig. 2*F*, 7th plot). However, changing the sugar source from dextrose to galactose, which is supposed to induce the degradation of plasmid segregation, did not further enhance plasmid loss (Fig. 2*F*, 6th and 8th plots). Based on all of these results, we concluded that using the Superloser vector for viral RNA launching reduces the spontaneous 5-FOA resistance and the additional 5-FOA selection step further improves the performance.

Recoding the Narnavirus Genome. Using the narnavirus launching system developed above, we attempted to examine the impact of massive as well as partial structural disruption of the viral RNA on its function (Fig. 3 *A* and *B*). To achieve this, we synonymously recoded, modifying codons of nucleotide region 282 to 1,945 encoding 555 out of 942 amino acids in the ScNV23S ORF using three different algorithms using the web-based tool GeneDesign (30, 31): the "optimized" algorithm that replaces each codon with the optimal codon for expression in *S. cerevisiae*, the "most different" algorithm that replaces each codon with the most optimal codon that is most dissimilar to the original codon, and the "least different" RSCU (relative synonymous codon usage) algorithm that seeks to replace as many codons as possible while minimizing disruption of the original average RSCU value for the sequence. Those sequence modifications do not change protein sequence but do change the RNA structure, and therefore, the stability and template activity of the viral RNA could be affected (Fig. 3*B*). To separately evaluate those effects, we quantified RNA level by two different procedures (Fig. 3*A*). Briefly, for RNA stability, we measured the RNA level transcribed from the launch vector by isolating RNA from yeast cells before 5-FOA selection, while for replication activity, we quantified the RNA level after eliminating the launching vector followed by inducing the replication by 1% KAc (Fig. 3*A*). We designated them "transcribed RNA" and "replicated RNA," respectively, however, considering that once the viral RNA is launched, the virus continues to replicate even without induction; therefore, the fraction of "transcribed RNA" may contain a certain amount of replicated RNA, and therefore, the measured RNA level might be an overestimate. The resultant RT-qPCR analysis showed that even before 5-FOA selection, all the modified ScNV23S variants reduced RNA level by at least 50%, and as a result, the replicating RNA level was drastically reduced to almost the limit of detection in every case (Fig. 3*C*). A similar analysis was applied to ScNV20S in which the RNA region 301 to 1,630 of the genome was codon-optimized to yeast codon usage (Fig. 3*D*). As in the case of ScNV23S, codon-modified ScNV20S also significantly reduced RNA stability as well as replicated RNA level (Fig. 3*E*). Furthermore, when we divided the recoded region into three segments to recode at a 500-nucleotide resolution (Fig. 3*D*), none of the six codon-modified ScNV20S variants retained RNA stability (Fig. 3*E*), implying that there is no single specifically required RNA structure. Rather, the results suggest a pervasive RNA folding requirement throughout the RNA that may be important for maintaining the viral RNA stability in vivo.

To challenge this hypothesis, we tested the effect of codon modification at the individual structural element level on viral RNA

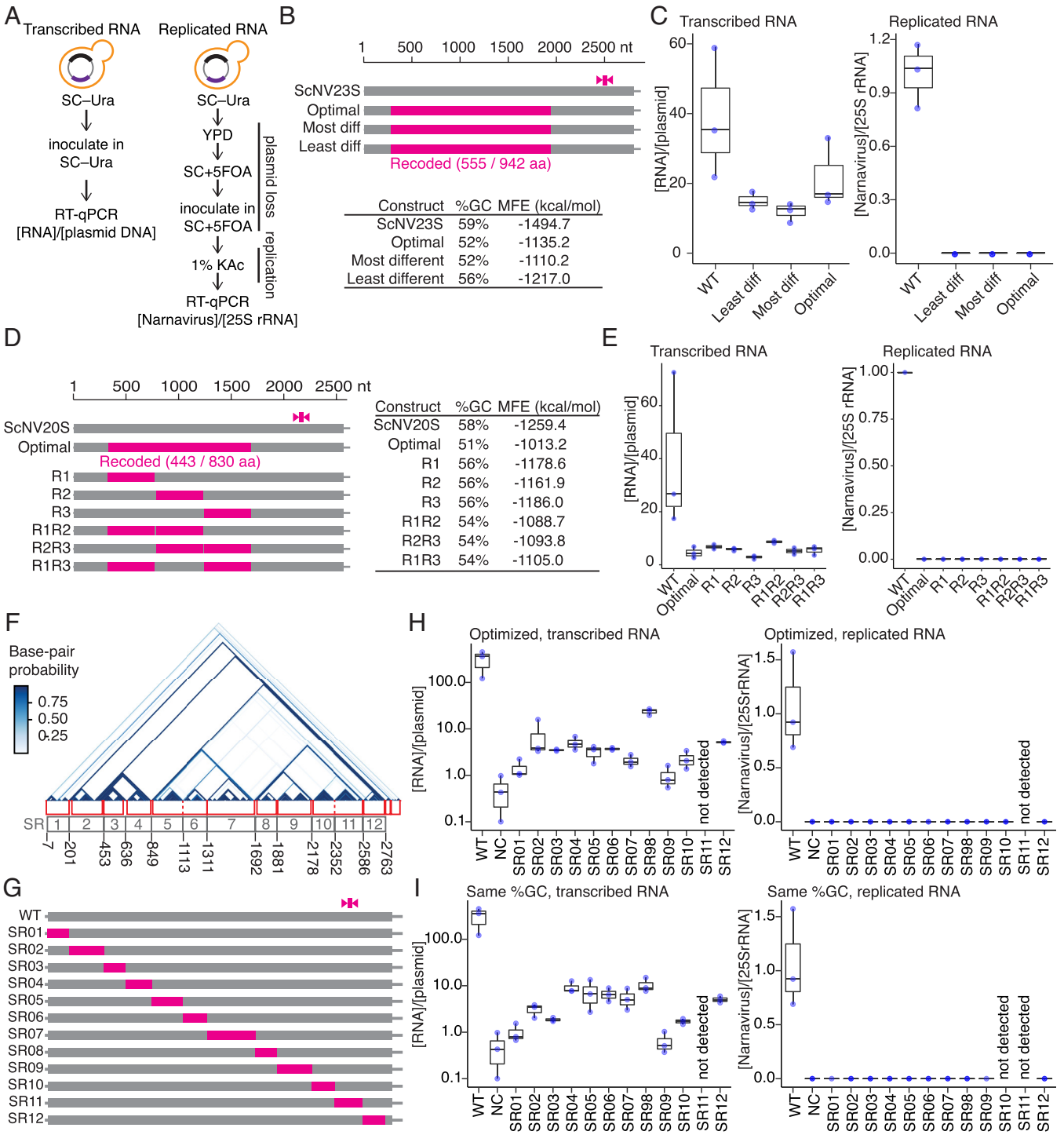

Fig. 3. Effect of recoding of narnavirus RdRP ORF. (*A*) Scheme of sample preparation for analyzing RNA transcribed from the plasmid (*Left*) and replicated RNA (*Right*). (*B*) Construct designs of recoded ScNV23S and the corresponding sequence information. Magenta boxes indicate the regions for pervasive codon modifications; gray boxes indicate wild-type nucleotide sequence. Arrowheads/square indicates positions targeted by qPCR primer/probe. (*C*) RT-qPCR analysis of recoded ScNV23S. Three independent colonies were tested (blue dots). (*D*) Construct designs of recoded ScNV20S and the corresponding sequence information. (*E*) RT-qPCR analysis of recoded *ScNC20S*. Three independent colonies were tested (blue dots). (*F*) Defining structural elements in the ScNV23S genome. The predicted RNA secondary structure of ScNV23S is visualized as a contact map, in which each base pair predicted by mfold is represented by a line colored according to the base-pair probability. The large and small triangles refer to possible long- and short-distance base-paring regions, respectively. Twelve structural elements were defined and indicated by red rectangles. Two of them were subdivided into two substructures (red dotted lines). (*G*) Construct designs of recoding ScNV23S at a structural element level. Recoded regions are indicated by magenta boxes. (*H*) RT-qPCR analysis of ScNV23S containing a structural element codon-optimized to *S. cerevisiae* (*H*) or codon-modified without changing the original GC content (*I*). *Left*, transcribed RNA; *Right*, replicated RNA. Three independent colonies were analyzed for each construct and are indicated by blue dots.

stability. Our computational prediction of RNA secondary structure defined 15 structural elements in the ScNV23S genome, of which 12 elements were found inside the CDS (SR01-12, Fig. 3*F* and *SI Appendix*, Fig. S2). Next, 24 constructs were designed in which one of the structural elements was codon-optimized by the following two different algorithms: i) codon-optimized to yeast codon usage and ii) codon-modified without changing the total GC content (Fig. 3*G*). All of the ScNV23S variants decreased

their transcript level to 10% or less and showed no evidence for replication (Fig. 3 *H* and *I*).

3′ Structural Elements in the ScNV23S Genome. As in the case of many RNA viruses, some *cis*-acting sequences for narnavirus replication were previously identified in the 3′ terminal region of ScNV20S and ScNV23S such as the 5-nt inverted repeats at their termini (5′-GGGGC…GCCCC-OH) and the first stem structure from the 3′ end (24, 25, 32, 33). However, the possible involvement of adjacent structural elements upstream of the known *cis*-elements at the 3′ end remains unexplored. In the ScNV23S genome, there are four stem-loop structures predicted in 160-nt from the 3′ end, of which the third hairpin loop structure from the 3′ end contains the stop codon (Fig. 4*A*). We first evaluate the importance of stem structures by site-directed mutagenesis. Adding a mutation to either side of the stem structure decreased the RNA levels for stems 2 and 3, but when sequences on both sides in those stems were exchanged simultaneously, thus restoring the stem formation potential, the RNA level was at least partially rescued, although only to less than half of the wild-type sequence abundance (Fig. 4*B*). This indicates that not only the 1st stem structure but also the upstream three stem structures seem to be essential for RNA maintenance. Note that it was not possible to design a restored 4th stem structure because the structure is entirely located within the CDS which restricts the availability of nucleotide substitutions without changing the protein sequence. Interestingly, the measured RNA level of each tested sequence is highly correlated with the predicted folding energy except for stem 3 mutants which might be too short to predict MFE precisely (*SI Appendix*, Fig. S5), indicating a biophysical link between their thermodynamical and biological stability as well as the potential use of RNA structure prediction as a reliable indicator to estimate viral RNA stability.

Next, we examined the importance of loop structures on RNA maintenance. Each of the four loop structures in the ScNV23S 3′ region was replaced with the NheI recognition sequence, GCUAGC, and RNA level was quantified before and after 5-FOA selection. The RNA level before eliminating the plasmid was not

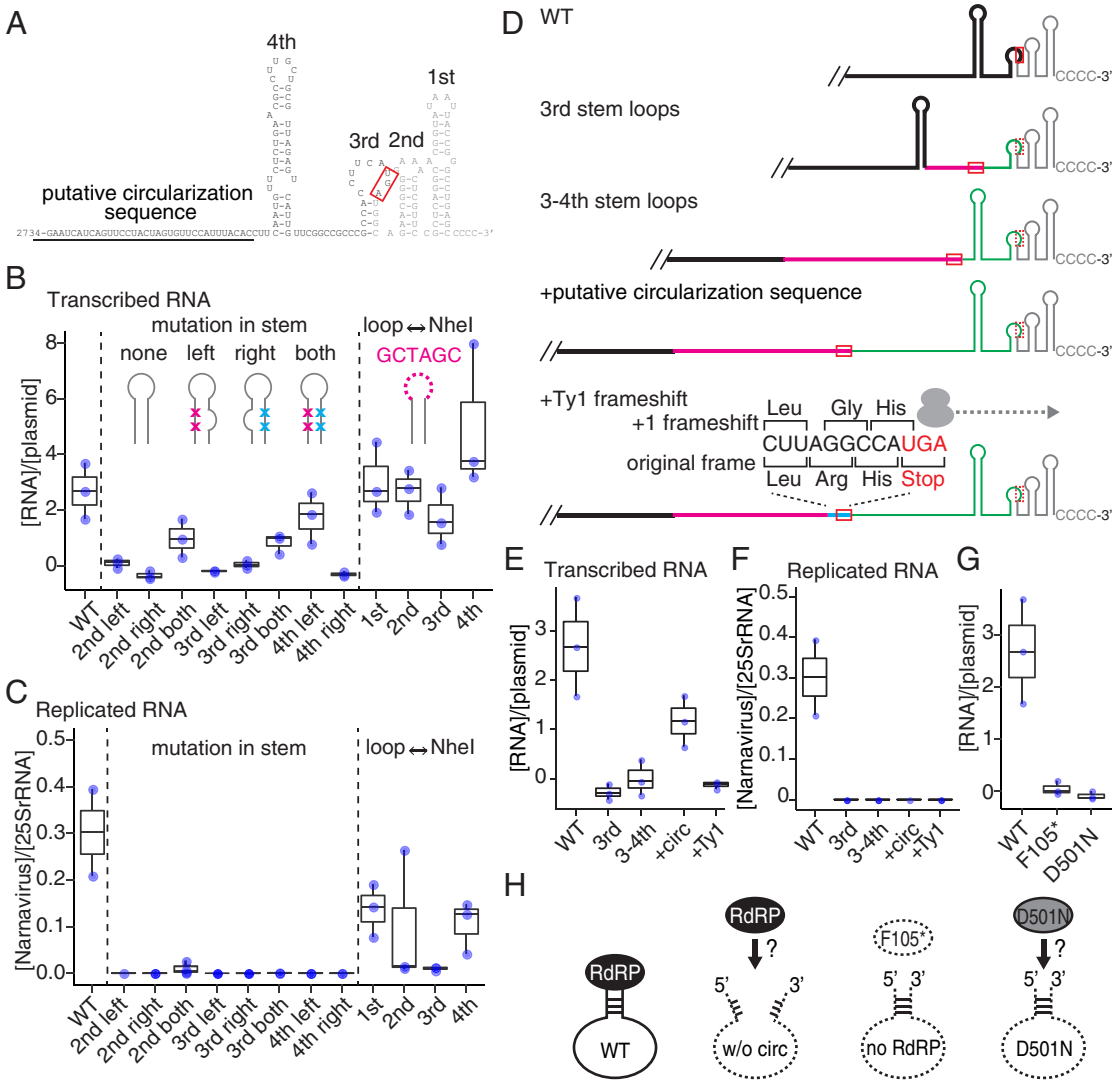

Fig. 4. Structural element dissection in ScNV23S 3′ region. (*A*) Schematic diagram of predicted secondary structure of 158 nucleotides from the 3′ end of ScNV23S. Underline, 37-nt sequence predicted to interact with the 5′ region of the genome (*SI Appendix*, Fig. S3). Red box, stop codon. Gray letters, 3′ UTR. (*B* and *C*) RT-qPCR analysis of a series of ScNV23S mutagenesis targeting one of the structural elements in the 3′-region. (*D*) Diagrams of refactored ScNV23S. Red box, stop codon; red dotted box, position of original stop codon not recognized as such in this context; green, 3′ UTR; magenta, codon-optimized to *S. cerevisiae*; gray, 3′ UTR. (*E* and *F*) RT-qPCR analysis of ScNV23S containing a refactored element. (*G*) Possible involvement of RdRP in stabilizing RNA genome before 5-FOA selection. RT-qPCR analysis of ScNV23S carrying a D501N catalytic mutation or an early stop codon inserted at F105. Three independent colonies were analyzed in all the RT-qPCR experiments (blue dots). (*H*) A hypothetical model of ScNV23S maintenance. The solid line indicates stable RNA; dashed lines represent unstable/degraded RNA. We hypothesize that i) lack of circularization, ii) lack of RdRP or iii) RdRP with weakened binding would lead to instability/degradation of the RNA.

significantly changed by those modifications except in the case of the 3rd loop disruption (Fig. 4*B*), implying no/smaller contribution of those loop structures on the RNA maintenance. However, examining the effect on RNA replication (i.e., measuring RNA abundance after 5-FOA selection, Fig. 4*C*) suggested a more severe impact of the loop substitutions, with loop mutants 1, 2, and 4 showing only 1/3rd the RNA abundance, and no replication of loop mutant 3. As for the 3rd loop modification, because the length of the stem is shorter than that of the others, the stem might be more easily destabilized by the loop modification.

Based on the site-directed mutagenesis, at least the four stem structures from the 3′ end seem to be essential for viral RNA function. To determine whether the stretch of those four elements serves as a sufficient *cis*-acting sequence, we attempted to disentangle the ORF/translation product from those stem loops partially or wholly included in the ORF, namely hairpins 3 and 4. We thus "refactored" the ScNV23S genome as shown in Fig. 4*D* (34). Briefly, to examine hairpins 3 and 4 for possible function, their sequence was recoded within the ORF C terminus to disrupt the RNA structures, whereas the corresponding native sequences of hairpins 3 and 4 were added downstream of the stop codon. Thus, the modified sequence retains all the four target RNA structures in the 3′ UTR while maintaining the protein coding sequence. Refactoring the 3rd and 4th stem-loop structures in this way reduced the RNA level significantly even before eliminating the plasmid. However, adding a 37-nt upstream sequence, which is predicted to interact with the 5′ end of the genome and thus designated as putative circularization sequence (*SI Appendix*, Fig. S3 *A–C*), restored the RNA level to around 50%, although this recovery was insufficient to support viral RNA replication (Fig. 4 *E* and *F*). We expected that further adding a ribosomal +1 frameshift sequence from the Ty1 retrotransposon just upstream of the stop codon, which is predicted to allow ribosome readthrough into hairpins 3 and 4, might help the viral RNA to fold into a stable conformation through translation-coupled folding; to the contrary, however, this modification abolished the observed recovery of transcript stability seen in the + putative circularization sequence (Fig. 4 *D* and *E*). We note that the failure of this construct to replicate/stabilize RNA might be simply due to the changes made to the amino acid sequence, making it difficult to interpret this negative result. Collectively, the results suggest that intramolecular interaction throughout this region is at least partially responsible for viral RNA stability and potentially that termination of translation of RdRP must occur in a way that positions it right at the 3′ end of the RNA. Given that the RdRP seemingly binds to both ends of the viral RNA (35), this binding may further stabilize a (noncovalently linked) circular stable RNA conformation and therefore contribute to its persistence. Indeed, when replacing Phe105 with a stop codon (F105Stop) in the ScNV23S genome to abolish RdRP production, the RNA level drastically decreased (Fig. 4*G*). Surprisingly, when adding a presumably catalytic-dead mutation at the conserved GDD motif among viral RdRPs (D501N), the RNA level was also decreased (Fig. 4*G*). This mutation in the RdRP is similar to one known to lose enzymatic activity but sustain template binding in hepatitis C virus (36). These results suggest that binding by the cognate RdRP is essential but may not be sufficient for maintaining the persistence of viral RNA in the host cell.

Discussion

We conducted a variety of site-directed mutagenesis studies on the two narnavirus genomes, ScNV20S and ScNV23S, ranging from limiting synonymous recoding to a single RNA structural element to recoding a large part of the genome. We also improved the existing pipeline for narnavirus launching and detection by modifying the launching vector and applying a liquid handler and RT-qPCR to achieve higher throughput in analyzing in vivo transcription and replication of the viruses. Our method for measuring viral RNA levels can expand our survey toward identifying host genetic and environmental factors involved in viral RNA maintenance and replication. In addition, the computational prediction of RNA secondary structure and the recoding scheme of each structural element enabled us to design a series of modified narnavirus genomes at a single structural element resolution.

Both ScNV20S and ScNV23S RNAs are unprotected by viral particles (20, 22, 25), but, instead, reside in the host cytoplasm as "naked" ribonucleoprotein complexes with their cognate RdRPs (22). Given that they lack 3′-poly(A) tails and perhaps lack 5′ cap structures (37) and thus may resemble intermediates in mRNA decay, how they can persist in the host cytoplasm without their genomes being degraded by the exonucleases involved in mRNA degradation pathways remains an interesting open question (33). Our study showed that recoding any of single structural elements internal to the genome drastically decreased the viral transcript level even in the presence of the launch vector that constitutively generates the RNA. In addition, all the mutants with reduced RNA stability showed no or considerably compromised replication competency. This suggests that even partial disruption of the RNA structure results in an almost complete breakdown of transcript stability in the host cell, thus preventing replication and transmission to other cells. We assume that throughout evolution narnaviruses acquired improved persistence in host cells by optimizing their primary RNA sequence to fold into both thermodynamically and biologically stable structures. This self-protection strategy from degradation might be unique to naked RNA viruses, and perhaps was lost with the evolution of capsid proteins, acquisition of which would relax this remarkably limiting constraint on genome structure, and perhaps thereby greatly accelerating the evolution of viral types.

Indeed, pervasive recoding typically does not lead to extreme sensitivity to RNA degradation in RNA viruses or related elements that encode nucleocapsid and/or capsid proteins. For example, a pervasively recoded polivirus genome retained infectivity (38), although its virulence was diminished, a theme that has now been echoed in a wide variety of RNA viruses (39, 40). While the emphasis of such studies has been on attenuating viral replication for the purpose of generating live vaccines, the fact remains that these massively recoded viruses retain infectivity and by and large replicate surprisingly well. In one dramatic example of this, Yarrington et al demonstrated that recoding around 90% of the Ty1 retrotransposon genome (which is also encapsidated) altered neither transcription nor translation levels (41). In addition, increasing studies have reported successful designs of self-replicating RNA vectors based on various RNA virus genomes for in vivo production of genes of interest (42–44). Folding into a particular stable RNA structure is perhaps the primary strategy that RNA species developed at early stages in genome evolution before "inventing" other generic forms of protection such as nucleocapsid and capsid proteins.

Our observation of reduced transcript level when adding an early stop codon in RdRP (Fig. 4*G*) is consistent with the previous report claiming that "the steady-state levels of 23S RNA transcripts unable to form complexes with p104 were consistently lower than those of transcripts able to bind to the protein" (33). Complex formation with the cognate protein seems to be essential to avoid viral RNA degradation; however, it remains unclear specifically what molecular

factors/states are involved in viral RNA maintenance. Another interesting point of view we found is possible involvement of a long-range RNA–RNA interaction to promote viral RNA stability (Fig. 4*E* and *SI Appendix*, Fig. S3). The formation of circularized RNA through base pairing (rather than a covalent bond) is already known to be essential for efficient replication of some RNA viruses such as polioviruses (45, 46) and flaviviruses (47), and might also be important for "naked" viral RNAs like narnaviruses to protect the genome from host nucleases. RNA circularization of ScNV20S has been observed directly by electron microscopy in the absence of proteins in vitro (48). Additionally, the lowest energy model for a genomic secondary structure predicted by RNAfold predicts an extended base-pairing interaction between sequences near the 5′ and 3′ termini (*SI Appendix*, Fig. S3), supporting that circularization can occur autonomously through long-range base-paring interactions. Formation of ribonucleoprotein complex with the cognate RdRP might assist this conformation as in the case of poliovirus ribonucleoprotein complex in which the RNA termini are circularized by a protein bridge consisting of the viral polymerase and possibly host proteins (45, 46).

In addition, the process of translation might be involved in narnavirus RNA maintenance as in the case of mosquito-infecting Culex Narnavirus 1 (CxNV1), a model species of "ambisense" narnaviruses. Premature truncation of the rORF (reverse ORF) resulted in a substantial loss of CxNV1 RNA, indicating a requirement of the presence of the rORF for RNA stability. The lack of rORF sequence conservation as well as an unusual ribosome footprint pattern in which the footprints are heavily concentrated at specific positions without triplet periodicity suggests that the act of translation is important rather than a specific protein product (7). It remains unknown whether translation-coupled RNA protection is involved in other narnaviruses. Further determination of complex states as well as host factors involved in viral RNA maintenance may allow us to design artificial RNA genomes as a platform for continuous evolution in vivo as well as an ideal model to study the origin of life.

Materials and Methods

Strains. For all the transformation experiments, haploid yeast knockout (YKO) strain derived from BY4741 (49, 50) carrying the *ski2* null mutation, which enhances the abundance of narnavirus RNAs (25, 48) (YKO ORF name YLR398C; *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ski2Δ*::KanMX), was used as the parental strain. For the expression profiles of narnavirus in Fig. 2, narnavirus-launched and plasmid-cured BY4741 strains provided by Dr. Rosa Esteban were used: strain 899 in which ScNV20S is launched through the plasmid pRE763 (26), while strain 961 in which ScNV23S is launched through the plasmid pMR33 (51).

Narnavirus Launching and Induction. Yeast cells transformed with a launching vector were selected on synthetic complete medium without uracil (SC–Ura) for single-colony isolation. The resultant plate was sequentially replica-plated to a YPD, followed by SC medium supplied with 0.1% (w/v) 5-FOA (SC+5FOA), to select for plasmid loss. Papillae grown on an SC+5FOA derived from a single colony were used for the RNA quantification experiments. Narnavirus induction was performed as follows: narnavirus-launched cells were grown in SC+5FOA at 30 °C for 2 d, transferred to the induction medium (1% potassium acetate, pH7.0), and then incubated for another 20 to 24 h at 30 °C to induce narnavirus RNA.

Plasmids and Synthetic DNAs. The *URA3*/2μ-based ScNV20S and ScNV23S launching vectors, pRE763 and pMR33, were provided by Rosa Esteban (26, 51). To construct the Superloser-based launching vectors, we derived the backbone from pDT137 (27). pDT137 was linearized by NotI and SacI for inserting ScNV20S amplified from pRE763 using the primer set oMFB498 and 499, or by EcoRI for inserting ScNV23S amplified from pMR33 using the primer set oMFB336 and 337. The resultant vectors were designated as pMFB186 and pMFB151, respectively. The launching plasmids for ScNV23S 3′-mutagenesis were made from pMFB151 by site-directed in vitro mutagenesis using Gibson assembly (52). The changes introduced into the ScNV23S RNA sequence are indicated in Dataset S6. For the recoding of ScNV20S and ScNV23S in Fig. 3, a gBlock synthetic DNA fragment corresponding to each codon-optimized sequence was cloned into MluI and AleI restriction enzyme sites of ScNV20S launching vector or BssHI and BsiWI sites of ScNV23S launching vector (summarized in Dataset S4). For recoding of ScNV23S at a single structural element resolution shown in Fig. 3, an eBlock synthetic DNA fragment corresponding to each codon-optimized sequence, PCR-amplified ScNV23S 5′- and 3′-fragments, and pMFB151 linearized by BstXI and SacI were assembled in yeast and then confirmed by colony PCR using primer sets targeting the left and right junctions (summarized in Dataset S5).

All plasmids used are listed in Dataset S2, and synthetic DNAs including primers, eBlocks, and gBlocks used are listed in Dataset S3. Synthetic DNAs were purchased from Integrated DNA Technologies (IDT, Coralville, IA).

In Vitro Transcription. ScNV20S and ScNV23S RNA were synthesized using the MEGAscript T7 Transcription kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Plasmid vectors pALI017 and pALI037 (37) linearized by SmaI and purified using ethanol precipitation were used as a template DNA for ScNV20S and ScNV23S RNA synthesis, respectively. In vitro transcribed RNA was incubated with 2U of TURBO DNase (Thermo Fisher Scientific) to digest the plasmid DNA and purified by phenol:chloroform precipitation. RNA concentration was measured by Qubit RNA BR Assay Kits (Thermo Fisher Scientific).

RNA Structure Prediction and Visualization. *Mitovirus* and *Narnavirus* reference genomes were obtained from NCBI Virus database (53) the sequence information is summarized in Dataset S1. Random sequences mentioned in Fig. 1*D* were generated by random shuffling of the wild-type ScNV23S genome, in order to maintain base composition. RNA secondary structure with MFE was predicted by RNAfold software in the ViennaRNA package (2.5.0) using default parameters (54). Base-pairing probability from RNAfold was used to generate the contact map.

RNA Structure Segmentation. In order to segment the whole ScNV23S genome into separate structural regions, we used Mfold-3.6 (http://www.unafold.org/mfold/software/download-mfold.php) to predict the RNA secondary structure using default parameters at a folding temperature of 30 °C. We used Mfold-3.6 instead of RNAfold because Mfold outputs MFE structures together with all suboptimal structures with at most 12.00 kcal/mol energy increment. The ensemble of optimal and suboptimal structures was further considered for structure segmentation. A dot plot was generated for each individual structure. We smoothed the dot plots to allow averaging across all structures. In the averaged dot plot, a filter was used to delineate the base-pairing regions, namely the stems. We only allowed base pairing within a range of 200 nucleotides to consider only local structures. Two base-pairing regions were combined into a structural region if they were adjacent or overlapped. The process was repeated until no further structural regions could be combined.

Codon Optimization. Codon optimization mentioned in Fig. 2 was conducted by a web-based tool GeneDesign (30, 31). Codons were optimized in four styles–optimal, most different, least different, and same GC. The "optimal" style recodes the sequence according to optimized codon usage in *S. cerevisiae*. The "most different" style replaces codons with the most dissimilar optimal codons. The "least different" style replaces as many codons as possible while maintaining the original RSCU score. The "same GC" style kept the GC content same as the original sequence.

Quantification of Narnavirus RNA. Cells were grown in a 96-well deep-well plate (Corning, CLS3961) supplied with a sterile 4-mm glass bead in each well to avoid cell clumping. One-milliliter cell cultures were collected by centrifugation (3,000 × g for 3 min at room temperature) and resuspended in 0.3 mL ice-cold Tris-buffer (50 mM Tris-HCl, pH8.0, 100 mM NaCl). 0.5 mm acid-washed glass beads (Sigma, G8772) were added to the cell suspension and then cells were disrupted by RETSCH MM400 Mixer Mill with 5 cycles of mixing at the top speed (60/s) for 30 s. The resultant supernatant was used for 96-well nucleic acid extraction with the combined use of Quick- –DNA/RNA Viral MagBead Automated Kit (Zymo Research, R2141) and CyBio Felix automated liquid handler (Analytik Jena). Two microliters of total nucleic samples was subjected to qPCR analysis in a 10-μL reaction volume using PrimeScript OneStep RT-qPCR Kit (TaKaRa, RR064). For the quantification of RNA transcribed from a launching plasmid, RT-qPCR was

performed with and without reverse transcriptase. The relative transcript level per plasmid was calculated by using the following formula:

$$\frac{([RNA] + [DNA])}{[RNA]} = 2^{\Delta\Delta Ct\,(RT(+)-RT(-))},$$

$$\frac{[RNA]}{[DNA]} = 2^{\Delta\Delta Ct\,(RT(+)-RT(-))} - 1.$$

Replicated RNA level was quantified by using 25S rRNA as a reference gene.

Data, Materials, and Software Availability. Codes and data for RNA secondary structure analysis are available (55). All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We are grateful to Rosa Esteban and Tsutomu Fujimura for extremely fruitful discussions as well as for sharing narnavirus launching plasmids and related yeast strains. We also thank Nieves Rodríguez-Cousiño for sharing the detailed procedure for narnavirus launching and detection. We thank Kaihang Wang for helpful discussions. J.S.B. and J.D.B. acknowledge funding from the NSF (collaborative grant MCB-1935366 to J.D.B. and collaborative grant MCB-1935355 to J.S.B.) for this project.

Author affiliations: [a]Department of Biochemistry and Molecular Pharmacology, Institute for Systems Genetics, NYU Langone Health, New York, NY 10016; [b]Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY 11201; and [c]Department of Biomedical Engineering, Whiting School of Engineering and School of Medicine, Johns Hopkins University, Baltimore, MD 21205

1. R. Esteban, T. Fujimura, "Narnaviruses (Narnaviridae)" in *Encyclopedia of Virology*, D. H. Bamford, M. Zuckerman, Eds. (Academic Press, Oxford, ed. 4, 2021), pp. 621–626, 10.1016/B978-0-12-809633-8.20941-2.
2. Y. Chiba, S. Oiki, T. Yaguchi, S. I. Urayama, D. Hagiwara, Discovery of divided RdRp sequences and a hitherto unknown genomic complexity in fungal viruses. *Virus. Evol.* **7**, veaa101 (2021).
3. Y. Sato *et al.*, A new tetra-segmented splipalmivirus with divided RdRP domains from Cryphonectria naterciae, a fungus found on chestnut and cork oak trees in Europe. *Virus Res.* **307**, 198606 (2022).
4. J. Jia *et al.*, Interannual dynamics, diversity and evolution of the virome in Sclerotinia sclerotiorum from a single crop field. *Virus. Evol.* **7**, veab032 (2021).
5. S. Sutela *et al.*, The virome from a collection of endomycorrhizal fungi reveals new viral taxa with unprecedented genome organization. *Virus. Evol.* **6**, veaa076 (2020).
6. A. Ruiz-Padilla, J. Rodriguez-Romero, I. Gomez-Cid, D. Pacifico, M. A. Ayllon, Novel mycoviruses discovered in the mycovirome of a necrotrophic fungus. *mBio* **12**, e03705-20 (2021).
7. H. Retallack, K. D. Popova, M. T. Laurie, S. Sunshine, J. L. DeRisi, Persistence of Ambigrammatic narnaviruses requires translation of the reverse open reading frame. *J. Virol.* **95**, e0010921 (2021).
8. S. Cook *et al.*, Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. *PLoS One* **8**, e80720 (2013).
9. A. M. Dinan, N. I. Lukhovitskaya, I. Olendraite, A. E. Firth, A case for a negative-strand coding sequence in a group of positive-sense RNA viruses. *Virus. Evol.* **6**, veaa007 (2020).
10. K. Kadowaki, H. O. Halvorson, Appearance of a new species of ribonucleic acid during sporulation in Saccharomyces cerevisiae. *J. Bacteriol.* **105**, 826–830 (1971).
11. B. Garvik, J. E. Haber, New cytoplasmic genetic element that controls 20S RNA synthesis during sporulation in yeast. *J. Bacteriol.* **134**, 261–269 (1978).
12. M. Wesolowski, R. B. Wickner, Two new double-stranded RNA molecules showing non-mendelian inheritance and heat inducibility in Saccharomyces cerevisiae. *Mol. Cell Biol.* **4**, 181–187 (1984).
13. R. B. Wickner, Double-stranded and single-stranded RNA viruses of Saccharomyces cerevisiae. *Annu. Rev. Microbiol.* **46**, 347–375 (1992).
14. Y. Matsumoto, R. B. Wickner, Yeast 20 S RNA replicon. Replication intermediates and encoded putative RNA polymerase. *J. Biol. Chem.* **266**, 12779–12783 (1991).
15. L. M. Esteban, N. Rodriguez-Cousino, R. Esteban, T double-stranded RNA (dsRNA) sequence reveals that T and W dsRNAs form a new RNA family in Saccharomyces cerevisiae. Identification of 23 S RNA as the single-stranded form of T dsRNA. *J. Biol. Chem.* **267**, 10874–10881 (1992).
16. N. Rodriguez-Cousino, L. M. Esteban, R. Esteban, Molecular cloning and characterization of W double-stranded RNA, a linear molecule present in Saccharomyces cerevisiae. Identification of its single-stranded RNA form as 20 S RNA. *J. Biol. Chem.* **266**, 12772–12778 (1991).
17. V. Lopez, R. Gil, J. Vicente Carbonell, A. Navarro, Occurrence of 20S RNA and 23S RNA replicons in industrial yeast strains and their variation under nutritional stress conditions. *Yeast* **19**, 545–552 (2002).
18. R. B. Wickner, H. K. Edskes, D. Bateman, A. C. Kelly, A. Gorkovskiy, The yeast prions [PSI+] and [URE3] are molecular degenerative diseases *Prion* **5**, 258–262 (2011).
19. A. V. Mardanov, A. V. Beletsky, T. N. Tanashchuk, S. A. Kishkovskaya, N. V. Ravin, A novel narnavirus from a Saccharomyces cerevisiae flor strain. *Arch. Virol.* **165**, 789–791 (2020).
20. W. R. Widner, Y. Matsumoto, R. B. Wickner, Is 20S RNA naked? *Mol. Cell Biol.* **11**, 2905–2908 (1991).
21. M. P. Garcia-Cuellar, L. M. Esteban, T. Fujimura, N. Rodriguez-Cousino, R. Esteban, Yeast viral 20 S RNA is associated with its cognate RNA-dependent RNA polymerase. *J. Biol. Chem.* **270**, 20084–20089 (1995).
22. A. Solorzano, N. Rodriguez-Cousino, R. Esteban, T. Fujimura, Persistent yeast single-stranded RNA viruses exist in vivo as genomic RNA.RNA polymerase complexes in 1: 1 stoichiometry. *J. Biol. Chem.* **275**, 26428–26435 (2000).
23. L. Vega, L. Sevillano, R. Esteban, T. Fujimura, Resting complexes of the persistent yeast 20S RNA Narnavirus consist solely of the 20S RNA viral genome and its RNA polymerase p91. *Mol. Microbiol.* **93**, 1119–1129 (2014).
24. R. Esteban, L. Vega, T. Fujimura, Launching of the yeast 20 s RNA narnavirus by expressing the genomic or antigenomic viral RNA in vivo. *J. Biol. Chem.* **280**, 33725–33734 (2005).
25. R. Esteban, T. Fujimura, Launching the yeast 23S RNA Narnavirus shows 5′ and 3′ cis-acting signals for replication. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2568–2573 (2003).
26. R. Esteban, L. Vega, T. Fujimura, 20S RNA narnavirus defies the antiviral activity of SKI1/XRN1 in Saccharomyces cerevisiae. *J. Biol. Chem.* **283**, 25812–25820 (2008).
27. M. A. B. Haase, D. M. Truong, J. D. Boeke, Superloser: A plasmid shuffling vector for Saccharomyces cerevisiae with exceedingly low background. *G3 (Bethesda)* **9**, 2699–2707 (2019).
28. D. C. Jeffares, A. M. Poole, D. Penny, Relics from the RNA world. *J. Mol. Evol.* **46**, 18–36 (1998).
29. M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
30. S. M. Richardson, S. J. Wheelan, R. M. Yarrington, J. D. Boeke, GeneDesign: Rapid, automated design of multikilobase synthetic genes. *Genome Res.* **16**, 550–556 (2006).
31. S. M. Richardson, P. W. Nunley, R. M. Yarrington, J. D. Boeke, J. S. Bader, GeneDesign 3.0 is an updated synthetic biology toolkit. *Nucleic Acids Res.* **38**, 2603–2606 (2010).
32. T. Fujimura, R. Esteban, Bipartite 3′-cis-acting signal for replication in yeast 23 S RNA virus and its repair. *J. Biol. Chem.* **279**, 13215–13223 (2004).
33. T. Fujimura, R. Esteban, The bipartite 3′-cis-acting signal for replication is required for formation of a ribonucleoprotein complex in vivo between the viral genome and its RNA polymerase in yeast 23 S RNA virus. *J. Biol. Chem.* **279**, 44219–44228 (2004).
34. L. Y. Chan, S. Kosuri, D. Endy, Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1**, 2005.0018 (2005).
35. T. Fujimura, R. Esteban, Interactions of the RNA polymerase with the viral genome at the 5′- and 3′-ends contribute to 20S RNA narnavirus persistence in yeast. *J. Biol. Chem.* **282**, 19011–19019 (2007).
36. V. Lohmann, F. Korner, U. Herian, R. Bartenschlager, Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. *J. Virol.* **71**, 8416–8428 (1997).
37. N. Rodriguez-Cousino, A. Solorzano, T. Fujimura, R. Esteban, Yeast positive-stranded virus-like RNA replicons. 20 S and 23 S RNA terminal nucleotide sequences and 3′ end secondary structures resemble those of RNA coliphages. *J. Biol. Chem.* **273**, 20363–20371 (1998).
38. J. R. Coleman *et al.*, Virus attenuation by genome-scale changes in codon pair bias. *Science* **320**, 1784–1787 (2008).
39. S. Mueller *et al.*, Live attenuated influenza virus vaccines by computer-aided rational design. *Nat. Biotechnol.* **28**, 723–726 (2010).
40. M. A. Martinez, A. Jordan-Paiz, S. Franco, M. Nevot, Synonymous virus genome recoding as a tool to impact viral fitness. *Trends Microbiol.* **24**, 134–147 (2016).
41. R. M. Yarrington, S. M. Richardson, C. R. Lisa Huang, J. D. Boeke, Novel transcript truncating function of Rap1p revealed by synthetic codon-optimized Ty1 retrotransposon. *Genetics* **190**, 523–535 (2012).
42. P. Aliahmad, S. J. Miyake-Stoner, A. J. Geall, N. S. Wang, Next generation self-replicating RNA vectors for vaccines and immunotherapies. *Cancer Gene Ther.* **22**, 1–9 (2022).
43. K. Lundstrom, Self-replicating vehicles based on negative strand RNA viruses. *Cancer Gene Ther.* **15**, 1–14 (2022).
44. M. A. Morse *et al.*, Clinical trials of self-replicating RNA-based cancer vaccines. *Cancer Gene Ther.* **10**, 1–9 (2023).
45. J. Herold, R. Andino, Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol. Cell* **7**, 581–591 (2001).
46. D. J. Barton, B. J. O'Donnell, J. B. Flanegan, 5′ cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. *EMBO J.* **20**, 1439–1448 (2001).
47. B. L. Nicholson, K. A. White, Functional long-range RNA-RNA interactions in positive-strand RNA viruses. *Nat. Rev. Microbiol.* **12**, 493–504 (2014).
48. Y. Matsumoto, R. Fishel, R. B. Wickner, Circular single-stranded RNA replicon in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7628–7632 (1990).
49. C. B. Brachmann *et al.*, Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1998).
50. G. Giaever *et al.*, Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387–391 (2002).
51. M. Ramírez Garrastacho, Análisis de la replicación del Narnavirus 23S RNA de Saccharomyces cerevisiae: Interacciones con el metabolismo del hospedador (Universidad de Salamanca, 2011).
52. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
53. E. L. Hatcher *et al.*, Virus Variation Resource - improved response to emergent viral outbreaks. *Nucleic Acids Res* **45**, D482-D490 (2017).
54. R. Lorenz *et al.*, ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
55. J. Cai, J. S. Bader, *RNA* analysis. *Zenodo*. https://doi.org/10.5281/zenodo.8007808. Deposited 5 June 2023.