Supplementary Materials for

Transcription polymerase-catalyzed emergence of novel RNA replicons

Nimit Jain, Lucas R. Blauch, Michal R. Szymanski, Rhiju Das, Sindy K. Y. Tang, Y. Whitney Yin, Andrew Z. Fire

correspondence to: afire@stanford.edu

This PDF file includes:

Supplementary Text
Supplementary Protocols
Figs. S1 to S21
Captions for Tables S1 to S10

Other Supplementary Materials for this manuscript include the following:

Tables S1 to S10

Supplementary Text

Role of 3' nucleotide additions in RNA replication

To distinguish between subterminal (Fig. 2B) and terminal initiation mechanisms, we analyzed the 5' and 3' sequence ends of RNA products from reactions initiated using templates with an extra 3' adenosine monophosphate. Under a terminal initiation model for such templates, uracil would be expected as the 5' base for complementary strand products. Further, for products with the same strand orientation as the starting template, an expectation with terminal initiation would be that a 3' consensus adenine is positioned in the sequences before the occurrence of diverse, T7 RNAP-catalyzed, 3' nucleotide additions. On the other hand, under a subterminal initiation model, both (i) uracil as the 5' base for complementary strand products and (ii) a 3' consensus adenine for the same strand products, would not be expected.

In our data, complementary strand products do not evidence 5' uracil above background levels (background measured using control chemically synthesized RNA oligos; a background of 5' nucleotide extensions was expected from reverse transcriptase activity during RNA-Seq library preparation) (fig. S6C). An interpretation of our observed 5' end sequence distributions is that guanine serves as the main 5' base on one strand and cytosine on the other strand, consistent with the 5' initiation nucleotide identities experimentally determined by Konarska and Sharp using two different assays (*3*). Furthermore, in our data, same strand products did not contain a 3' consensus adenine (fig. S7). 3' nucleotide additions by T7 RNAP were not positioned after a possible 3' consensus adenine; instead, diverse 3' nucleotide additions were detected prior to the expected position of a 3' consensus adenine. Thus, analysis of both 5' and 3' sequence ends of RNA products supports a subterminal initiation model over terminal initiation.

We note that previously published chromatography data are consistent with our findings regarding the significance of 3' nucleotide additions in RNA replication by T7 RNAP. The high frequency of 3' nucleotide additions in replicating RNA populations may explain why Konarska and Sharp observed all four nucleotides at the 3' end of X RNA using a radioactivity-based assay [Fig. 7D in (*4*)]. Furthermore, a role of 3' nucleotide extensions could potentially have been masked in previous studies on T7 RNAP-RNA replication because the RNA templates were prepared using run-off transcription of synthetic DNA oligos, which is known to result in RNA products with 3' nucleotide extensions [e.g. (*9*, *10*)].

We further note a slight gel mobility difference between Y2 RNA replication products and chemically synthesized Y2 RNA oligos (Fig. 2A, fig. S6A) on our denaturing gels [10% TBE-urea (29:1 acrylamide/bis-acrylamide)]. The mobility difference may be collectively accounted for by (i) the different 5' chemical ends of replication products (5'-triphosphate) and RNA oligos (5'-hydroxyl) (fig. S6B), and by (ii) 3' nucleotide extensions longer than one nucleotide in replication products.

Requirement of 2-way- and 4-way- repeats for efficient RNA replication

2-way- and 4-way- repeats confer a fitness advantage for RNA replication by T7 RNAP. However, RNA templates with distortive mutations that would disrupt perfect complementarity in the 2-way- or 4-way- repeats can (at least in some cases) still be replicated, as evidenced by (i) strong correlation between frequencies of distortive mutations on one strand and frequencies of their complementary mutations on the other strand (fig. S8), and (ii) concordance of distortive mutations between the two halves of RNA dimers (fig. S12). The capability of templates with distortive mutations to be replicated shows a lack of rigid RNA structure requirements for replication, and has implications for replicating RNA evolution: RNAs could evolve gradually through single sequence changes at a time.

Additionally, we note that for the $Y2_1$ degenerate library in Fig. 3B, the second most abundant 4 base combination was not Watson-Crick but was a single sequence change away from the most abundant 4 base combination (which was a 4-way Watson-Crick base combination). The specific single sequence change in the second most abundant 4 base combination could still allow pairing between the 4-way repeat units through a GU wobble base-pair, at least for one of the replicating RNA strands.

Kinetics of RNA synthesis using the X_1 and $Y2_1$ degenerate libraries as templates are shown in table S3.

Our suggestion that the role of the 2-way- and 4-way- repeats is structural rather than sequence-specific is based on observations that a given replicon sequence can retain replication competence when recoded with diverse 2-way- and 4-way- Watson-Crick base combinations (Fig. 3). Sequence biases for replication of particular 2-way- and 4-way- Watson-Crick base combinations are present, however, and can be noted from Fig. 3.

We also note that while 4-way Watson-Crick base combinations [(A,U,A,U),(U,A,U,A),(C,G,C,G),(G,C,G,C)] preserve both 2-way- and 4-way- repeats (Fig. 3B), we were able to distinguish a 4-way repeat requirement from a 2-way repeat requirement in the experiments of Fig. 3B because base combinations preserving the 2-way repeat but not the 4-way repeat [e.g. (A,A,U,U), (C,C,G,G) etc.] were less abundant post-replication compared to the 4-way Watson-Crick base combinations.

Interrupted rolling circle mechanism for RNA concatemer synthesis

We performed several quantitative analyses to assess the sequence agreement between RNA dimer halves. We found that the observed sequence agreement between dimer halves was much more frequent than would be expected based on a bi-templated synthesis model (fig. S11). These results suggest that uni-templated synthesis is the dominant mechanism for formation of RNA dimers.

We had obtained RNA dimers starting with mixtures of monomer templates containing intentionally randomized bases at specific positions. In evaluating sequence variants located

outside the intentionally randomized bases in RNA dimers, we found that the concordance of variants between the two dimer halves was more frequent by 4.5-7 fold than would be expected based on the variants occurring independently in each dimer half (fig. S12). To give a sense of the magnitude of this concordance: for most sequence variants, concurrent incidence in both dimer halves was more frequent than incidence in either half alone. These results again support a uni-templated synthesis mechanism for RNA dimer formation.

From examining previously published data on the RNA concatemers of X RNA (*3*), we note that an interrupted rolling circle model quantitatively explains the RNase T1 cleavage patterns observed for these RNA concatemers. A previous report had hypothesized an apparent rolling-circle mechanism operating on single-stranded linear DNA oligos transcribed by T7 RNAP (*53*). But in that report, only a single template sequence was used per reaction and therefore, the data shown were also consistent with a mechanism for RNA concatemer formation involving multiple template molecules.

A structural interpretation of our interrupted rolling circle model may be that upon completion of a round of template copying, the 5' and 3' ends of a replicating RNA monomer template are close to each other in space at the active site of T7 RNAP. The proximity of the template ends in space may facilitate jumping of T7 RNAP from the 5' to 3' end.

The mechanism generating the extra nucleotides observed at the junction between the two halves in RNA dimers is not fully known. The extra nucleotides at dimer junctions could be a result of 3' nucleotide additions to RNA products by T7 RNAP as it jumps from the 5' to 3' end of the RNA template and/or a result of the copying of the extra nucleotides present at the 3' end of the monomer template.

An apparent difference between the RNAs replicated by T7 RNAP and RNAs that have been proposed to follow a symmetric rolling circle model for replication by transcription polymerases [viroids in the family *Avsunviroidae* (*13*) and human Hepatitis delta virus (*14*)] is that the latter category of replicons encode ribozymes that can self-cleave in RNA concatemers to yield monomers. For the T7 RNAP-RNA replication system, while we cannot exclude the possibility that a minor fraction of RNA monomer templates are generated indirectly via self-cleavage of RNA concatemers, we note that a substantial fraction of monomers appear to have 5'-triphosphate (fig. S6B), consistent with the monomers being direct products of RNA synthesis.

Potential relevance of an interrupted rolling circle model to viroid replication

Current mechanistic models for replication of viroids in the family *Avsunviroidae* involve RNA concatemer intermediates produced by rolling circle synthesis using circular RNA templates. Linear RNA molecules have also been detected alongside circular RNAs for viroids of the *Avsunviroidae* family (*27*, *28*). It has been proposed that the linear RNA molecules may be active as templates for instructing RNA synthesis (*28*) but how linear RNAs might template synthesis of RNA concatemers is unclear.

An interrupted rolling circle mechanism with linear RNA templates offers a plausible means for RNA concatemer synthesis. To assess the potential role of an interrupted rolling circle mechanism in viroid replication, we examined published data for two viroids, avocado sunblotch viroid (ASBVd) (*27*) and peach latent mosaic viroid (PLMVd) (*28*). Both ASBVd and PLMVd belong to the *Avsunviroidae* family of viroids, and are replicated in the chloroplasts of infected plants. ASBVd appears to be replicated by a chloroplastic RNA polymerase similar to T7 RNAP (*8*). ASBVd and PLMVd populations contain particular 5' triphosphate-bearing, monomer-length, linear RNA sequences for both strand orientations. Data on both initiation and termination of RNA synthesis to produce these 5'-triphosphate containing molecules, are more parsimoniously explained by a linear template model rather than a circular template model.

With respect to initiation of RNA synthesis (or 5' end specification): For both ASBVd and PLMVd, the measured 5' initiation site for the (+) strand corresponds to a site within a few nucleotides of the 3' end of a subset of linear (-) molecules present in the RNA population. Similarly, the 5' initiation site for the (-) strand corresponds to a site within a few nucleotides of the 3' end of a subset of linear (+) molecules in the population. In a circular template model, such positioning for the 5' ends of the (+) and (-) strands would be considered coincidental, with an additional source of specificity such as particular structural or sequence motifs (*27*, *28*) required to explain the initiation site positioning. In contrast, in a linear template model, the measured 5' ends of the (+) and (-) strands would be expected simply based on full-length copying.

With respect to termination of RNA synthesis (or 3' end specification): The presence of a defined set of 5'-triphosphate containing, monomer-length, linear molecules in ASBVd and PLMVd populations requires an explanation for precise 3' end generation. In a circular template model, RNA 3' end generation can be explained by positing specific termination signals for RNA synthesis or by particular RNA cleavage events *in vivo*. In contrast, in a linear template model, RNA 3' end generation can be explained more simply by the termination of RNA synthesis upon reaching the template 5' end.

Our analysis suggests that it would be valuable in future work to experimentally assess the relative contributions of linear and circular templates in viroid RNA replication, especially for viroids in the *Avsunviroidae* family.

Origin of replicating RNAs through partial instruction from DNA seeds

Before conducting the no-template-added, high concentration T7 RNAP reactions in drop format, we first tested whether our microfluidic assay could support replication of our characterized chemically synthesized RNA templates at low concentrations of T7 RNAP. Templated RNA replication catalyzed by T7 RNAP in drops was evident using (i) gel electrophoresis analysis, whereby RNA synthesized cumulatively in a pool of drops could be visualized, and using (ii) a fluorescence imaging-based drop-by-drop assay of RNA synthesis, with inclusion of a nucleic-acid binding dye into the drops. In the latter approach, dilution of the starting RNA template allowed us to track the percentage of drops that were fluorescent after reaction incubation as a function of the starting RNA template concentration, akin to digital droplet PCR (fig. S16).

For the RNAs synthesized in no-template-added, high concentration T7 RNAP drop reactions, we also conducted functional tests to assess replication-competence. Specifically, aggregated drop reactions were used in bulk as templates in fresh, microliter-scale, low concentration T7 RNAP reactions and the resulting RNA pools sequenced. The numerous RNA species from the initial no-template-added drop reactions that were amplified in the bulk, low concentration T7 RNAP reactions exhibited typical sequence and structural hallmarks of replicating RNAs (fig. S18): (i) 2-way repeats, (ii) 4-way repeats and (iii) GG and CC end sequences outside the 2-way repeats: one strand containing two G nucleotides at or close to both the 5' and 3' ends (and therefore, the complementary strand containing two C nucleotides at or close to both the 5' and 3' ends). We concluded that novel replicating RNAs can be isolated from no-template-added drop reactions.

Of note, no-template-added tube and no-template-added aggregated drop reactions migrated differently on denaturing gels. The tube reactions appeared mostly as well-defined bands corresponding to particular replicating RNA species (e.g. fig. S1B). The aggregated drop reactions appeared as smears (fig. S17), reflecting the rich diversity of RNA products that was also evident upon high-throughput sequencing.

We performed the analyses presented in Fig. 5 and fig. S20C as follows. For each sequenced pool from an aggregated drop reaction or tube reaction, we performed a global, sequence-agnostic analysis and grouped all the detected sequences into RNA species. For each of the aggregated drop reactions, a subset of species contained complementary RNA sequences with GG and CC end sequences located outside a 2-way repeat configuration. Within this subset of RNA species, two distinguishable clusters of species were observed, corresponding to species with long and short 2-way repeats. Based on previous experimental results (Fig. 1, Fig. 3, fig. S18), we identified as replicating RNAs from all drop and tube reactions, RNA species that contained two sequence hallmarks: (i) long 2-way repeats, and (ii) GG and CC end sequences located outside the 2-way repeats (with the molecules containing the GG and CC end sequences being complementary). These two sequence hallmarks were also found to be sufficient to identify the predominant RNA species in cases where the reaction products migrated as well-defined bands on denaturing gels (i.e. tube reactions that had been set up in parallel as part of the experiment). It should be noted that other RNA species in the aggregated drop reactions that we are currently excluding from analysis (e.g. species with short 2-way repeats or species without

the GG and CC end sequences) could also be competent for replication. Our current knowledge of replicating RNA sequence features stems primarily from tube-based replication assays which are inherently competitive in nature. Compartmentalizing the volume of a tube reaction into smaller drop reactions could lead to better detection of replicating RNA species with divergent sequence features.

The chemical space of nucleic acids that can seed emergence of novel RNA replicons is not fully known. Although our experiments provide evidence for the origin of replicating RNAs from DNA seeds, it is foreseeable that particular RNA molecules could also work as seeds in certain circumstances (*54*). For example, we might expect any RNA that mimics an intermediate product involved in the proposed model in Fig. 5D to serve as a seed. Furthermore, our assays do not currently allow us to gauge relative seeding efficiencies for different types of DNA molecules (single-stranded versus double-stranded, or with differing length, sequence identity or end configuration such as 3' overhang versus 5' overhang versus blunt ended for dsDNA seeds). As we obtained replicating RNAs matching our complex seed pool both before and after treatment of the seed pool with hot alkali, both single-stranded and double-stranded DNA molecules may be competent as seeds.

It is important to appreciate the difference between (i) a replicating RNA originating from a seed and (ii) being able to detect a replicating RNA as having originated from a seed. We can only confidently assign replicating RNAs to initiating seeds when the detected seed matches are long, and essentially mismatch- and gap-free. Such high-quality seed matches were observed for only a subset of replicating RNAs. The lack of a significant seed match to a replicating RNA could be for several reasons, including: (i) the initial seed used in generating the replicating RNA may have contributed only a short sequence, (ii) the replicating RNA may have diverged in sequence from its seed due to extensive mutation and selection, (iii) the seed sequence may be absent from our current databases, and (iv) the replicating RNA could conceivably have originated through alternative mechanisms such as non-templated, *de novo* polymerization of NTP substrates (*29*).

Some details of the mechanistic scheme proposed in Fig. 5D are also worth clarifying: (i) The RNA product in the first step of the model ("Transcription") contains a sequence stretch matching the DNA seed (red box in Fig. 5D) but may additionally contain novel 5' and 3' end sequences generated by T7 RNAP (black stubs flanking the red box); (ii) The first round of RNA-templated 3' extension may be primed by nucleotides that were copied from the DNA seed and/or by extra nucleotides added by T7 RNAP to the 3' end of the transcribed RNA product; (iii) The second round of RNA-templated 3' extension is depicted in Fig. 5D as comprising denaturation of a hairpin RNA intermediate (which contains two units of the 4-way repeat) followed by self-templated 3' extension. Denaturation of the hairpin RNA intermediate could also occur alongside (rather than before) self-templated 3' extension, similar to intended stem-loop DNA formation during loop-mediated isothermal amplification of DNA by DNA polymerases (*55*); (iv) While the two rounds of RNA-templated 3' extension are depicted as being intramolecular in Fig. 5D, the possibility of RNA-templated intermolecular 3' extension cannot be excluded; (v) More than two rounds of RNA-templated 3' extension could also occur [e.g. sequence in the loop region of the putative long hairpin structure of replicating RNAs could (at least in some cases) be derived from an additional round of RNA-templated 3' extension];

and (vi) RNA-templated synthesis of new RNA chains could occur at several intermediate steps before the formation of a full-length replicating RNA.

In terms of relevance for other biological systems, we note that DNA-dependent RNA polymerases divergent from T7 RNAP can also demonstrate the biochemical activities involved in our replicon emergence model (Fig. 5D), including apparently promoter-less transcription of DNA to RNA [mammalian RNAP example: (*56*) and references therein], RNA-templated 3' extension of RNA [mammalian RNAP example: (*57*), yeast RNAP example: (*58*) and references therein] and RNA-templated *de novo* initiation of new RNA product molecules [bacterial RNAP examples: (*59*, *60*)].

Sequence composition of replicating RNAs

We analyzed the sequence composition of replicons in our repertoire (table S5). We found replicating RNAs to have an AU-rich sequence character (fig. S21), which could potentially play a role in replication by facilitating separation of complementary strands from each other and/or by facilitating shape-shifting via the 4-way repeat [for parallels of AT-rich sequences favoring similar types of structural transitions in DNA, see e.g. (*61*, *62*)]. Beyond any role in RNA replication, the AU-rich sequence composition of replicons could also reflect sequence biases of the biochemical steps involved in the initial template evolution process (Fig. 5D).

Temperature dependence of high concentration T7 RNAP reactions

We note that high concentration T7 RNAP reactions exhibit a strong temperature dependence. Reactions (set up in bulk in tubes) that were maintained for a length of time at room temperature before being incubated at 37°C, appeared as smears on denaturing gels. We further investigated this observation by comparing in parallel, high concentration T7 RNAP reactions immediately incubated at 37°C with reactions immediately incubated at 25°C. Reactions at 37°C showed, as expected, gel densities that differed between reactions (corresponding to formation of different replicons in different reactions) whereas reactions at 25°C all appeared as smears on denaturing gels. We have not extensively characterized the RNA products synthesized at room temperature but some sequencing results indicate that the time spent by a reaction at room temperature is correlated with the count of homopolymeric RNA sequences [specifically, poly(rA) and poly(rU)] detected in the corresponding sequenced pool.

DNase treatment of no-template-added T7 RNAP reactions

We compared RNA synthesis in no-template-added, high concentration T7 RNAP reactions, with and without addition of DNase. At an earlier time point (~4-5 hours) into the reactions at 37°C, all reactions with DNase were clear without any visible precipitate and these reactions exhibited no RNA products detectable by gel electrophoresis. In contrast, positive controls (reactions without DNase) were visibly turbid and exhibited gel densities that differed between reactions (signature of replicating RNA evolution). At a later time point (~23-24 hours) into the reactions at 37°C, all reactions irrespective of DNase treatment were visibly turbid. However, for this later time point, the gel electrophoretic pattern of reactions with and without DNase still appeared different. While the reactions without DNase showed gel densities that differed between reactions, the +DNase reactions all appeared as smears. The sensitivity of RNA synthesis to DNase treatment observed for the earlier time point is consistent with our model for RNA synthesis from trace amounts of DNA seeds in no-template-added reactions but we cannot rule out inhibitory effects of DNase addition that are separate from DNA cleavage. For the later time point results, one possible interpretation is that DNase addition causes selection for novel types of RNA species that evolve from DNA seeds resistant to DNase. Another possible interpretation is that by reducing RNA synthesis from contaminating DNA seeds, other processes such as RNA synthesis from RNA seeds or *de novo* non-templated polymerization of NTPs become manifest.

RNA replication by the DNA-dependent RNA polymerase of bacteriophage T3

We found that T3 RNA polymerase can replicate an RNA species with a reference sequence similar to Y2 RNA. The capability of T3 RNA polymerase to replicate RNA was also noted by Biebricher and Luce (*5*).

Supplementary Protocols

Fig. 2A- and fig. S6-specific supplementary protocols

To each of the RNA oligos AF-NJ-219 and AF-NJ-220, adenosine 3',5'-diphosphate (pAp) was added using T4 RNA ligase 1 (*63*) as follows: 90 µl of reaction volume containing 50 pmol of RNA oligo was denatured at 95°C for 3 minutes followed by snap cooling on ice for 3 minutes. The reaction was removed from ice and the following reagents were quickly added: 10 µl of 100 µM pAp (in water), 15 µl of 10x T4 RNA ligase reaction buffer, 15 µl of 10 mM ATP, 15 µl of 100% DMSO and 5 µl of T4 RNA ligase 1 (50 units). Reaction incubation was at 16°C for 22.25 hours in a thermal cycler. The reaction was stopped by addition of SDS and EDTA, followed by an extraction with 1:1 phenol-chloroform.

We used serial dilution to quantitatively compare T7 RNAP reaction yields from three template types (Fig. 2A, fig. S6A and data not shown): (i) Y2 RNA synthetic oligos with an extra 3' adenosine monophosphate, (ii) Y2 RNA synthetic oligos without an extra 3' nucleotide and (iii) gel-extracted Y2 RNA monomer replication products. In these assays, RNA oligos with an extra 3' adenosine monophosphate were far more potent than oligos without an extra nucleotide in generating replicating populations, with yields from 16-fold dilution of extra 3' adenosine monophosphate-containing oligos comparable to yields from undiluted oligos which did not contain an extra 3' nucleotide. The third template type—gel-extracted Y2 RNA monomer replication products—yielded roughly similar amounts of reaction products after ~16-32 fold dilution compared to undiluted RNA oligos with an extra 3' adenosine monophosphate[1].

Quantification of gel intensities was performed using the raw image data with AlphaView software (ProteinSimple). For each reaction lane, gel intensity was quantified within a bounding box made from approximately 52 to 60 nucleotides (RNA oligo input bands at ~50 nucleotides were excluded so as not to have signal from the input template). The bounding boxes did not contain any saturated pixels. The average intensity from "blank" bounding boxes on the same gel was used for background subtraction.

For treatment of Y2 RNA replication products with RppH or SAP (fig. S6B), RNA was first denatured at 95°C for 3 minutes followed by snap cooling on ice for 3 minutes. Buffer components and enzymes were added subsequently. Buffer compositions for the phosphatase treatments were based on manufacturer recommendations. Phosphatase reactions were incubated at 37°C for 1 hour followed by heat inactivation at 65°C for 20 minutes. Prior to loading on gels,

[1] Several possibilities could account for the lower template efficiency of RNA oligos with an extra 3' adenosine monophosphate compared to the gel-extracted Y2 RNA monomer replication products, including (i) an uncharacterized template requirement [e.g. particular dependence on a type of RNA structure or on the 5' chemical end of the RNA (synthetic RNA oligos have 5' hydroxyl ends whereas replication products have 5' triphosphate ends)], (ii) a more efficient value for a characterized template requirement (e.g. 3' nucleotide extensions other than a single adenosine monophosphate may be more efficient for instructing RNA synthesis), and (iii) an uncharacterized growth advantage due to the complex ensemble character of the Y2 RNA replication products (see e.g. fig. S8) versus the synthetic RNA oligos.

RNA was isolated by addition of SDS and EDTA, 1:1 phenol-chloroform extraction and ethanol precipitation.

Fig. 3-specific supplementary protocols

Replication reactions and sequencing for the X_1 (AF-NJ-257) and $Y2_1$ (AF-NJ-258) libraries were performed in duplicate with similar results. Starting RNA oligo template concentrations for replication of the X_1 and $Y2_1$ libraries were 2 ng/µl and 4 ng/µl, respectively.

The pre-replication RNA pools for the X_2, X_3, X_4 and $Y2_2$ libraries were prepared by T7 RNAP-catalyzed DNA transcription of DNA oligos AF-NJ-200, AF-NJ-201, AF-JTG-11 and AF-JTG-13, respectively. In these DNA transcription reactions, final concentrations of AF-NJ-200 and AF-NJ-201 were 25 nM, and of AF-JTG-11 and AF-JTG-13 were ~2.4 ng/µl.

Prior to RNA replication, the transcribed X_2 and X_3 RNA pools were treated with TURBO DNase (3 µl TURBO DNase in a 50 µl reaction with 1x TURBO DNase buffer) at 37°C for 1 hour, followed by addition of SDS and EDTA, 1:1 phenol-chloroform extraction and ethanol precipitation.

Fig. 5-, fig. S17- and fig. S20C-specific supplementary protocols

Covaris shearing of DNA: DNA (in TE buffer, pH 8) was sheared to a size range of 100-300 base-pairs as assessed by agarose gel electrophoresis. Sheared DNA was purified using the Zymo Clean and Concentrator kit[2].

Restriction digestion[3]*:* 75 µl reactions with either MnlI (7.5 µl) or Hpy188III (6 µl), DNA and 1x CutSmart buffer were incubated at 37°C for 2 hours. Digests were monitored to reach near completion by agarose gel electrophoresis. Digested DNA fragments were purified using the Zymo Clean and Concentrator kit[4].

Bacteriophage lambda DNA was Covaris sheared. A Fire lab plasmid pPD122.03 was mini-prepped using the ZymoPURE Plasmid Miniprep kit, which includes an RNase A digestion step

[2] Column purification of DNA seeds using the Zymo Clean and Concentrator kit is expected to impose a lower limit size cutoff on the recovered DNA fragments.

[3] Hpy188III and MnlI were chosen as restriction enzymes because the two enzymes are expected to generate, on average, fragments of roughly similar size as fragments generated by Covaris shearing. Additionally, these two enzymes allow for generation of a diverse pool of DNA seeds because: (i) the recognition sequences and/or cleavage sites of the two enzymes contain degenerate bases, (ii) the two enzymes leave different kinds of overhangs (Hpy188III leaves 5' overhangs and MnlI leaves 3' overhangs), and (iii) the two enzymes have different relationships between the cleavage site and recognition sequence (Hpy188III cuts at its recognition sequence whereas MnlI cuts a few nucleotides away from its recognition sequence).

[4] To minimize denaturation of short dsDNA fragments, heat inactivation was not used for stopping the restriction enzyme reactions.

(RNase A containing-buffer ZymoPURE P1 was stored at 4°C for maximal retention of activity per manufacturer guidelines). The plasmid was then Covaris sheared. *S. cerevisiae* genomic DNA was restriction digested separately with MnlI and with Hpy188III.

Genomic DNA was prepared from the nematode strains using a standard protocol involving SDS-Proteinase K treatment followed by phenol-chloroform extraction and ethanol precipitation. Genomic DNA preps (DNA amounts up to 7 µg/prep) were treated with 30 µg of RNase A (ThermoFisher) at pH 7.4 at 42°C for 2 hours (no salt added for RNase A treatment), followed by removal of RNase A using Proteinase K-SDS treatment and 2 extractions with 1:1 phenol-chloroform. No gel density corresponding to RNA was visible by agarose gel electrophoresis following RNase A digestion. *C. elegans* DNA was then Covaris sheared, *C. remanei* DNA digested with MnlI and *C. brenneri* DNA digested with Hpy188III.

Our designed DNA seed pool consisted of seven types of DNA seeds (percentage contribution by mass given): (i) Sheared lambda phage genomic DNA (7%), (ii) Sheared *C. elegans* genomic DNA (7%), (iii) Sheared DNA from the plasmid pPD122.03 (7%), (iv) MnlI digested *C. remanei* genomic DNA (20%), (v) Hpy188III digested *C. brenneri* genomic DNA (15%), (vi) MnlI digested *S. cerevisiae* genomic DNA (19%), and (vii) Hpy188III digested *S. cerevisiae* genomic DNA (25%).

After pooling the seven types of DNA seeds together, the combined DNA seed pool was treated with 100 units of RNase I in the presence of 100 mM NaCl at pH 8 at 37°C for 1 hour. RNase I was removed using 0.2% SDS treatment followed by 2 extractions with 1:1 phenol-chloroform[5]. A "No RNase I control" was used to confirm that RNase I treatment did not lead to loss of DNA.

The DNA seed pool was then split into three equal parts: (i) No further treatment (except for addition of TURBO DNase buffer to 1x final concentration), (ii) Treatment with 3 µl TURBO DNase (in a 50 µl reaction with 1x TURBO DNase buffer) at 37°C for 1 hour, and (iii) Heating with sodium hydroxide (0.2 N; reaction volume was 10 µl) at 70°C for 1 hour. For neutralization of the sodium hydroxide, 20 µl 200 mM Tris-HCl at pH 7 was added.

After the respective treatments to the three parts of the DNA seed pool, SDS and EDTA were added to each part, followed by extraction with 1:1 phenol-chloroform and ethanol precipitation.

The efficacy of TURBO DNase treatment of the DNA seed pool was assessed by measuring DNA concentrations for the first (no DNase treatment) and second (+DNase treatment) parts of the seed pool. DNase treatment was found to reduce DNA concentration by ~50 fold.

T7 RNAP reactions were set up in drop and tube format for four experimental conditions in parallel: (i) unseeded, (ii) seeded with DNA pool, (iii) seeded with DNase-treated DNA pool and (iv) seeded with hot alkali-treated DNA pool. For the "seeded with DNA pool" condition, the volume seeded with the first part of the DNA seed pool (neither DNase nor NaOH treated) gave a final DNA seed reaction concentration of ~47 femtograms per µl (estimated to correspond to

[5] To minimize denaturation of short dsDNA fragments, heat inactivation was not used for stopping the RNase I reaction.

~10-15 molecules of DNA seeds per droplet); an equivalent volume of the second and third parts of the DNA seed pool was seeded for the "seeded with DNase-treated DNA pool" and "seeded with hot alkali-treated DNA pool" conditions, respectively. Each replicate of drop reactions for an experimental condition consisted of ~50 µl total volume (drops + oil) and took ~5 minutes for generation.

An MS2 spike-in for RNA-Seq was created by fragmentation of bacteriophage MS2 genomic RNA in a solution of 5 mM Na_2CO_3, 45 mM $NaHCO_3$ and 1 mM EDTA at 95°C for 30 minutes (*64*). MS2 fragments in the 70-90 nucleotides size range were gel-extracted and subsequently 3' dephosphorylated by T4 PNK treatment in 100 mM MES-NaOH (pH 5.4), 10 mM $MgCl_2$, 10 mM beta-mercaptoethanol and 300 mM NaCl, at 37°C for 6 hours (*64*); this was followed by purification using the NEB Monarch RNA Cleanup kit (NEB #T2030S), and then by an extraction with 1:1 phenol-chloroform and ethanol precipitation. 60 picograms of the prepared MS2-spike in was added to the aggregated drop reaction products for sequencing, and 300 picograms to the tube reaction products.

In Fig. 5A, 220 RNA species are shown on the left scatter plot, and 204 on the right scatter plot. For each seeded or unseeded condition, RNA species from two different aggregated drop reactions (corresponding to two time points) are shown together on the scatter plots.

For Fig. 5B, RNA species from aggregated drop reactions seeded with the DNA pool or the hot alkali-treated DNA pool were analyzed. Further, only RNA species with >=26 bases matching to our DNA pool are depicted in the histogram in Fig. 5B because matches in this length range were absent for RNA species from the negative controls (the "unseeded" and "seeded with DNase-treated DNA pool" conditions). Approximately 22% of the total number of RNA species identified from aggregated drop reactions for the "seeded with DNA pool" and "seeded with hot alkali-treated DNA pool" conditions had >=26 bases matching to our DNA pool sequences.

The RNA species shown in Fig. 5C were all isolated from drop reactions, either from the "seeded with DNA pool" condition or from the "seeded with hot alkali-treated DNA pool" condition.

Fig. S16-specific supplementary protocols

Four experimental conditions were set up in parallel: (i) +Template, -T7 RNAP; (ii) -Template, +T7 RNAP; (iii) +Template, +T7 RNAP; (iv) +Template [diluted 10-fold compared to condition (iii)], +T7 RNAP.

SYBR Gold was included in reactions for all four conditions at a final concentration of 1x. AF-NJ-223 was used as template for conditions (i), (iii) and (iv) at a final concentration of 0.1 pM, 0.1 pM and 0.01 pM, respectively. Reactions were kept covered with aluminium foil during incubation.

Bright-field and fluorescence images of drops were acquired in 30 micron tall microfluidic wells using an epifluorescence microscope (Nikon Ti-U) equipped with an electron multiplying CCD camera (Andor). We used an excitation filter with transmission centered at 470 nm and an

emission filter with transmission centered at 525 nm. An exposure time of 0.2 s was used during imaging.

Percentage drops fluorescent for a field of view was calculated by using the fluorescence and bright-field images for the field of view. Specifically, percentage drops fluorescent was calculated as 100 times the ratio: (the number of drops detected in the fluorescence image) / (the number of drops detected in the bright-field image). Images for all four experimental conditions were processed using the same parameters. Automated detection of drops was checked by visual inspection.

We note the results of applying Poisson (single-hit and two-hit) models to the drops fluorescence data shown in fig. S16. As we don't yet know the quantitative parameters characterizing the efficiency of the T7 RNAP-RNA replication process, we *a priori* assume (with either model) that the effective number of template copies per drop may be different from the number that we targeted. Practically speaking, under our assumption, we use observed data for the 3.31-targeted-template-copies-per-drop condition to calculate the effective number of templates per drop and then use that to make a prediction of percentage drops fluorescent for the 0.39-targeted-template-copies-per-drop condition.

Under such a modeling framework, if replication could proceed starting with a single template molecule (single-hit model), then using the measurements obtained with the 3.31-targeted-template-copies-per-drop condition, the percentage drops fluorescent predicted by Poisson statistics for the 0.39-targeted-template-copies-per-drop condition would be 5.2 ± 0.6 (mean ± standard deviation), close to the observed value of 6.6 ± 1.1. In contrast, if replication could only proceed starting with two or more template molecules (two-hit model), the percentage drops fluorescent predicted for the 0.39-targeted-template-copies-per-drop condition would be 1.0 ± 0.1, which deviates from observation.

We note that previous reports in the digital PCR literature support our assumption that the effective number of templates per drop need not be the same as the targeted number. For a related discussion of the underestimation of RNA copy number in reverse-transcription digital PCR experiments, see e.g. (*65*).

Further work in the future would be required to fully test the assumptions in our modeling framework and more definitively determine whether replication can initiate using a single template molecule.

Best practices for conducting T7 RNAP-catalyzed RNA replication reactions

Best laboratory practices for minimizing cross-contamination when working with nucleic acid amplification technologies [e.g., (*66*)] also apply to the study and use of T7 RNAP-catalyzed RNA replication. Amplification of contaminating templates could be harder to control with T7 RNAP-catalyzed RNA replication compared to PCR because (i) no primers are required for RNA replication, and (ii) amplification proceeds continuously during RNA replication as opposed to in discrete cycles during PCR. Amplification of contaminating RNA replicons that are not part of an input template pool but are pre-existing in the laboratory can be minimized

using droplet microfluidics as contaminants could be confined to a few drops. We further highlight key best practices for studying T7 RNAP-catalyzed RNA replication using bulk tube reactions below:

- To prevent contamination of T7 RNAP preps with RNA replicons, we highly recommend that the polymerase preps be isolated in a facility which does not receive any shipments from the facility where experiments on RNA replication have been or are being conducted. Contamination of polymerase preps with a pre-existing replicon will lead to subsequent no-template-added, high concentration T7 RNAP reactions consistently yielding that particular replicon because templated replication occurs more efficiently than evolution of novel replicons [see e.g., (*5*)].

- Maintain a catalogue of which RNA replicon sequences have already been isolated in the laboratory and when these were isolated. If a no-template-added, high concentration T7 RNAP reaction yields a sequence similar to what has previously been observed in the laboratory, then it cannot be ascertained whether the new reaction witnessed molecular evolution or amplified a pre-existing template.

- When studying templated RNA replication, conduct reactions at low concentration of T7 RNAP and for short durations of time (~few hours). Also perform no-template-added controls in parallel and check that no products are detected for these controls.

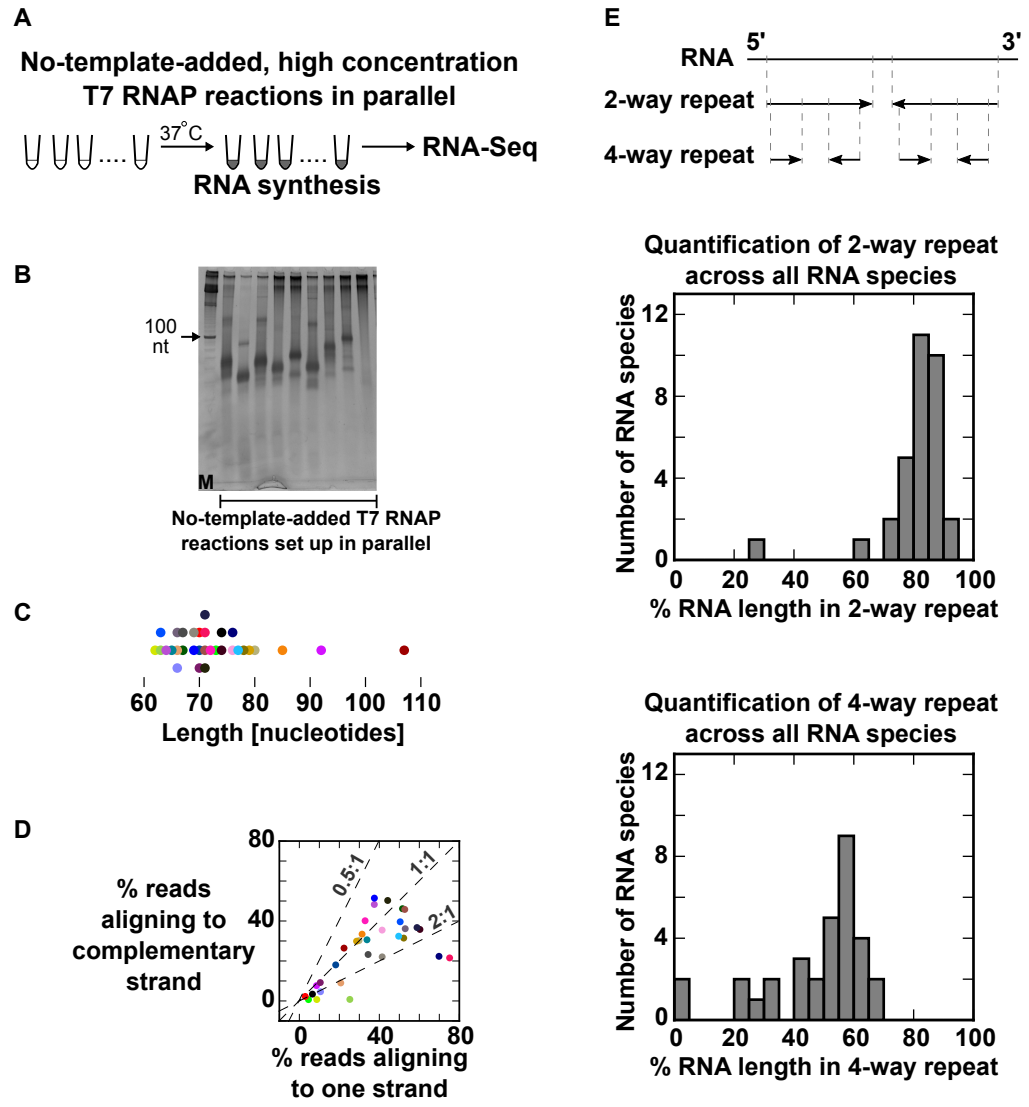


Fig. S1

Characterization of RNA species isolated from no-template-added, high concentration T7 RNAP reactions set up in parallel. (**A**) Experimental schematic to investigate RNAs synthesized in no-template-added reactions. (**B**) Representative denaturing gel image illustrates the different migration of products from no-template-added T7 RNAP reactions that had been set up in parallel. M=marker (denatured 10 base-pair DNA ladder), nt=nucleotides. (**C-E**) Further analysis of the RNA species from Fig. 1. (**C**) Length distribution of reference sequences for the RNA species. (**D**) RNA species are constituted by sequences of both strand orientations. For each RNA species (individual points on scatter plot), the plot shows the fraction of reads from the source reaction aligning to the species reference sequence (x axis) and to the reverse complement of the species reference sequence (y axis). Diagonal lines (0.5:1, 1:1 and 2:1) are shown as visual aids. Color coding for RNA species in panels (C) and (D) is the same as in Fig. 1. (**E**) Quantification of 2-way- and 4-way- repeat lengths for all RNA species from Fig. 1.

Fig. S2. Nimit Jain et al.

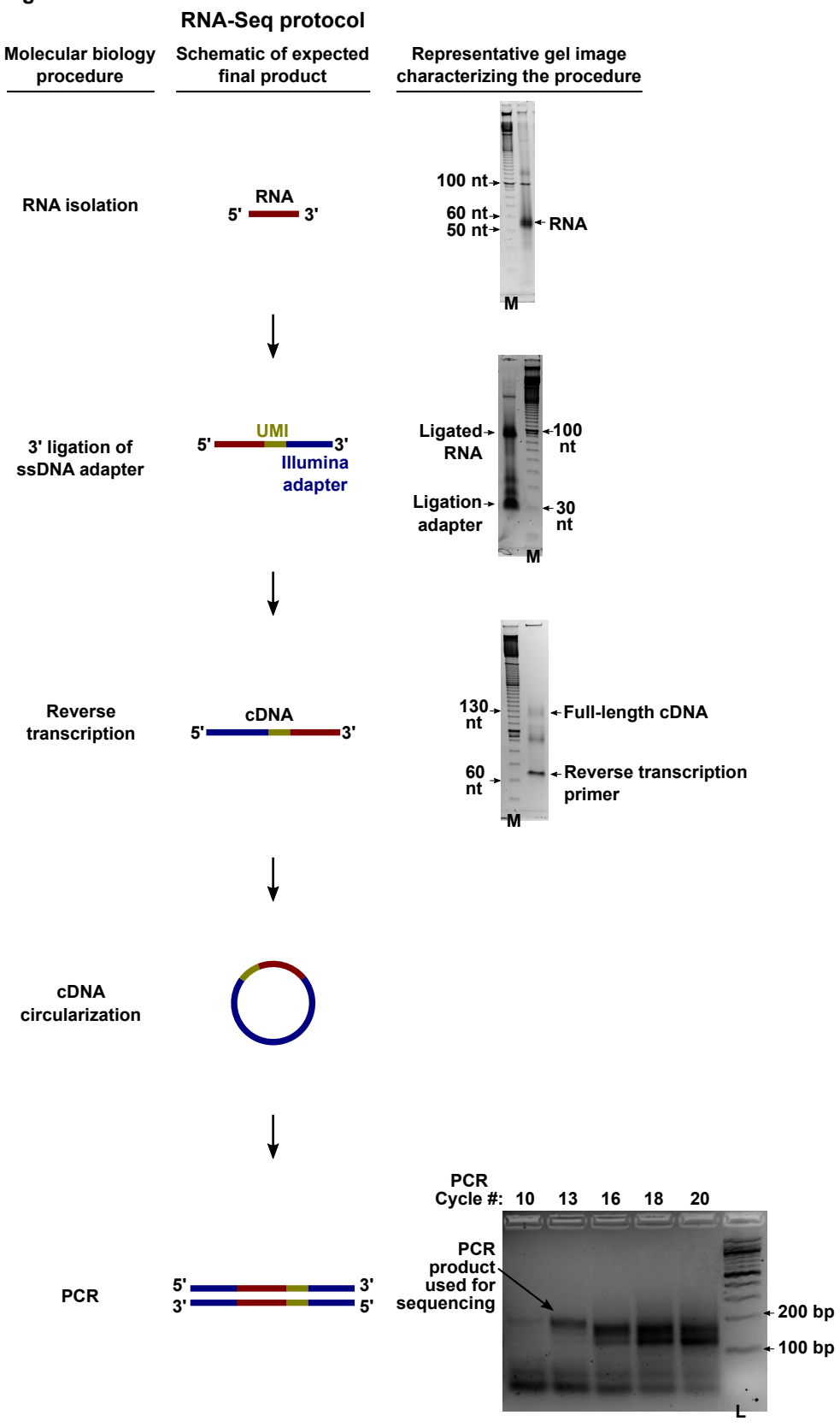
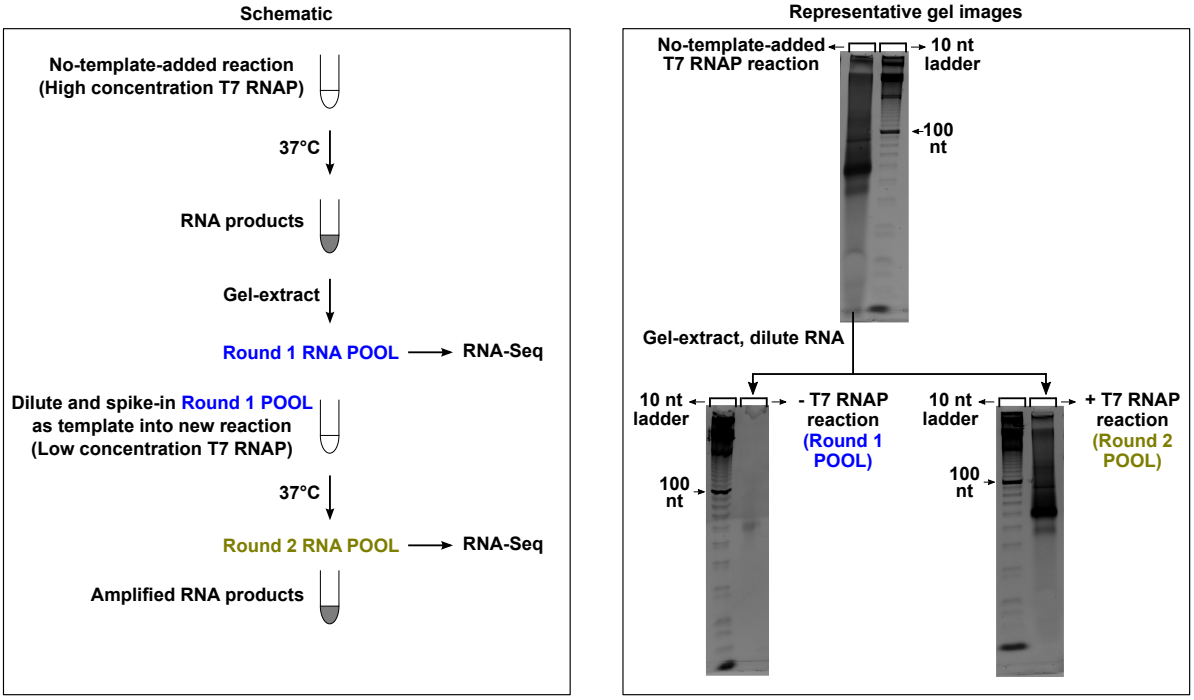


Fig. S2

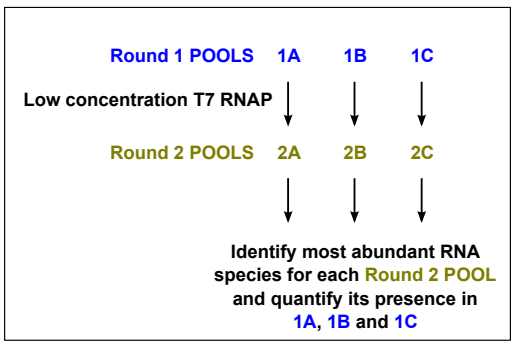
Schematic of RNA-Seq protocol. Representative gel images at various steps of the protocol are shown. UMI=Unique Molecular Identifier (a degenerate 6- or 8-nucleotide molecular barcode), M=marker (denatured 10 base-pair DNA ladder), L=100 base-pair DNA ladder, bp=base-pair, nt=nucleotides.

Fig. S3. Nimit Jain et al.

A Amplification of RNA species upon dilution into fresh T7 RNAP reactions



B Sequence correspondence between Round 1 POOL and Round 2 POOL when distinct RNA species were amplified in parallel



Quantification for the most abundant Round 2 POOL RNA species

Output Pool	% Output reads from most abundant RNA species	Specificity of most abundant output RNA species to input pools (scale from 0 to 1)		
		1A	1B	1C
2A	85%	0.995	0.002	0.002
2B	62%	0.003	0.994	0.003
2C	87%	0.001	0.011	0.988

Fig. S3

Sustained and templated propagation of RNA species originally isolated from no-template-added, high concentration T7 RNAP reactions. (**A**) Regeneration of RNA species upon dilution into fresh, low concentration T7 RNAP reactions. Gels for the (-) and (+) T7 RNAP reactions with the diluted *Round 1* RNA pool as template were processed in parallel. nt=nucleotides. (**B**) Templated growth of RNA species. Three *Round 1* RNA pools (originally isolated from no-template-added reactions) were propagated in parallel. The *Round 2* products

from a particular reaction corresponded in sequence to the *Round 1* RNA pool used as template for that reaction. Sequences for the most abundant RNA species present in the three *Round 2* pools are listed in table S2.

Fig. S4. Nimit Jain et al.

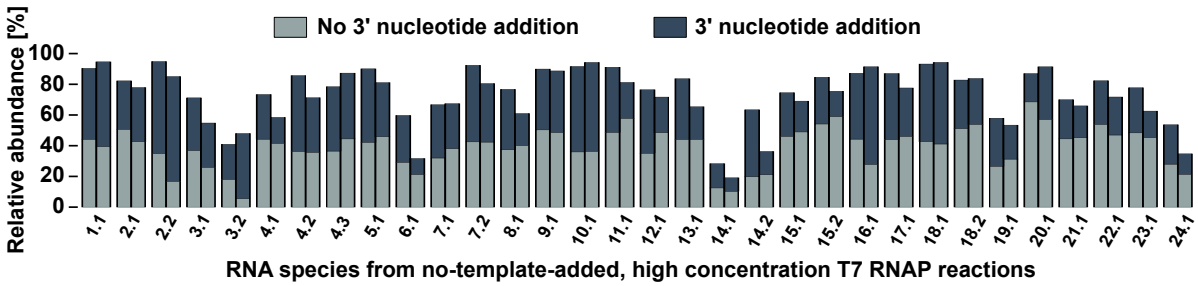


Fig. S4

Frequent addition of nucleotides at the 3' end in RNA species from no-template-added T7 RNAP reactions. RNA species from Fig. 1 further analyzed here. For each RNA reference sequence (first bar for each RNA species) and its reverse complement (second bar), the percentage of reads terminating (at positions -2, -1 and 0 from the 3' end) without further nucleotide additions ("*No 3' nucleotide addition"* in gray) is shown alongside the percentage of reads terminating with nucleotide additions ("*3' nucleotide addition"* in navy).

Fig. S5. Nimit Jain et al.

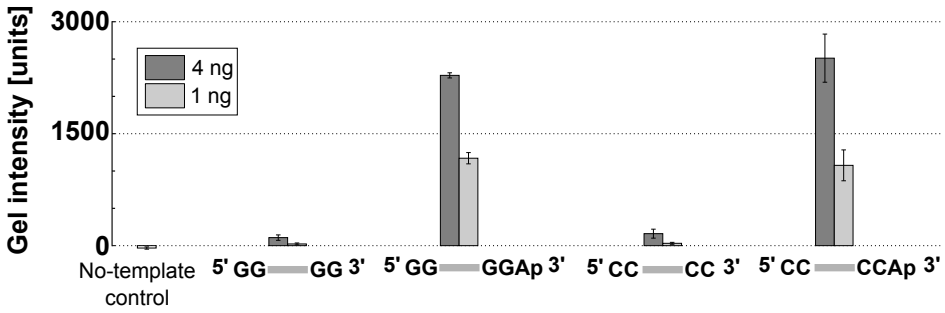


Fig. S5

Quantification of gel data from the experiment in Fig. 2A. Bar plot shows background-subtracted average gel intensity for duplicate reactions for each template type, with the whiskers representing the range of the duplicates.

Fig. S6. Nimit Jain et al.

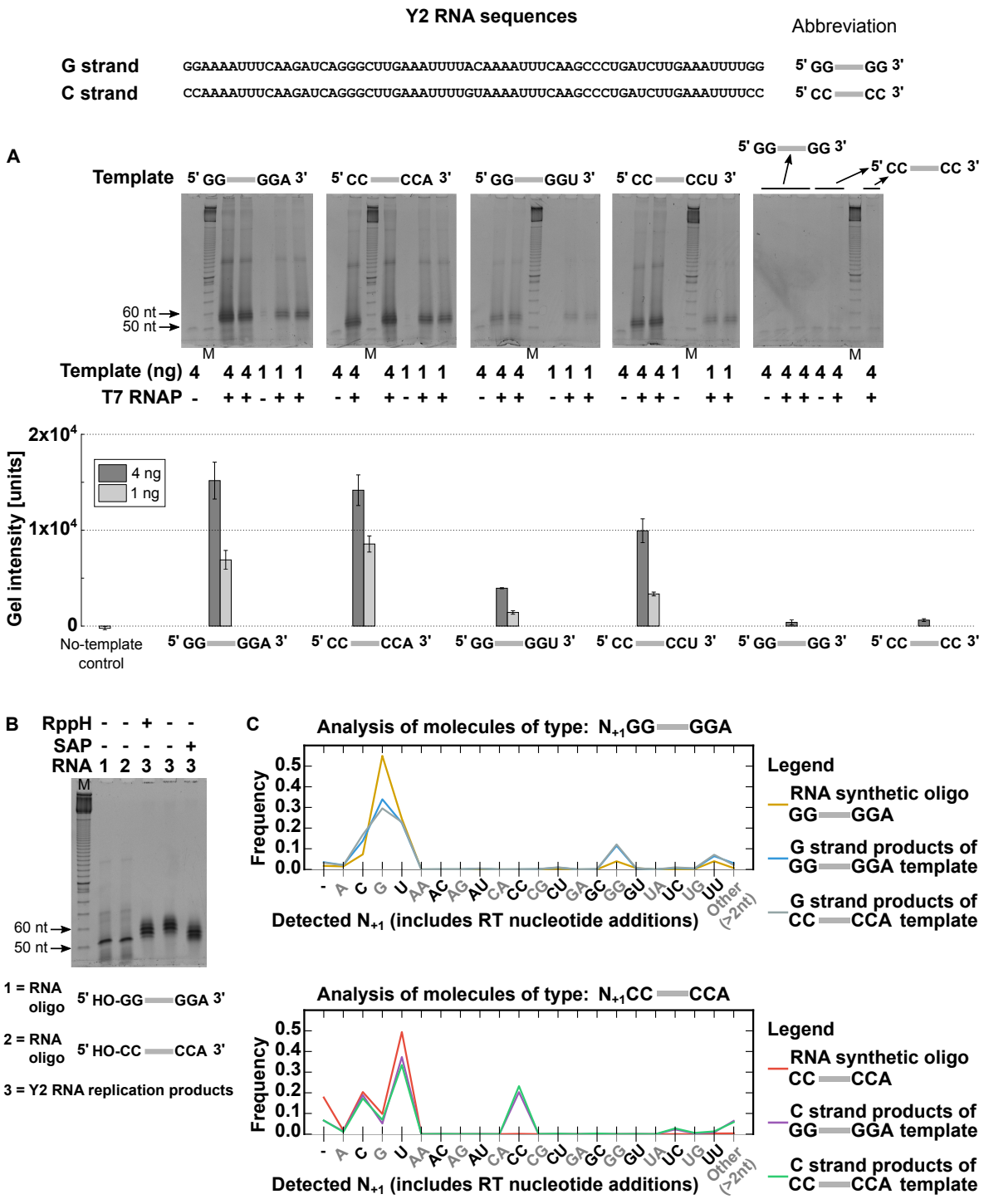


Fig. S6

Role of 3' nucleotide additions in RNA replication by T7 RNAP. (**A**) Gel-based assay showing increased T7 RNAP reaction products after chemical addition of a single adenosine monophosphate or uridine monophosphate to the 3' ends of Y2 RNA G and C strands. M =

marker (denatured 10 base-pair DNA ladder), nt = nucleotides, ng = nanograms. All gels were processed in parallel. Bar plot shows background-subtracted average gel intensity for duplicate reactions for each template type, with the whiskers representing the range of the duplicates. (**B**) The RNA 5' chemical end partly accounts for differences in electrophoretic mobility between Y2 RNA replication products (5'-triphosphate) and chemically synthesized Y2 RNA oligos (5'-hydroxyl). RppH = RNA 5' Pyrophosphohydrolase, SAP = Shrimp Alkaline Phosphatase, M = marker (denatured 10 base-pair DNA ladder), nt = nucleotides, OH = hydroxyl. (**C**) Sequence distributions at 5' ends of Y2 RNA synthetic oligos and Y2 RNA replication products. Complementary strand products (e.g. G strand products of CC—CCA template or C strand products of GG—GGA template) do not contain 5' uracil above background levels observed for synthetic oligos, supporting a subterminal initiation model over terminal initiation. A background of 5' nucleotide extensions in the detected sequences was expected from reverse transcriptase activity during RNA-Seq library preparation. RT = reverse transcriptase.

Fig. S7. Nimit Jain et al.

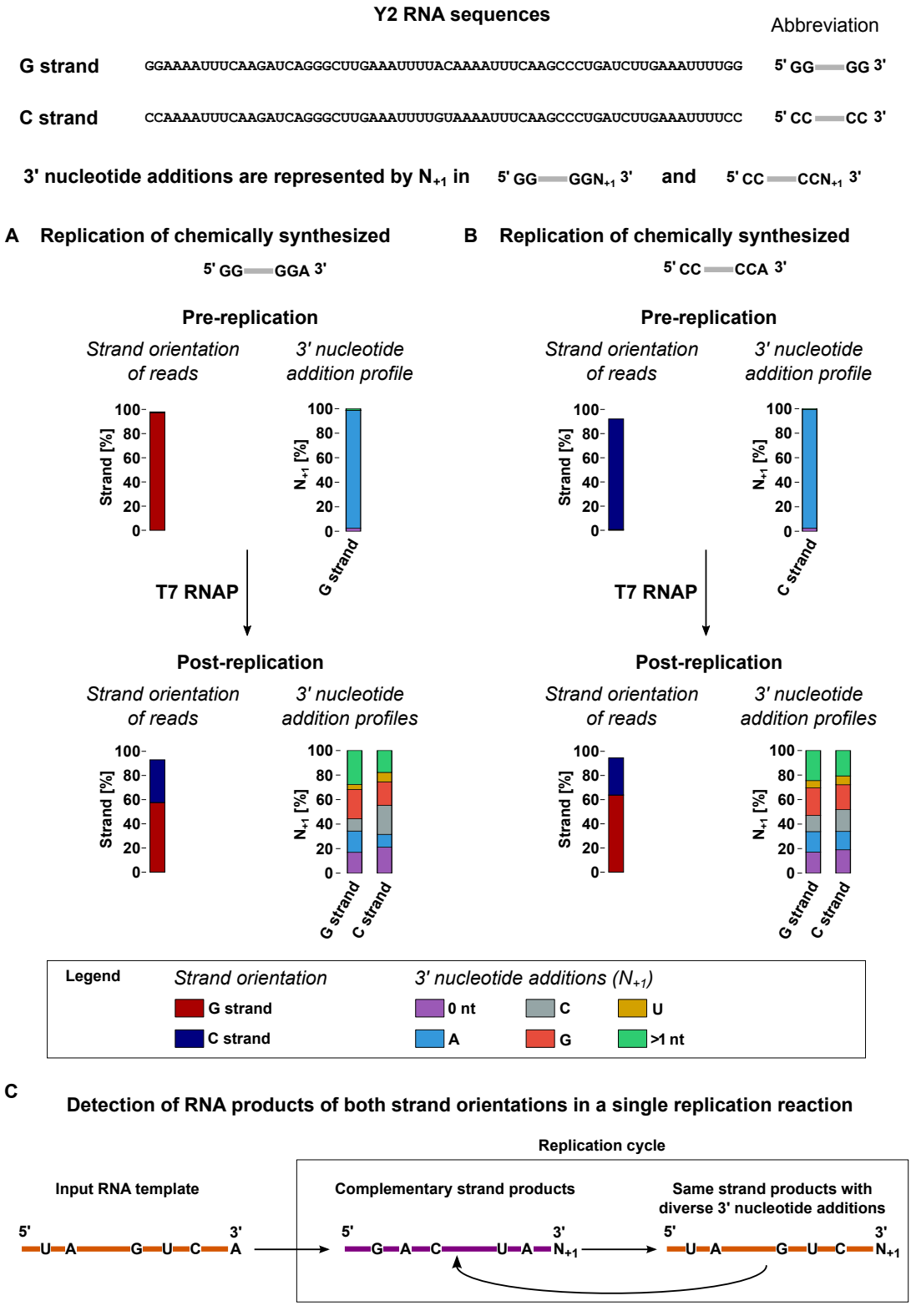


Fig. S7

High-throughput sequencing-based readout showing a key signature of RNA replication: synthesis of RNA molecules of both strand orientations in the same reaction starting with (**A**) chemically synthesized Y2 RNA G strand with an extra 3’ adenosine monophosphate or (**B**) chemically synthesized Y2 RNA C strand with an extra 3’ adenosine monophosphate. nt=nucleotides. (**C**) Schematic to explain how newly synthesized RNA products of both strand orientations can be identified in the same T7 RNAP reaction.

Fig. S8. Nimit Jain et al.

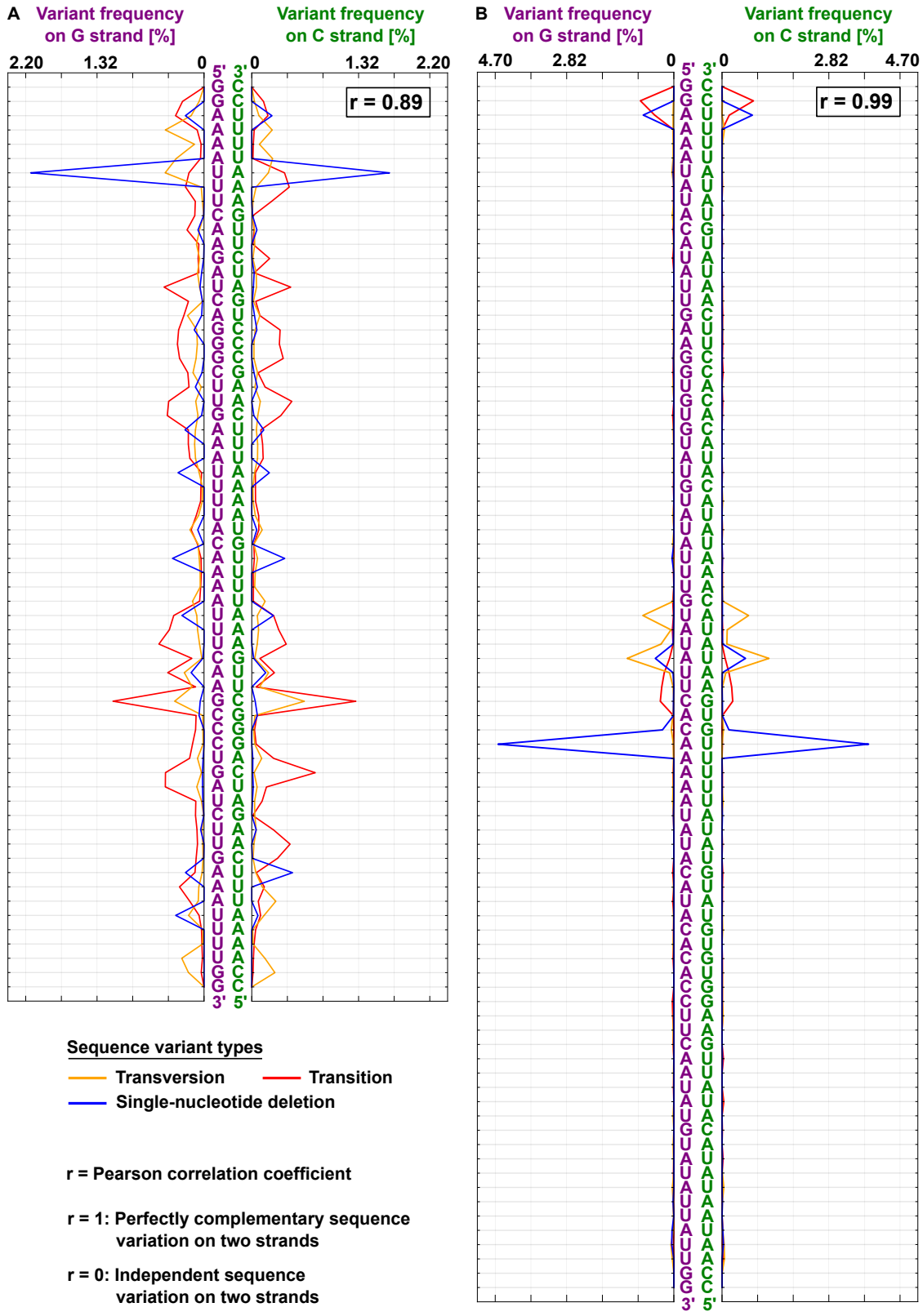


Fig. S8

Replicating RNA populations consist of multiple replication-competent sequences. We assessed replication of RNAs with sequence variants compared to reference G and C strands. Copying of RNAs with sequence variants is expected to result in complementary sequence variant profiles for the two replicating RNA strands. The degree of complementarity may be quantitatively assessed using the sample Pearson correlation coefficient. Plots in (A) and (B) show the distributions of sequence variants for two amplified RNA populations: (**A**) RNA species obtained from a templated T7 RNAP reaction starting with chemically synthesized Y2 RNA G strand with an extra 3' adenosine monophosphate, and (**B**) RNA species 2.1 from Fig. 1. Full-length sequences obtained from the RNA populations were used for analysis. Frequencies at which sequence variants were detected are shown per position for three distinct types of variants: transitions (A→G, C→U, G→A, U→C), transversions (A→C or U, C→A or G, G→C or U, U→A or G) and single-nucleotide deletions. Antiparallel symmetry between the sequence variants on the two strands (complementary variation) and values close to 1 for the sample Pearson correlation coefficient support the hypothesis that templates with sequence variants can be replicated by T7 RNAP. 95% confidence intervals for the sample Pearson correlation coefficient were estimated by non-parametric bootstrapping to be 0.76-0.96 for the RNA population in (A) and 0.96-0.999 for the population in (B).

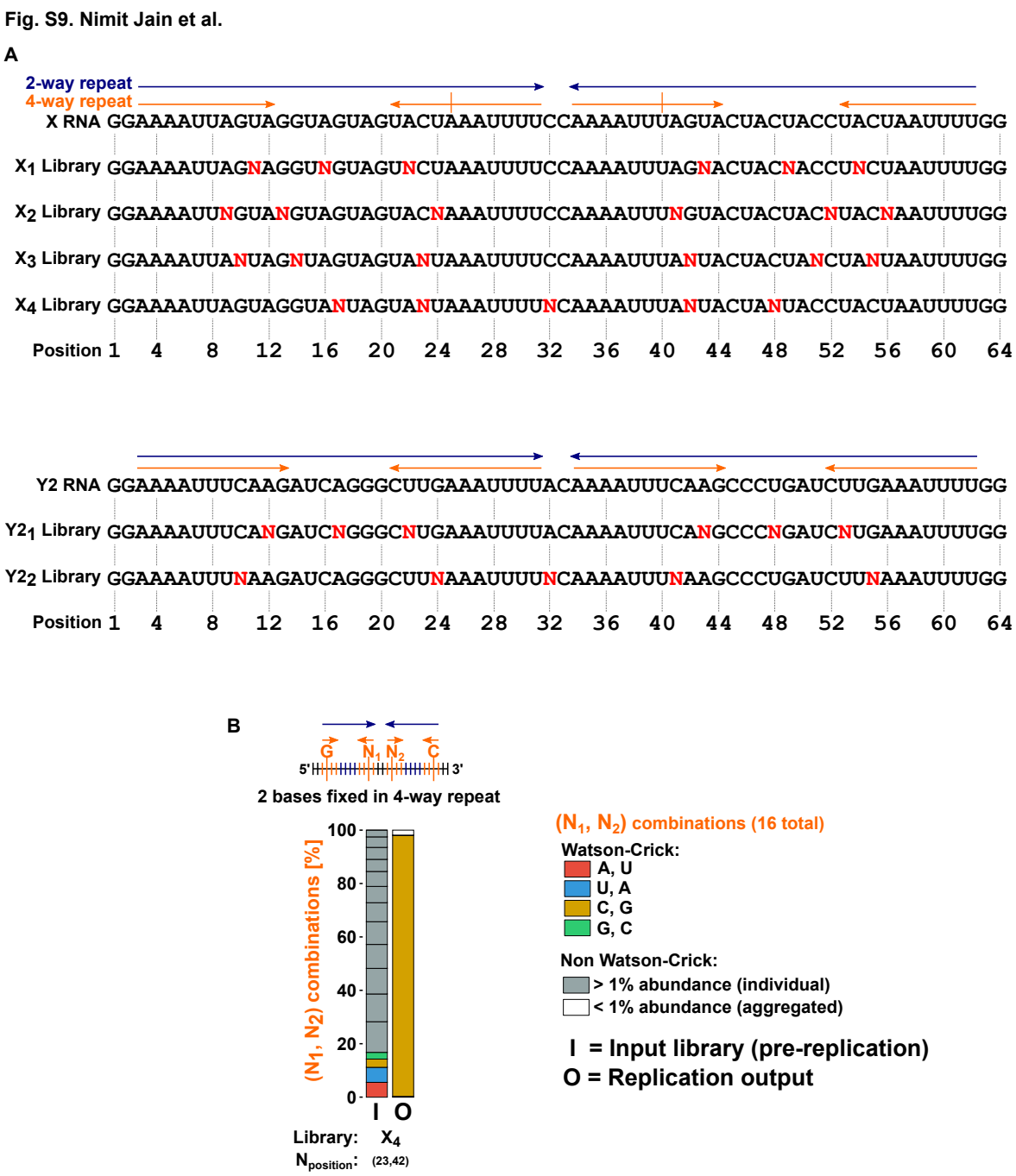


Fig. S9

Test of the function of 2-way- and 4-way- repeats in RNA replication by T7 RNAP. (**A**) Six degenerate libraries (X_1-X_4, $Y2_1$-$Y2_2$) were constructed by randomizing the base identities at a subset of sequence positions in either X RNA or Y2 RNA. G strand sequences for X and Y2 RNA are shown, with putative 2-way- (blue) and 4-way- (orange) repeats. X RNA has an imperfect 4-way repeat (vertical orange bars show sequence insertions). Positions chosen for base randomization in X RNA and Y2 RNA are shown for each degenerate library as "N" in red.

(**B**) Another test of 4-way repeat requirement that supplements the results of Fig. 3B. The X RNA-derived degenerate library X_4 contained randomized bases (denoted by "N_1" and "N_2") at only two of the four potentially base pairing positions in the 4-way repeat (specifically, positions 23 and 42 of X RNA were chosen for sequence randomization). The X_4 library was used as template in a T7 RNAP reaction, and RNA populations before replication (input "I") and after replication (output "O") were sequenced. Post-replication, only the 2 base combination (C, G) was dominant at the positions with initially randomized bases, leading to the 4-way Watson-Crick base combination (G,C,G,C) in the 4-way repeat. Color coding of bar graphs: different Watson-Crick base combinations are shown in unique colors; abundant (>1%) non-Watson-Crick base combinations are shown individually in gray; infrequent (<1%) non-Watson-Crick base combinations are summed together and shown in white.

Fig. S10. Nimit Jain et al.

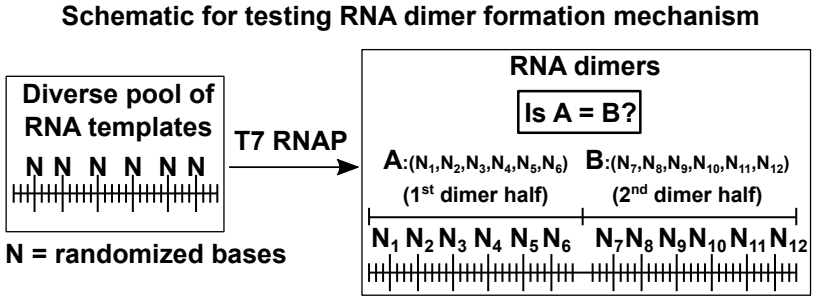


Fig. S10

Experimental scheme to assess mechanisms for RNA dimer synthesis. When RNA dimers are obtained using a diversity of monomer sequences in the same T7 RNAP reaction, uni- and bi-templated synthesis mechanisms have distinct predictions for sequence agreement between the two halves of RNA dimers (half 1=half 2 for uni-templated synthesis; half 1=half 2 in proportion to the template concentration for bi-templated synthesis). Experiments were performed in duplicate with each of two starting diverse monomer pools, X_1 and $Y2_1$. Each monomer pool contained randomized bases at a distinct set of six positions (denoted by "N"). Base identities at these six positions were used for calculating sequence agreement between the two dimer halves.

Fig. S11. Nimit Jain et al.

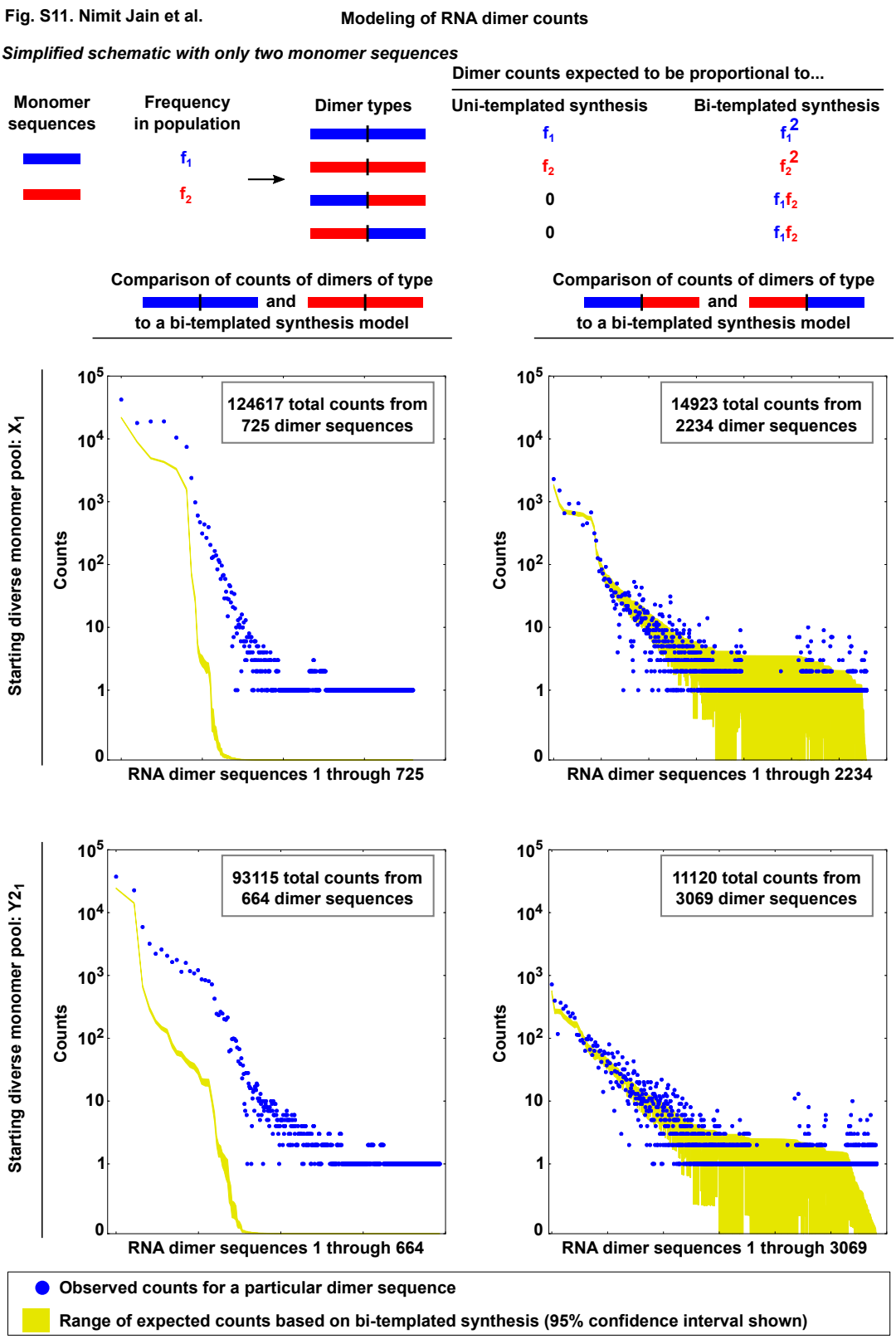
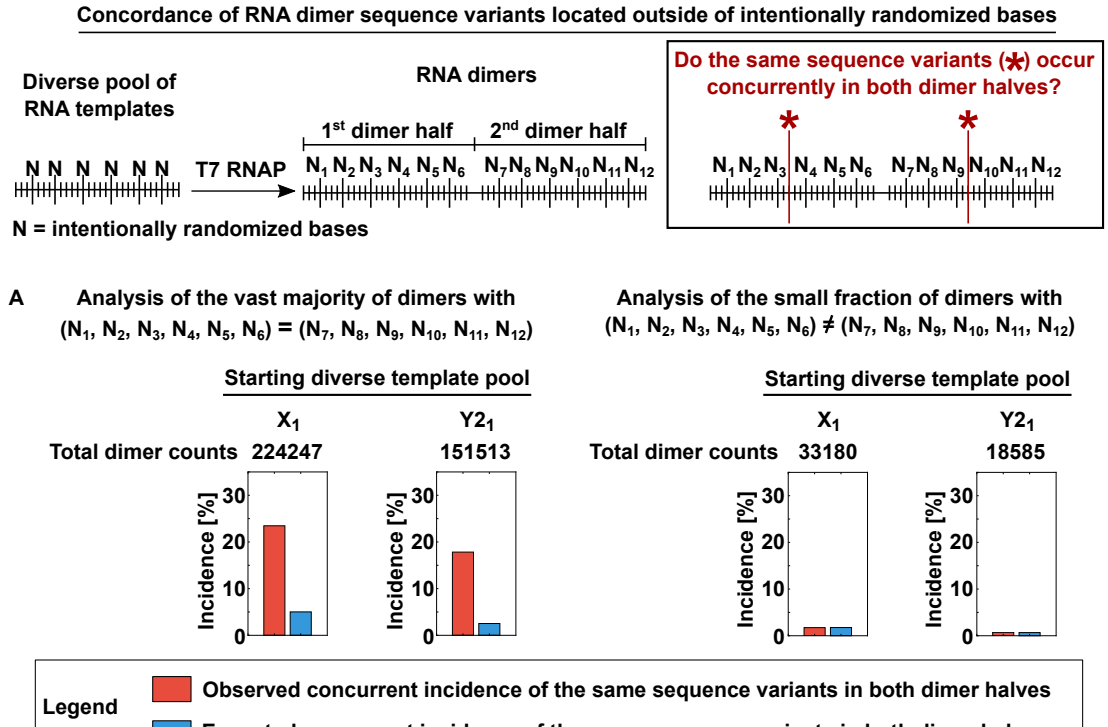


Fig. S11

Further evidence for uni-templated synthesis being the dominant mechanism for generation of RNA dimers. This figure supplements Fig. 4B. In the schematic at the top, blue- and red- colored bars represent different monomer sequences, which may have one or more mismatches with respect to each other. Observed counts shown are for dimers obtained starting with the diverse monomer template pools X_1 and $Y2_1$. Each monomer pool contained randomized bases at a distinct set of six positions. Individual dimer sequences are plotted at different coordinates along the x axis. The vast majority of dimer sequences obtained from the X_1 and $Y2_1$ pools were concordant, i.e. had perfect sequence agreement between the first and second dimer halves. The observed counts for these concordant dimers are shown in the left plots (each blue dot represents a particular dimer sequence), along with the 95% confidence interval for counts expected from bi-templated synthesis generating the concordant dimers (yellow area). The consistent overrepresentation of observed concordant dimer counts over expected counts, across a diversity of dimer sequences, supports a uni-templated mechanism. Conversely, such overrepresentation was not observed when analysis was performed on the small fraction of dimer sequences where there was sequence disagreement between the first and second dimer halves (plots on the right).

Fig. S12. Nimit Jain et al.



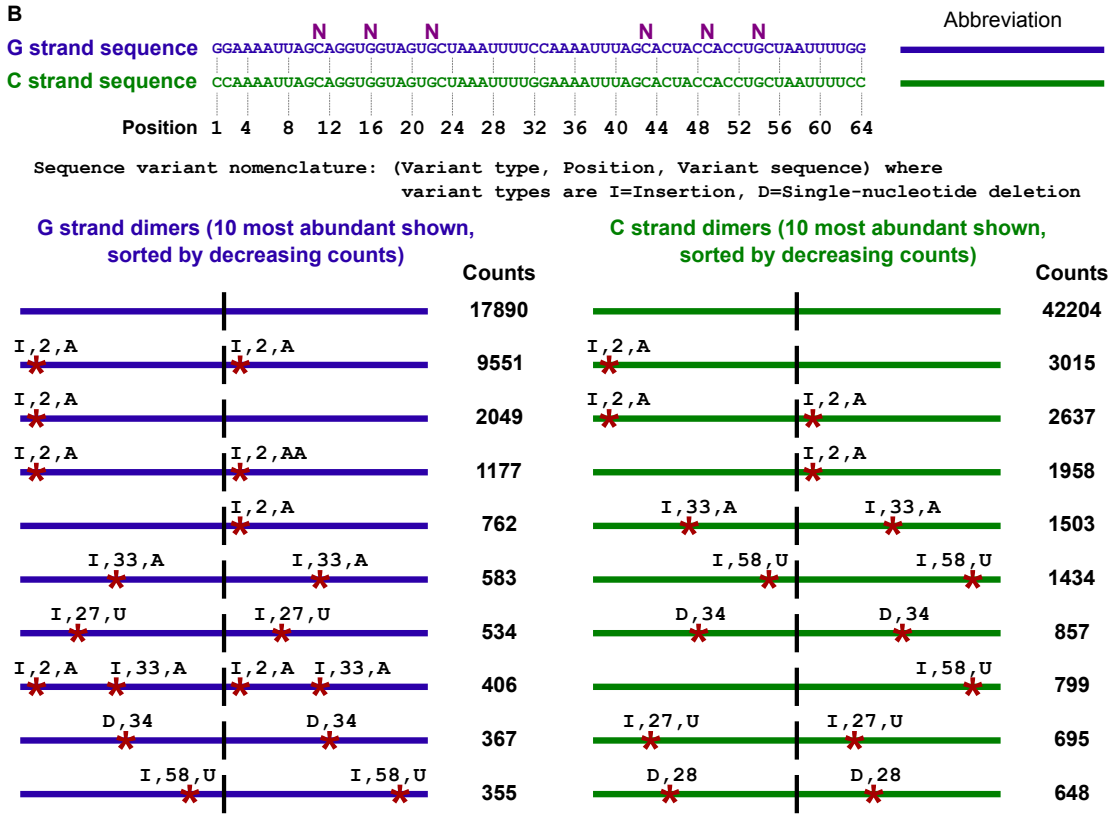
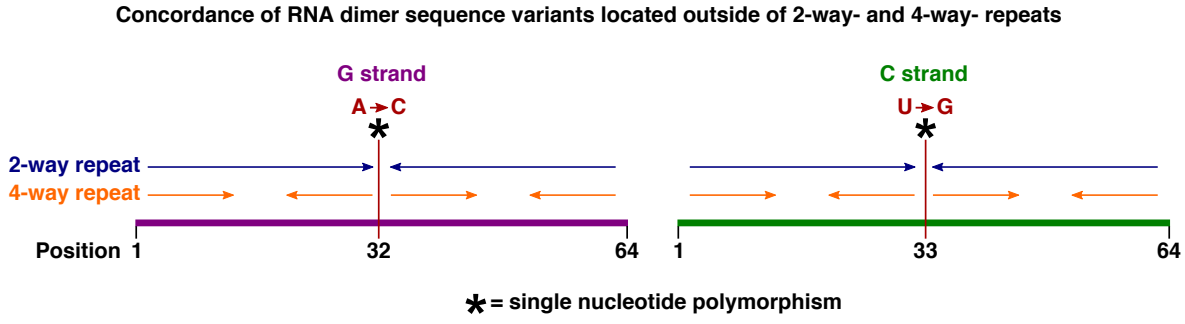


Fig. S12

Uni-templated synthesis of RNA dimers is further supported by concordance of sequence variants between dimer halves. RNA dimers were obtained starting with the diverse monomer template pools X_1 and $Y2_1$; each pool contained intentionally randomized bases at a distinct set of six positions (denoted by "N"). For this figure, sequence variants refer to polymorphisms in RNA dimers located outside the intentionally randomized bases. (**A**) Cumulative analysis of RNA dimers for sequence variant concordance. Plots on the left show analysis for the vast majority of dimers obtained from the X_1 and $Y2_1$ pools, with perfect sequence agreement between the six randomized base positions in the two dimer halves. For such dimers, the observed concurrent incidence of the same sequence variants in both dimer halves (red bars) was more frequent by 4.5 fold (X_1 pool) or 7 fold ($Y2_1$ pool) compared to the null hypothesis* (blue bars). Conversely, increased concurrent incidence of sequence variants compared to the null hypothesis* was not evident when analysis was performed on the small fraction of dimer sequences with sequence disagreement between the six randomized base positions in the two dimer halves (plots on the right). * = Null hypothesis was that sequence variants occur concurrently in the two dimer halves by random chance based on the frequencies of the sequence variants in the population. (**B**) Concurrence of sequence variants in the two dimer halves is also evident when all dimer sequences are listed in decreasing order of counts. To exemplify this, the ten most abundant G strand (blue) and C strand (green) dimer sequences are shown for an RNA template from the X_1 pool. The "N" in purple above the sequences shows the positions of intentionally randomized bases in the X_1 pool.

Fig. S13. Nimit Jain et al.



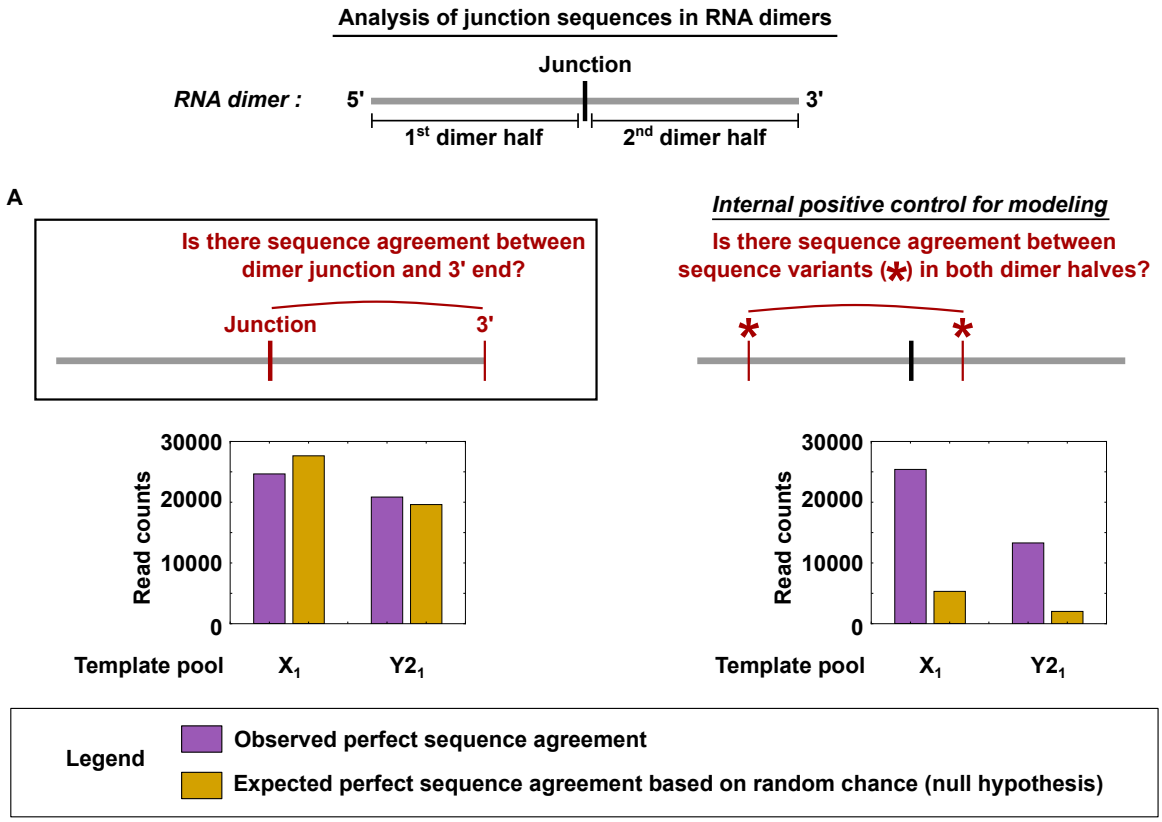
Summary of dimer counts

Template				
(1)	18	40	0	0
(2)	658	267	1	0
(3)	105	241	0	0
(4)	181	647	0	1
(1)	144	186	0	0
(2)	1922	618	1	3
(3)	393	527	0	0
(4)	480	1133	1	1

Fig. S13

Additional evidence for concordance of sequence variants between RNA dimer halves. Also see fig. S12. Dimer counts are shown for four templates from the $Y2_1$ diverse pool (template sequences in table S4). Counts are categorized based on the presence or absence of specific sequence variants (A→C at position 32 on G strand, U→G at position 33 on C strand) in the two dimer halves. Purple = G strand, Green = C strand.

Fig. S14. Nimit Jain et al.



B

Abbreviation

G strand sequence GGAAAAUUAGCAGGUGGUAGUGCUAAAUUUUCCAAAAUUUAGCACUACCACCUGCUAAUUUUGG

C strand sequence CCAAAAUUAGCAGGUGGUAGUGCUAAAUUUUGGAAAAUUUAGCACUACCACCUGCUAAUUUUCC

G strand dimers (10 most abundant shown, sorted by decreasing counts)

Junction	3' end	Counts
UU GGGGG AA UU	GGA	517
UU GGGGG AA UU	GGG	376
UU GGGGG AA UU	GG	320
UU GGGGG AA UU	GGGG	307
UU GGGGG AA UU	GGC	274
UU GGGUGG AA UU	GGA	247
UU GGGAGG AA UU	GGA	241
UU GGCGG AA UU	GGA	239
UU GGGUGG AA UU	GGG	238
UU GGUGG AA UU	GGG	224

C strand dimers (10 most abundant shown, sorted by decreasing counts)

Junction	3' end	Counts
UU CCUCCC AA UU	CC	3401
UU CCUCCC AA UU	CCC	3198
UU CCUCCC AA UU	CCA	1680
UU CCCCC AA UU	CC	1467
UU CCUCCC AA UU	CCG	1434
UU CCCCCC AA UU	CC	930
UU CCUCCC AA UU	CCCC	749
UU CCCCC AA UU	CCA	686
UU CCCCC AA UU	CCC	680
UU CCCCC AA UU	CCG	605

Fig. S14

Analysis of junction sequences between the two halves of RNA dimers. (**A**) Observed sequence agreement (purple bars in left plot) between the dimer junction and 3' end was close to what would be expected based on the junction sequence distribution and 3' end sequence distribution being independent of each other (mustard bars in left plot). Data shown are for dimers obtained starting from the X_1 and $Y2_1$ diverse RNA monomer pools. Each pool contained intentionally randomized bases at a distinct set of six positions. Dimers used for analysis here had perfect sequence agreement between the six randomized base positions in the two dimer halves. The greater-than-expected concordance of sequence variants (located outside the intentionally randomized bases) between RNA dimer halves served as an internal positive control (based on figs. S12 and S13 results) for our sequence agreement calculations (right plot). (**B**) Dimer junction and 3' end sequences for an example RNA template from the X_1 pool. G strand sequence of the example template shown in blue and C strand sequence in green.

Fig. S15. Nimit Jain et al.

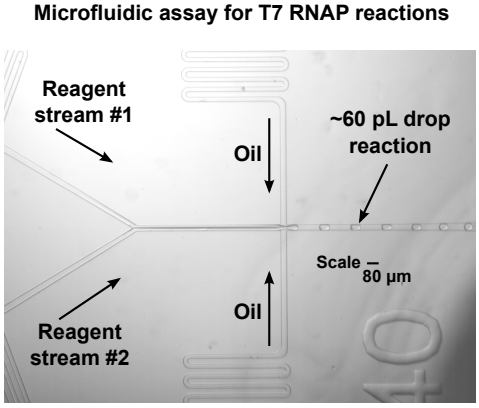


Fig. S15

Microfluidic drop generation setup for T7 RNAP-catalyzed RNA replication reactions. One reagent stream was used to flow in nucleoside triphosphates (NTPs) and when stated, RNA or DNA templates. The other reagent stream was used to flow in T7 RNAP. pL = picoliters, µm = microns.

Fig. S16. Nimit Jain et al.

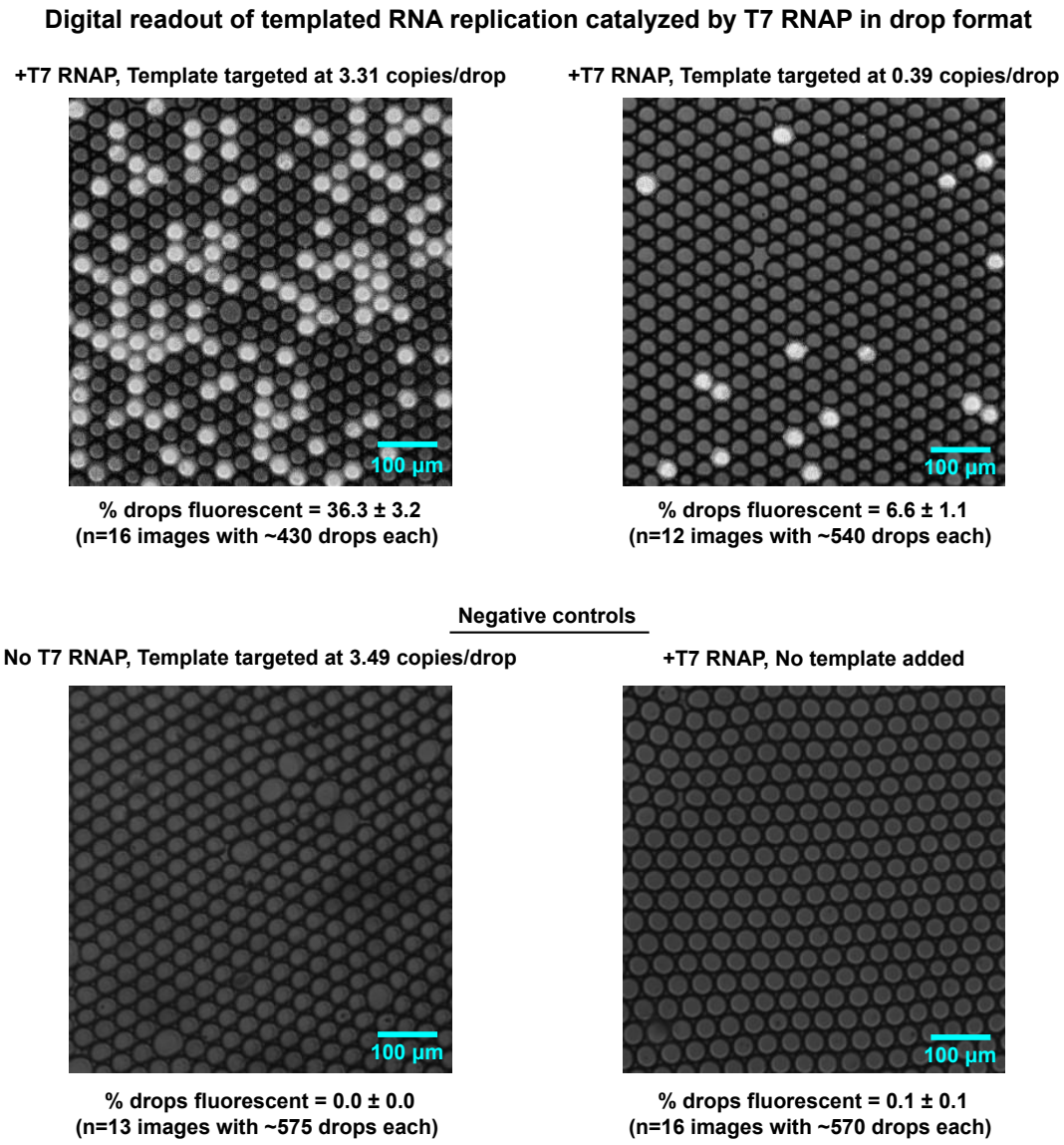


Fig. S16

Digital droplet RNA replication. Chemically synthesized Y2 RNA G strand with an extra 3' adenosine monophosphate was used as template. Reactions were conducted at low concentration of T7 RNAP (table S8). Bright, fluorescent drops contain a high concentration of RNA, consistent with RNA replication. % drops fluorescent reported as (mean ± standard deviation). µm = microns.

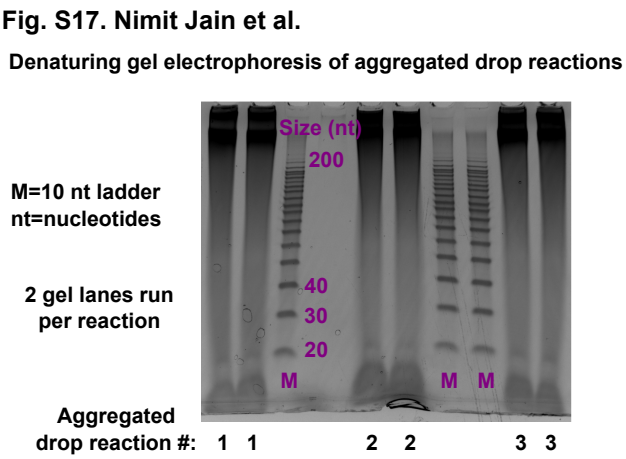


Fig. S17

Migration of aggregated drop reactions on denaturing gels. Reactions were conducted at high concentration of T7 RNAP (table S8). Aggregated drop reactions shown correspond to: (i) no-template-added (reaction 1), (ii) seeded with a DNA pool consisting of DNA from nematodes, yeast, phage and a plasmid (reaction 2), and (iii) seeded with the DNA pool, with the DNA pool having been pre-treated with DNase (reaction 3).

Fig. S18. Nimit Jain et al.

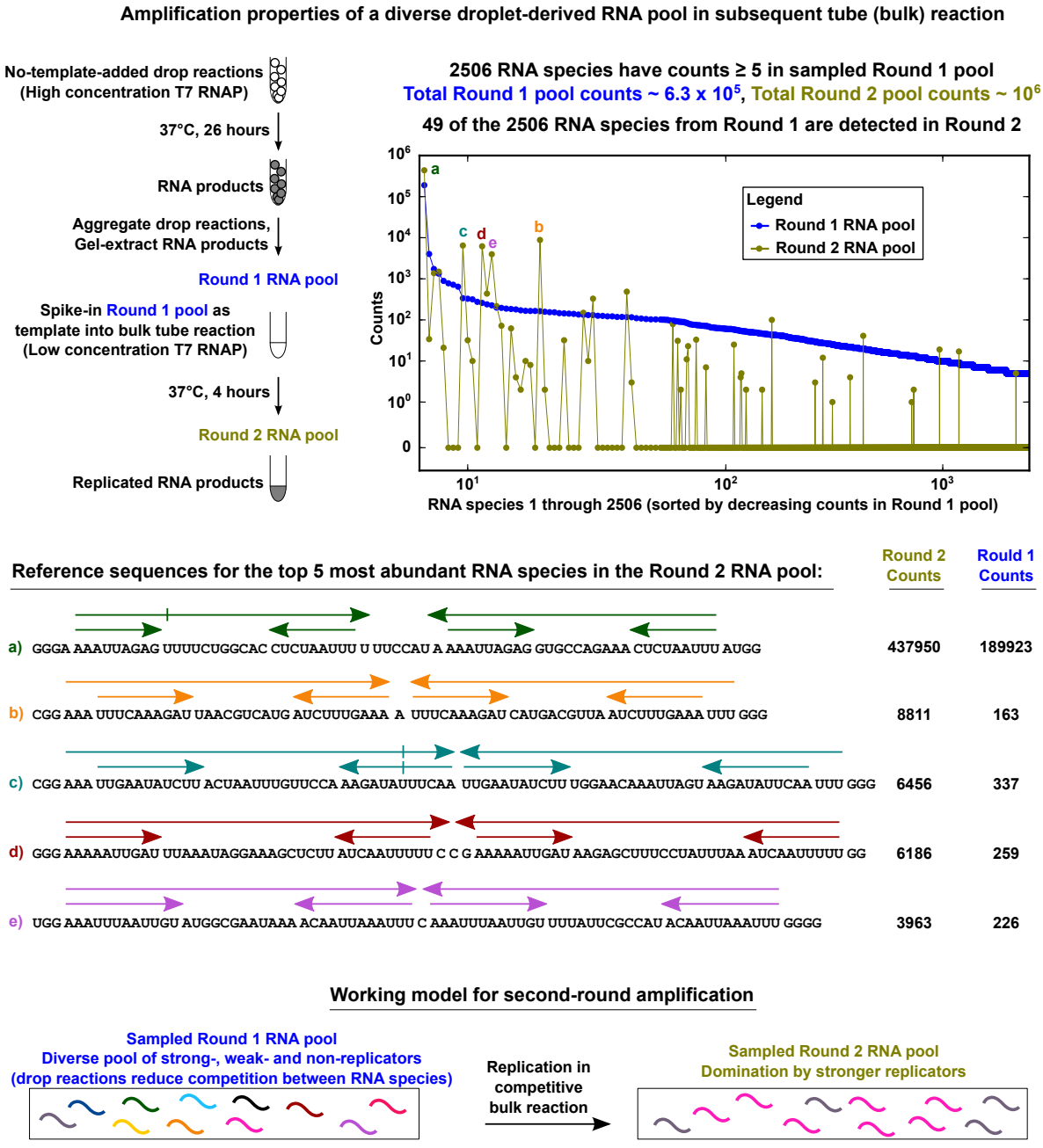


Fig. S18

Novel replicating RNAs can be isolated from no-template-added, high concentration T7 RNAP reactions set up in microfluidic droplets. A gel-extracted sample of aggregated drop reactions (*Round 1 RNA pool*) was used in bulk as template in a 10 µl low concentration T7 RNAP reaction (products called *Round 2 RNA pool*). Both *Round 1*- and *Round 2- RNA pools* were characterized by RNA-Seq. As expected from competition between RNA species during amplification with the *Round 1 pool* as template, most RNA species from the *Round 1 pool* were not detected in the *Round 2 pool*. The predominance in the *Round 2 pool* of a small subset of

species from the *Round 1 pool* demonstrates the capability of this subset of species to replicate and further, to amplify more efficiently compared to the other species from the *Round 1 pool* that were not detected in the *Round 2 pool*. In addition, the predominant species in the *Round 2 pool* exhibited typical sequence and structural hallmarks of RNAs replicated by T7 RNAP (e.g. 2-way repeats and 4-way repeats). The top five most abundant RNA species in the *Round 2 pool* are shown as examples. Arrows above each RNA sequence represent 2-way- and 4-way- repeats, with vertical bars along the arrows indicating sequence disagreements between repeat units.

Fig. S19. Nimit Jain et al.

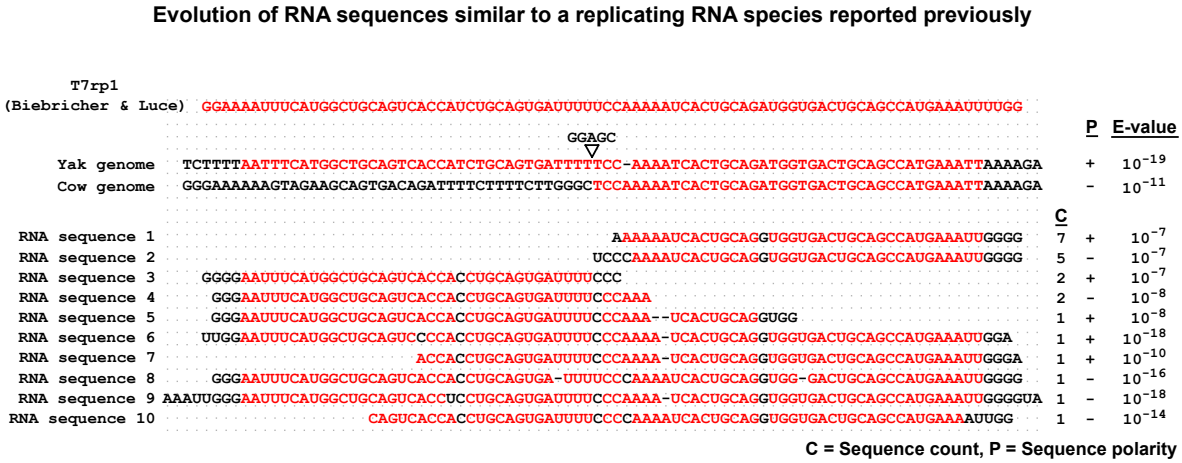


Fig. S19

Evolution of RNA sequences similar to the T7rp1 replicating RNA reported by Biebricher and Luce (*5*)**.** Bases matching in alignments to T7rp1 are shown in red. Sequences with the same strand orientation as T7rp1 are assigned polarity ("P") of plus (+); sequences complementary to T7rp1 are assigned polarity of minus (-). T7rp1 strongly matches cow and yak genomes. The 10 RNA sequence examples shown were isolated as follows. First, we generated no-template-added, high concentration T7 RNAP drop reactions. Bovine serum albumin (BSA) was included in the reactions during drop generation. The drop reactions were aggregated, and a gel-extracted sample of the aggregated reactions was then used in bulk as template in a 10 µl low concentration T7 RNAP tube reaction. Sequences shown were detected from this second-round tube reaction. E-values are based on BLAST statistics (*18*) for alignment to a sequence database consisting of T7rp1 and the NCBI nt database.

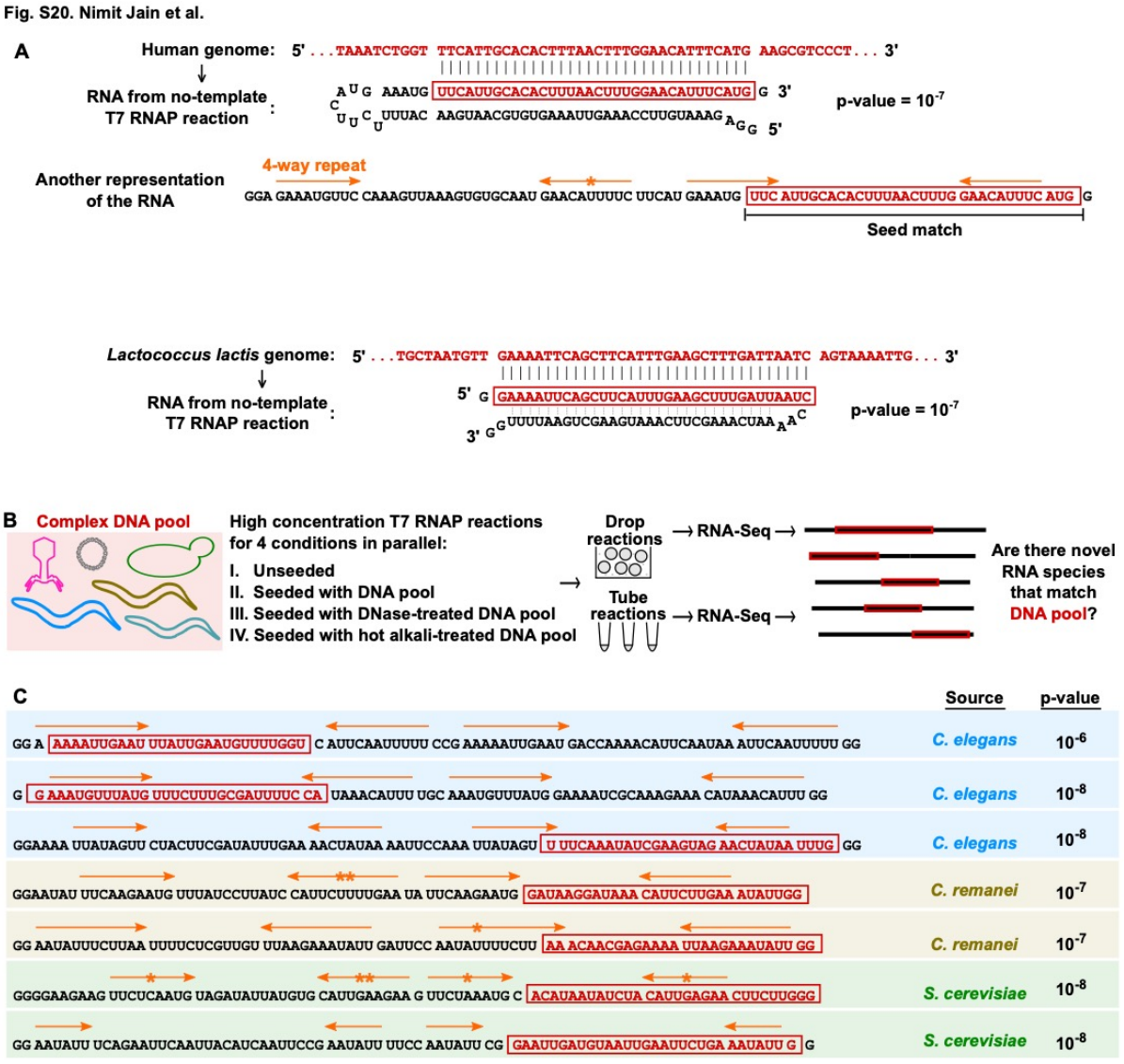


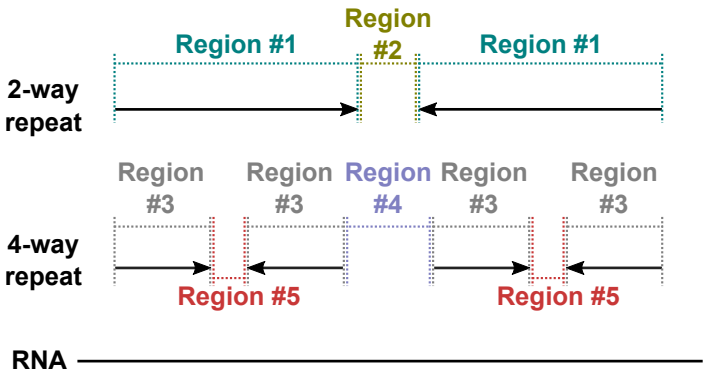
Fig. S20

RNAs replicated by T7 RNAP can originate through partial instruction from DNA seeds. (**A**) RNA species isolated from no-template-added T7 RNAP reactions can match known genomes. An RNA species matching the human genome (top) and another RNA species matching the genome of *Lactococcus lactis* (bottom) are shown. Candidate G strand sequences are shown for the RNA species. p-values are based on alignment to the RefSeq Genomic database (*19*). The long hairpin shown for each RNA species is a predicted structure. Best match to a known genome is shown in a red box. 4-way repeats are shown as orange arrows, with orange asterisks indicating sequence disagreements between 4-way repeat units. (**B**) Experimental schematic to test the hypothesis that replicating RNAs can originate through partial instruction from DNA seeds. A complex DNA pool (consisting of DNA derived from three nematode species, yeast, bacteriophage lambda and a plasmid) was used to seed high

concentration T7 RNAP reactions. Controls performed in parallel were: “unseeded,” “seeded with DNase-treated DNA pool,” and “seeded with hot alkali-treated DNA pool.” Bulk tube- and microfluidic droplet- reactions were set up in parallel for each experimental condition, followed by RNA-Seq and bioinformatic analysis of synthesized RNA species. (**C**) More examples of RNA species that originated from different sources in our designed DNA pool. This panel supplements Fig. 5C. Candidate G strand sequences are shown for RNA species (extra 3’ nucleotides apart from possibly extra guanosine monophosphates are omitted). With the exception of the third RNA listed in this panel, the shown RNA species were all isolated from drop reactions, either from the “seeded with DNA pool” condition or from the “seeded with hot alkali-treated DNA pool” condition. The third RNA example was isolated from a tube reaction for the “seeded with hot alkali-treated DNA pool” condition. Annotation of RNAs as in panel (A). Long 2-way repeats, though present in the RNA species, are not shown for simplicity. p-values are based on alignment to a database consisting of sequences in our DNA pool.

Fig. S21. Nimit Jain et al.

Quantification of AU-richness of different regions of replicating RNAs



Region	Description	% AU
1	2-way repeat	78.2 ± 6.2
2	Sequence between 2-way repeat units	63.2 ± 33.6
3	4-way repeat	82.9 ± 9.6
4	Sequence between second and third 4-way repeat units	63.7 ± 31.0
5	Sequence within 2-way repeat but between 4-way repeat units	65.3 ± 15.2

p-value table

	1	2	3	4	5
1		10^{-18}	10^{-17}	10^{-20}	10^{-52}
2			10^{-28}	0.82	0.23
3				10^{-31}	10^{-76}
4					0.32

Fig. S21

RNAs replicated by T7 RNAP have AU-rich sequence composition. %AU content (expressed as mean ± standard deviation) was calculated for different regions of replicating RNAs. Sequences from table S5 were used for analysis. p-values were calculated pair-wise between the different replicating RNA regions using a two-sided Welch's t-test for the null hypothesis that different regions have the same average %AU content.

Captions for Tables S1 to S10

Table S1. Reference sequences for the RNA species described in Fig. 1.

Table S2. Sequences of RNA species described in fig. S3.

Table S3. Kinetics of RNA synthesis in replication reactions with the X_1 and $Y2_1$ degenerate libraries as templates. Product yields were determined using gel electrophoresis by quantification of gel-image intensities within gel-lane regions expected to contain monomer-length products. A serial dilution series of reference standards was used for quantification on gels. Each time point of replication reactions was quantified up to three times, by loading different dilutions on gels. Reported yields are averages of the multiple measurements made for each time point. See table S8 for reaction conditions for X_1 and $Y2_1$ replication.

Table S4. Template sequences from the $Y2_1$ degenerate library used for the analysis in fig. S13.

Table S5. Sequences of RNA species described in Fig. 5 and fig. S20C. Candidate G strand sequences are listed. Extra 3' nucleotides apart from possibly extra guanosine monophosphates are omitted.

Table S6. Key reagents/equipment used in our study.

Table S7. Oligonucleotide sequences used in our study.

Table S8. Reaction conditions for the various sets of T7 RNAP reactions in our study.

Table S9. RNA-Seq library preparation parameters for samples in our study.

Table S10. Brief description of the functionality of the code deposited on GitHub.