Nanopore direct RNA sequencing (DRS) of MS2 bacteriophages in *E. coli* throughout its life cycles reveals a complex transcriptional activity to control and maintain its growth

Authors: Ni Ni and Gaetan Burgio

Affiliation: Division of Genome Sciences and Cancer, The Shine Dalgarno Centre for RNA Innovation, the John Curtin School of Medical Research, the Australian National University, Canberra, Australia.

Corresponding author: Email: Gaetan.burgio@anu.edu.au

Abstract

The RNA bacteriophage MS2 is an RNA phage that infects the bacterium *E. coli* and is one of the most studied and prototypical model phages in molecular biology and microbiology. Previous research revealed complex translational control and fine-tuning for MS2 replication. However, the dynamics of its transcriptional activity and replication during the life cycles within the bacteria remain elusive. Here, we employed Nanopore Direct RNA sequencing (DRS) to investigate the transcriptome and epitranscriptome landscape of the MS2 in infected *E. coli* throughout multiple life cycles. We discovered that MS2 phages sustain a high level of transcriptional activity required for replication. We found large amounts of subgenomic small transcripts from RNA degradation, Nanopore DRS bias, and transcripts containing the *coat*-encoding region, required for virion assembly. We found hybrid reads due to the error-prone activity of the MS2 replicase complex by the template-switching mechanism. We evidenced that RNA modification is conserved throughout the entire life cycle in full-length transcripts without the acquisition of new modifications, whereas small transcripts did acquire newly modified sites. The conserved sequence and secondary structure (U-rich hairpin) of Ψ installation sites were the most amenable to RNA modification, and one site aligned well with host RluA-mediated installation. Overall, our investigation reveals a more complex transcriptional dynamics of MS2 phages than anticipated within *E. coli* to maintain its growth and replicate under host pressure.

Introduction:

Viruses of bacteria (bacteriophages) are the most abundant organisms on Earth (Clokie et al., 2011). Phages play an important role in the regulation of biological processes and have been harnessed for biotechnology applications (Zhang and Wu, 2020). RNA bacteriophages (commonly known as RNA phages), such as MS2, R17, PP7 and Q-beta, are single-stranded RNA viruses that infect a variety of bacteria (Bollback and Huelsenbeck, 2001). It is thought that the ssRNA phages are the oldest lineage of RNA viruses still in existence (Wolf et al., 2018). These have proven to be useful models for comprehending basic molecular mechanisms such as RNA genome replication, translational control, and gene regulation (Poot et al., 1997; Rolfsson et al., 2016).

MS2 belongs to the family *Fiersviridae*, and is an icosahedral, non-enveloped, F-specific, positive-sense single-stranded RNA (+ssRNA) lytic coliphage (Salas & de Vega, 2008). The ssRNA phages adopt a simple infectious cycle that includes host adsorption, genome entry (genomic RNA penetration), genome replication, phage assembly, and host lysis to release new bacteriophage progeny for successive infections. MS2 phage replicase β subunit and host proteins (ribosomal protein S1 subunit, EF-Tu subunit (elongation factor thermo unstable), and EF-Ts subunit (elongation factor thermos stable)) form an RNA-dependent RNA polymerase (Rdrp) complex required for replication (**Figure 1A**) (Wagner, Weise, and Mutschler, 2022). Full-length negative single-stranded RNA intermediates are generated to serve as the template for the synthesis of positive genomic RNA (gRNA) by the replicase complex (**Figure 1A**). Among phages, ssRNA phage MS2 has one of the smallest and simplest genomes with only possessing four ORFs in its 3,569 nt genome that also serves as an mRNA: *mat* encodes maturation protein (MP), which enables attachment to the F pilus and RNA entry into the host (Meng et al., 2019); *coat* encodes the coat protein (CP), which forms the viral capsid; *lys*

encodes lysis protein (L), which allows for host lysis (Chamakura, Tran, and Young, 2017); and *rep* encodes catalytic subunit of replicase (Rep), which is an RNA-dependent RNA polymerase (Wagner, Weise, and Mutschler, 2022) **(Figure 1B)**.

Previous research revealed RNA–protein interactions that regulate the precise timing and strength of viral protein production throughout the bacteriophage life cycle (Dykeman, 2024; Twarock et al., 2024). Despite years of research on MS2 replication and assembly, little is known about the dynamics of transcription of MS2 during its life cycle. Like many other viruses, MS2 RNA encodes an error-prone replicase (Wagner, Weise, and Mutschler, 2022). The replicase complex is repressed from the CP, which binds to specific sites on the positive-strand RNA (Peabody, 1990). While the interplay between different phage proteins resulting in entry, replication, assembly, and encapsidation has been relatively well studied in *in vitro* settings (Rolfsson et al., 2016; Wagner, Weise, and Mutschler, 2022), the dynamics in transcription throughout the MS2 life cycle remain elusive. Specifically, the unanswered questions remain on how MS2 transcriptional activity rapidly increases to sustain the production of the virion, the nature and the extent of the byproducts the error-prone replicase produces during the MS2 life cycle, and how the dynamics of acquisition of *de novo* mutations and RNA modification occur in the bacterial cell.

To investigate the MS2 transcriptome dynamics across time, Direct RNA sequencing using Oxford Nanopore sequencing enables the sequencing of long RNA sequences encompassing the full-length transcript and the identification of RNA modification in infected cells. Previous work has provided valuable insights into the dynamics of transcription, RNA splicing (Kim et al., 2020), recombination (Viehweger et al., 2019), and RNA modification landscapes of various viruses (Viehweger et al., 2019).

53

54 Here, we employed Nanopore DRS to investigate the transcriptome and epitranscriptome
55 landscape of the MS2 in infected *E. coli* throughout its life cycle. We discovered that MS2
56 phages sustain a high level of transcriptional activity within the bacterial cell. We found sub-
57 genomic transcripts resulting from either RNA degradation or Nanopore DRS bias.
58 Interestingly, we found the presence of small sub-genomic transcripts encoding the full-length
59 coat protein ORF. The MS2 replicase is error-prone, which frequently generates hybrid RNAs
60 across time by the template-switching mechanism. We evidenced that RNA modification is
61 conserved throughout the entire life cycle in full-length transcripts without the acquisition of
62 new modifications, whereas small transcripts did acquire modified sites. The conserved
63 sequence and secondary structure (U-rich stem-loops) of Ψ installation sites were the most
64 amenable to RNA modification and likely mediated by host pseudouridine synthase RluA.
65 Together, using DRS RNA Nanopore sequencing, we have dissected the dynamics of MS2
66 transcriptional activity in bacterial cells across time.

67

68
69 **Methods:**
70
71 **Bacteria, phage strains, and culture conditions**

72 The bacteria host *E. coli* DSM 5695 (also called *E. coli* W1485) (F^+) and MS2 bacteriophages
73 DSM 13767 were obtained from the DSMZ-German Collection of Microorganisms and Cell
74 Culture. The bacteria were maintained in NZCYM media (10.0 g/L Casein hydrolysate, 5.0 g/L
75 NaCl, 1.0 g/L Casamino acids, 5.0g/L Yeast extract, 2.0 g/L MgSO4 X 7H2O) with maltose
76 (2g/L) and incubated at 37ºC overnight. MS2 phages were maintained and propagated in SM
77 buffer (NaCl 100 mM, MgSO4 X 7H2O 8 mM, Tris-HCl 50 mM, gelatin 0.01% (w/v), pH 7.5).
78 MS2 plaque assays were performed in a double-layer Agar. Briefly, 100 µl overnight culture
79 of DSM 5695 bacteria and 100 µl MS2 in SM buffer were mixed with 5 ml NZCYM soft agar

80 with maltose (2g/L). Plates were incubated at 37ºC overnight (~16 h). MS2 phages were
81 harvested. The top layer was disrupted after pouring 5 ml SM buffer, and the lysate was
82 centrifuged at 2,800g for 30 minutes at room temperature. The lysate was filtered using a 3 ml
83 syringe with a 0.22 μM filter. The precise viral stock titer (PFU/ml) of the phage (MS2) was
84 determined.

85

86 **MS2 phage infection and RNA isolation**

87 The bacteria *E. coli* (DSM 5695) liquid culture growth infected with bacteriophage MS2 was
88 generated by multimode plate reader IID VICTOR Nivo (PerkinElmer), start OD 600 of
89 overnight DSM 5695 bacteria culture was around 0.2-0.3. Each data point was the result of at
90 least three biological replicates. DSM 5695 *E. coli* cells (overnight culture) were infected with
91 MS2 at a multiplicity of infection (MOI) of 200 and incubated for 6 h at 37 °C. Infected DSM
92 5695 cells and non-infected DSM 5695 cells (i.e., mock-infected cells) were harvested (~10^{10}
93 bacterial cells were prepared as the initial material) at 0h, 20 and 40 minutes, 3h, and 6h post-
94 infection (p.i.). At the 0h time point, immediately after adding the phages, we isolated the total
95 RNA, and the bacterial cells were not washed. Total RNAs were extracted from infected *E.*
96 *coli* samples (TRIzol [Invitrogen]) following the manufacturer's instructions.

97

98 **Reverse transcription quantitative PCR (RT-qPCR) of MS2 phage genes**
99 The primer pairs for RT-qPCR were designed specifically to the MS2 gene regions using
100 RefSeq Representative Genome Database (Organism limited to *Escherichia coli*) with the
101 NCBI Primer-BLAST tool. The *E. coli cysG* gene (housekeeping gene) was selected as the
102 reference gene. Briefly, reverse transcription for 100 ng whole RNA extracted was performed

103 with SuperScript™ IV Reverse Transcriptase [Invitrogen] with random primers (hexamers)
104 [Invitrogen] at 50°C, then qPCR was performed with PowerTrack™ SYBR Green [Invitrogen]
105 using specific sets of primers (**Suppl. Table S12**) for specific phage genes (400 nM primers
106 with annealing temperature 60°C) with input cDNA (serially diluted to determine the qPCR
107 primer efficiency) according to the manufacturer's protocol an visualized by QuantStudio™
108 12K Flex Real–Time PCR System. Data analysis using the ΔΔCT method was performed to
109 quantify the relative expression of MS2 phage genes.

110

111 ***In vitro* transcription of MS2 RNA**

112 One µg of total RNA was purified from MS2-infected *E. coli* cells. Reverse transcription
113 (SuperScript™ IV Reverse Transcriptase [Invitrogen]) was performed with each phage-
114 specific RT primer (RTprimer1-6) (**Suppl. Table S12**). Templates for *in vitro* transcription were
115 prepared by PCR with each phage-specific PCR primer pair (i.e., IVT-frag1-F & IVT-frag1-R
116 primers with 5' terminal T7 RNA polymerase promoter region) followed by agarose gel
117 purification (Wizard SV Gel and PCR Clean-Up System kit [Promega]), *in vitro* transcription
118 using T7 polymerase (MEGAshortscript™ T7 Transcription Kit [Thermo Fisher]) and RNA
119 purification (Monarch RNA Cleanup Kit [NEW ENGLAND Biolabs]) with re-suspension in
120 ddH2O. The oligonucleotides used in this study are listed in the **Suppl. Table S12**.

121

122 **Poly(A) tailing of RNA and beads purification**

123 Total RNAs and IVT RNAs were heat incubated at 65°C for three min and cooled on ice for
124 one min before polyadenylating RNAs using the *E. coli* poly(A) polymerase (NEB M0276S
125 [NEW ENGLAND Biolabs]). Briefly, 10 µg RNA, 5 units poly(A) polymerase, 2 µL 10X
126 reaction buffer, and 1 mM ATP in a total reaction volume of 20 µL were incubated for 45 min

127 at 37°C. To stop and clean up the reaction, poly(A)-tailed RNAs were purified following the
128 Agencourt AMPure XP magnetic beads (1:1.8 RNA to beads ratio) according to the
129 manufacturer's protocol and resuspended in 10 µL of ddH2O. The integrity of total RNA from
130 *E. coli* was assessed via a Bioanalyzer (Agilent).
131
132 **Direct RNA nanopore sequencing**
133 The Direct RNA Sequencing Kit (SQK-RNA004 by Oxford Nanopore) was used to prepare
134 the libraries and sequence the native poly(A)-tailed RNA library according to the
135 manufacturer's protocol and including the reverse transcription step to generate RNA:cDNA
136 hybrids. Sequencing runs were performed using PromethION flow cells (FLO-PRO004RA).
137 The Flow Cell Wash Kit (EXP-WSH004) was used to remove the library between flow cell
138 washes and library reloading, following the manufacturer's instructions.
139
140 **Basecalling and read mapping**
141 The Computational workflow for processing the Oxford RNA-direct Nanopore sequencing
142 datasets was shown (**Suppl. Figure S1**). Basecalling was performed using MinKNOW
143 software (Nanopore) using the Super accurate model. The Fast5 files were filtered. Reads with
144 a minimum PHRED score of 10 were removed. Fast5 files were then converted to pod5 using
145 pod5_file_convert script using the –one-to-one option. Basecalling on pod5 files was
146 performed with Dorado (v 0.9.1) using the rna004_130bps_sup@v5.1.0 basecalling model for
147 m5C, m6A_DRACH, Inosine_m6A, PseuU. Reads were mapped to the reference sequence
148 NC_001417.2 using minimap2 (Li, 2018) using the following parameters: -ax map-ont -y --
149 secondary=no -k 4 -w 10. Samtools (Li et al., 2009) was used to retain the mapped reads either
150 on the positive strand (Samtools view –F2324 -b) or negative strand (Samtools view –f 16 –b)
151 and reverse complemented. Reads were sorted and indexed. Mapped read lengths were counted

using genomecov function from Bedtools (Quinlan and Hall, 2010). Other metrics, such as alignment identity, were determined using MarginStats (Jain et al., 2015) and based on a similar workflow previously described (Workman et al., 2019).

Gene expression analysis

To map the reads to transcripts, a transcript reference sequence was generated using the fasta reference sequence and a gff file converted from genebank format. Reads were mapped to the transcript using minimap2 with the same options as previously. Mapped reads were counted with Salmon (Patro et al., 2017) using the options: --ont –P 48 –l U. Gene expression analysis was performed with Deseq2 (Love, Huber, and Anders, 2014) under R 4.4.0 using time as a variable. Gene expression was normalised and compared for each gene across time.

Read coverage analysis

Read length was determined with Samtools and separated into multiple fractions of similar read lengths. Reads were mapped to the NC_00141702.2 genome (DSM 13767 genome assembly lacked the 3' end of the contig, preventing its use as a reference genome) and sorted using Samtools. Reads were also mapped to the bacterial host strain DSM 5695 for read count. Read coverage within a 5 bp interval for reads starting/ending versus the total number of read at these specific window positions using Samtools and a custom bash script. Stop and start codons were determined using a motif search and overlayed to the relative coverage plot. The secondary RNA structure was determined using RNAfold. Hybrid reads were mapped to the NC_00141702.2 genome using blastn. The top hits e.values > 10^8 were retained. The query adjacent reads were retained for further analysis, and the target read locations were determined on the positive and negative strands and mapped back to the NC_00141702.2 genome.

177 **Detection of single nucleotide and structural variants**

178 To detect single nucleotide variants and indels, we performed the analysis using Clair3 (Zheng
179 et al., 2022) and Sniffles2 (Smolka et al., 2024). We included the following flags for Clair3
180 analysis: --platform="ont" --model_path=r941_prom_sup_g5014 –include_all_ctgs –
181 no_phasing_for_fa. For indels detection, we performed the analysis using Sniffles2 with
182 default parameters. The files were filtered and converted using Bcftools (Danecek et al., 2021)
183 in vcf format.

184

185 **RNA modification analysis**

186 Basecalled reads from Dorado were mapped to the NC_00141702.2 genome using Minimap2,
187 sorted, and indexed using Samtools. Basecalling for modification was determined using
188 Modkit pileup function using –max-depth 2000000 option and a modification threshold > 80%.
189 Differential methylation profile was determined using modkit dmr function. The Nanopore
190 DRS technique has the limitation of high false-positive rates (FPR), and the IVT samples were
191 prepared to reduce the FPR. Only Stoichiometry (test) - Stoichiometry (IVT) > 0 was included
192 in our study.

193

194 **Statistical analysis**

195 Statistical analysis was conducted. The mean ± standard deviation (SD) was shown. The
196 differentially expressed genes were identified by assessing the statistical significance via the
197 adjusted *p*-value after accounting for multiple hypothesis testing (Benjamini-Hochberg
198 method). The threshold for statistical significance was set at P value < 0.05, *, P value <0.01,
199 **. The Benjamini-Hochberg method was performed to control the false discovery rate (FDR
200 correction) for post-transcriptional modifications (Benjamini and Hochberg, 1995). Graphs

were generated, and data were visualized using ggplot2 from R (version 4.4.0), GraphPad Prism (version 10.4.1), and Matplotlib from Python (version 3.12.8).

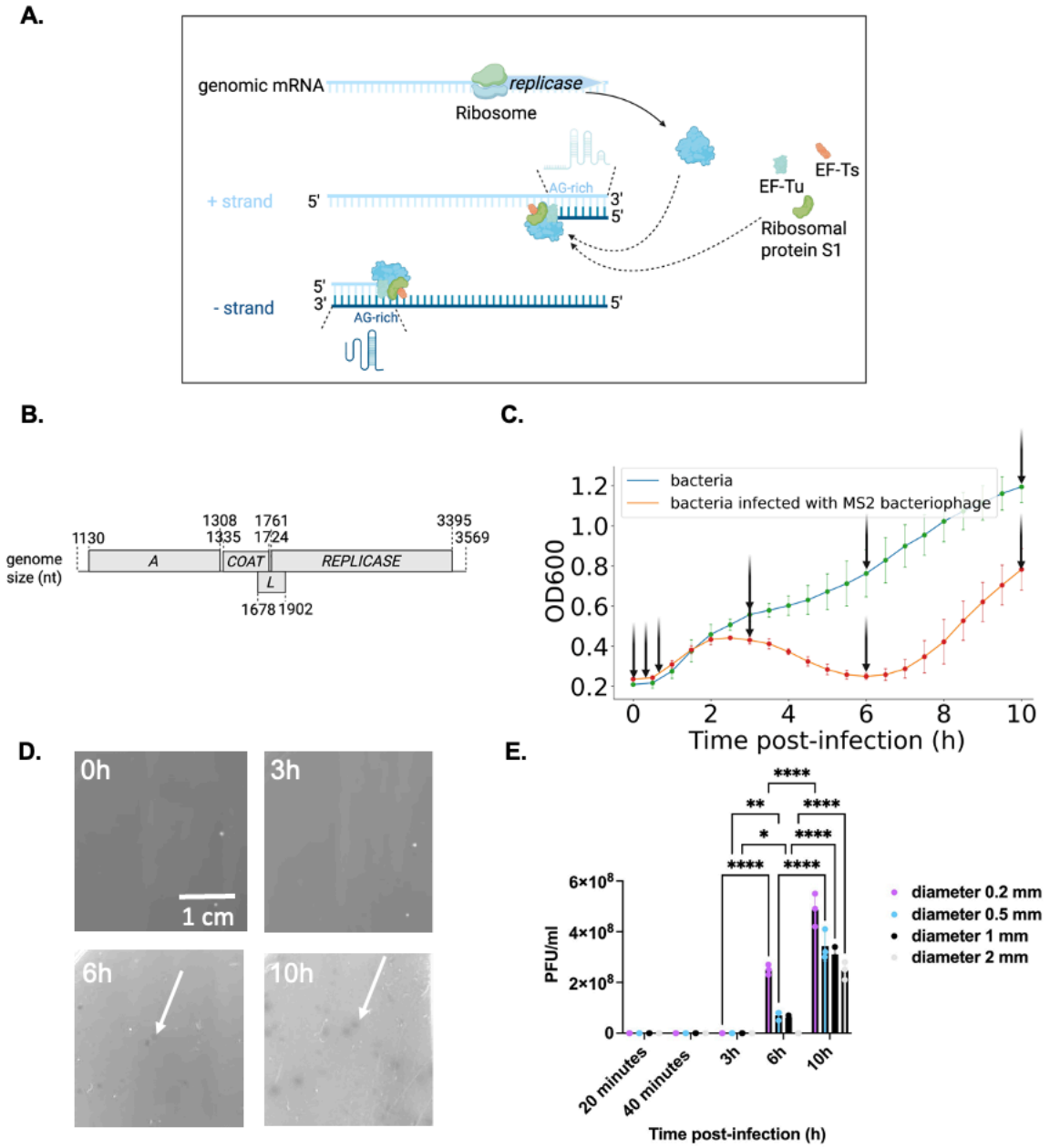
Sequencing data Files and code availability

Raw sequencing data files are available at the NCBI Sequence Read Archive (SRA) under the following accession ID: Bioproject ID PRJNA1243726 and BioSamples SAMN47628658 to SMNA47628577. The custom source codes used for the computational analysis have been deposited in the following Github repository: https://github.com/gburgio/MS2_Nanopore_DRS.

Results:

MS2 bacteriophage replicated within half an hour in *E. coli*

To determine the course of infection in *E. coli* host over 10 hours post inoculation (p.i.), we infected the bacterial cells with MS2 phage and monitored *E. coli* growth over time in liquid culture at the multiplicity of infection (m.o.i) of 200 (**Figure 1C**). MS2 formed plaques from 6 hours p.i., indicating its lytic nature (**Figure 1D**). The size of MS2 plaques ranged from 0–2 mm in diameter within 10 hours (**Figure 1 D-E**). Overall, the data indicated that the time points 20 minutes (before the first burst), 40 minutes (after the first burst), 3h (maximum growth), and 6h (lowest growth) p.i. captured a single and multiple life cycles, resulting in plaque formation. We chose these time-points to assess the dynamics of phage transcription across a single and multiple life cycles.



226

227

228 **Figure 1. Characteristics of MS2 phage A.** MS2 replication pattern. The MS2 replicase
229 complex contains host factors EF-Tu, EF-Ts, and S1 ribosomal protein. **B.** Schematic
230 presentation of the MS2 full-length genomic RNA coding regions organisation and
231 composition. *A*, maturation protein-encoding gene; *COAT*, coat protein-encoding gene; *L*, lysis
232 protein-encoding gene; *REPLICASE*, replicase protein-encoding gene. NTD, N-terminal

233 domain; CTD, C-terminal domain. **C.** Growth curves of bacterial cells (DSM5695 *E. coli*)
234 infected with MS2 RNA phage at MOI = 200 (red line) and bacterial cells without infection
235 (blue). **D.** MS2 plaques on the host *E. coli* DSM5695. Plaques were pointed with white arrows.
236 **E.** Plaque PFU and size change. All pair-wise comparisons were performed after observing a
237 significant Analysis of Variance (ANOVA) with * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$. Data
238 are presented as means ± SD (n = 3).
239

240 **MS2 phage maintained a high transcriptional activity across time**

241 To capture MS2 transcriptional changes across time, we performed long-read Nanopore direct
242 RNA sequencing (DRS) runs on total RNA extracted from DSM5695 *E. coli* cells infected with
243 MS2. We mapped over 17,6 million reads over a PHRED quality score threshold of 10. Out of
244 7.5 million reads were mapped on the MS2 genome, we found an average read length spanning
245 from 668 to 954 nt (**Suppl. Table S1)**. Reads over 90% of the MS2 transcript were included as
246 full-length reads (**Suppl. Table S1**). Immediately after infection (0h p.i.), only full-length
247 RNAs were detected (which accounted for only 1.63% of the total viral reads) (**Figure 2A and**
248 **Suppl. Table S2**) and almost no observable negative-strand transcripts (**Suppl. Figure S2C)**.
249 A sudden increase in the ratio of viral RNAs to bacterial RNAs occurred from 20 minutes
250 (1.05%) to 40 minutes (56.13%), suggesting the first cycle burst occurred at ~30 to 40 minutes
251 post-inoculation (**Figure 2A and Suppl. Table S2)**. Gene expression analysis was conducted
252 to assess variation between different time points. From the Principal Component Analysis
253 (PCA), the observed variation on the first two PC axes accounted for 88% of the total variance
254 (**Figure 2B**). Three main groups were identified: the 0h and 20 minutes p.i. on the PC1 negative
255 axis and the infected group over 40 minutes p.i. on the positive PC1 axis (**Figure 2B**). Among
256 the infected group, the 40-minute time point group was on the PC2 positive axis, whereas the
257 3 and 6 hours p.i. were clustered on the negative PC2 axis **(Figure 2B)**. The different time

258 points exhibited distinct clusters on the correlation heatmap (**Suppl. Figure S2A**), confirming
259 significant transcriptional changes of MS2 across time. Next, we analyzed the differential
260 expression (DE) of the four MS2 ORFs across time (**Figure 2C**). We found an upregulation of
261 the *mat* and *cap* ORFs while the *rep* ORF was downregulated (**Figure 2C**). Interestingly, we
262 found a peak expression of the *lys* gene at 40 minutes p.i., capturing a potential phage burst.
263 We confirmed by RT-qPCR similar results to the Nanopore DRS (**Suppl. Table S3**). Compared
264 with 0h p.i., we importantly found a significant fold increase of > 400-fold at 3 and 6 hrs p.i.
265 of the coat protein (**Suppl. Table S3**). The results also indicated the relatively higher expression
266 of phage RNA of the coat-protein ORF than the maturation-protein ORF at the same time point
267 (**Suppl. Table S3**), confirming our observations from Nanopore DRS. Overall, it suggests a
268 significant up regulation of the coat protein and fine tuning of the ORFs expression to regulate
269 MS2 transcription across time.
270
271 **Read mapping revealed a bias in coverage towards the coat protein ORF**
272 During replication, MS2 forms an intermediate negative transcript that serves as a template for
273 transcription/translation of the ORFs (Wagner, Weise, and Mutschler, 2022). We separated the
274 positive and negative strands, mapped reads to capture the intermediate state of viral
275 replication. Next, we assessed the read coverage across time for all reads (**Figure 2D-E**),
276 positive and negative strands (**Figure 3A-B**). We observed a coverage bias towards the *coat*
277 ORF at 3h and 6h p.i. and the 3' end of the viral genome on the overall transcript (**Figure 2D),**
278 further confirming our results. As the nanopore DRS has the 3' bias due to directional
279 sequencing from the 3' ends of RNAs (Jain et al., 2022), we deduced that coverage was biased
280 only towards the *coat* ORF in the MS2 transcriptome.
281
282 **Emergence and dominance of short transcripts across time**

While quantifying read counts, we noticed over 60-fold increase of the full-length transcripts (> 3000 nt) across time (**Figure 2 F-J**), consistent with the gene expression data. Intriguingly, we noticed the emergence and dominance of a large population of short reads (< 800 nt) from 40 minutes p.i. (**Figure 2 H-J and Suppl. Figure S2 F-K)** on positive strand**,** ranging from 37 % to 55 % of the read count (**Suppl. Table S4**). A possible explanation is that these reads are the accumulation of degraded RNA via bacterial lysis. As such, the number of these short reads would increase in the phage and the bacteria across time. We quantified the proportion of these short reads in the host genome, and we found a similar percentage variable from 82% to 92% of the total read count, with a decrease to 82 - 85% from 20 minutes to 3 hours p.i. (**Suppl Figure S4 and Suppl Table S5**). It suggests that the increase in the proportion of the short-read count found in phage is not correlated with bacterial lysis. To quantify a potential RNA degradation of these reads mapped to the phage transcript, we performed a quality control of the samples and the reads consisting in measuring the RNA integrity number (RIN) (**Suppl. Table S6**), the quality score versus read length and the number of mismatches per read to the reference transcript, proxy for read quality (**Suppl Figure S3**). The RIN score shows a low amount of RNA degradation with a median of 7.8 ±0.6 (**Suppl. Table S6)**. To miminise the effect of RNA degradation in our analysis, we opted for a more stringent cutoff from the literature and retained reads over a PHRED score of 10. While N50 statistics show a slight reduction in the N50 length across time, from a median length dropped from 1,432 nt at 0 hrs p.i. to 1,185 nt at 6 hrs p.i.

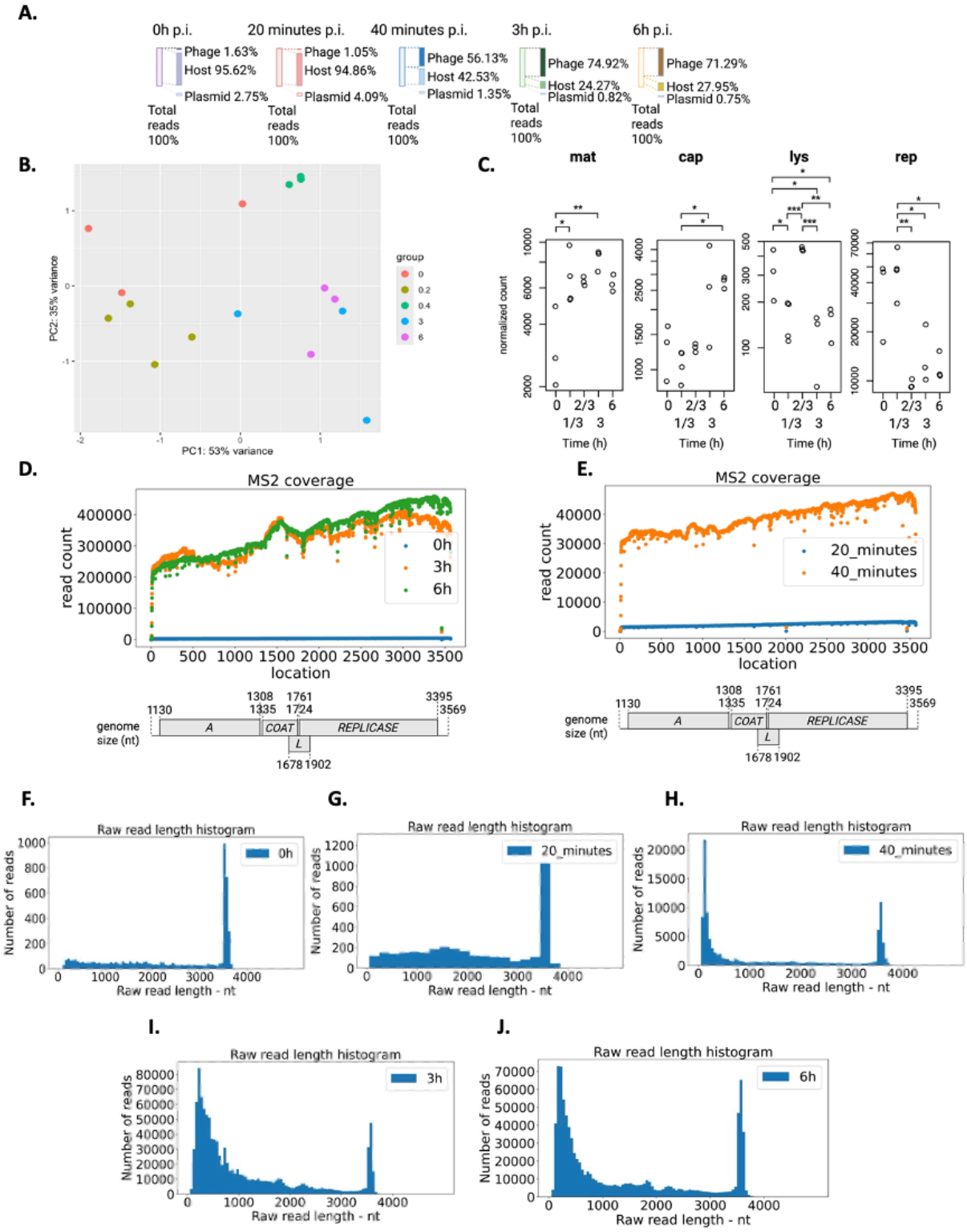


Figure 2. Statistics of Oxford Nanopore Technologies (ONT) direct RNA sequencing data via computational approaches and transcription pattern of MS2. A. Read counts from

312 Nanopore direct RNA sequencing of total RNA from DSM5695 *E. coli* cells infected with MS2
313 phage at different time points. **B**. Principal component analysis (PCA) plot of each MS2
314 sample. **C**. Differential expression patterns for four MS2 genes at different time points based
315 on normalised read count. * p <0.05, ** p < 0.01, *** p < 0.001. **D**. Coverage map for MS2
316 total reads at 0 to 6 hours p.i., **E**. 20 and 40 minutes p.i., **F**. MS2 raw read length distribution
317 at 0h p.i. **G**. At 20 minutes p.i. **H**. At 40 minutes p.i. **I**. At 3h p.i. J. At 6h p.i.
318
319 **AG-rich regions were responsible for the initiation of small transcripts**
320 Previous works described the ability of the replicase to generate small spontaneous template-
321 directed RNA species (Chetverin, Chetverina, and Munishkin, 1991; Wagner, Weise, and
322 Mutschler, 2022). We hypothesized that these small transcripts could be either RNA species
323 resulting from Nanopore DRS 3' bias and library preparation, small transcripts required for
324 replication, and/or products of RNA degradation from bacterial lysis.
325
326 To further investigate the nature of these small transcripts, we quantified their coverage across
327 the MS2 genome (**Figure 3 A-B**). We found an enrichment over 2-fold of the 5' end and the
328 *coat* ORF (~18.8% of total positive-strand reads with read length <800 nt at 3h) on the positive
329 strands (**Figure 3A and Suppl. Table S7),** confirming the initial mapping and RT-qPCR results
330 (**Suppl. Table S3**). We also found enrichment of the 3' end of the full transcript, likely due to
331 the DRS bias. Interestingly, transcripts mapped to the 3' end of the *rep* ORFs were enriched on
332 the negative strand (**Figure 3B**), whereas part and the 3' end of the *mat* ORF lacked coverage
333 on the positive strand (**Figure 3A**). To detail the mapping of these < 800 nt transcripts, we
334 performed a genome-wide mapping of the start and end of these transcripts within a 5 bp size
335 window and calculated the fold changes relative to the overall coverage. We noticed peaks of
336 significance on the positive and negative strands (**Figure 3 C-D**). We notably observed hotspots

337 at ~1,000-1,300 bp within the *mat* ORF and hotspots of end sites at ~1740-1745 nt after the
338 c*oat* ORF and at 2125-2130 nt within *rep* ORF (**Figure 3 C-D**).
339
340 We next postulated that specific RNA signals (ribosome binding sites/codon start) are the key
341 determinants for MS2 transcript initiation of these < 800 nt transcripts. We investigated several
342 start site peaks by assessing the genomic composition and secondary structure of their
343 corresponding genomic locations. We found AG-rich regions near these hotspots on the
344 negative strand. Using Clustal alignment and MEME, we found that the MEME motif (**Suppl.**
345 **Figure S5 A-E**) of these regions superimposed to known ribosomal binding sites (RBS) (Ng,
346 Wenfa (2019) (**Suppl. Table S8**). Together, they suggest the production of small transcripts
347 during the phage life cycles.
348
349 **A population of subgenomic reads is produced during the MS2 life cycles**
350 To replicate its own transcript, MS2 generates a negative-strand transcript, which initiates the
351 replication from the 3' to the 5' end (Wagner, Weise, and Mutschler, 2022). Prior work on
352 Nanopore sequencing technology has demonstrated that degraded RNAs/random cleavages
353 during library preparation exhibit a trimmed 3' end (Prawer et al., 2023). We postulated that
354 transcripts mapped starting from the 3' end resulted from Nanopore DRS 3' end bias since the
355 sequence starts from the 3' end, whereas transcripts mapped from the 5' end were degraded
356 transcripts. We aligned these short reads to the MS2 transcript. We found three types of aligned
357 reads (**Figure 3E-F**). The first block aligned from 3' to 5' end accounted for 78.42% of the read
358 count (length ranging 500nt – 900 nt) at 20 minutes p.i. (**Suppl. Table S9**) and reduced in
359 proportion after the first cycle. These reads corresponded to Nanopore 3' end bias. The second
360 block aligned from 5' to 3' end accounted for degraded RNA and increased over time, likely
361 due to active lysis in bacteria and potentially DRS sample preparation. Previous studies

362 identified that the coat protein subunit binds alongside the RNA template to mediate and
363 regulate the viral assembly and packaging during replication (Valegård et al., 1994).
364 Interestingly, a third block of aligned transcripts aligned to the *coat* and the *lys* ORFs and
365 increased across time (**Figure 3F and Suppl. Figure S6B**). This observation was in accordance
366 with our previous results on the gene expression analysis backed by RT-qPCR (**Figure 2D**).
367 Within one life cycle, we observed an increased count of transcripts containing the coat protein
368 encoding region, from 1.92% for 0h, 2.73% for 20 mins, and 3.38% for 40 mins (**Suppl. Table**
369 **S9**). For transcripts containing both the coat and lysis protein encoding region, we observed an
370 increase of 0.14% at 20 minutes to 1.23% at 40 minutes (**Suppl. Table S9**). We found no clear
371 alignment patterns with MS2 on the negative strand, suggesting that, unlike the full-length
372 transcripts, these small transcripts did not produce a negative strand (**Suppl. Figure S6 D-F,**
373 **Suppl. Table S10**). Overall, it suggests that these small transcripts throughout the life cycle(s)
374 vizualised are a combination of degraded RNA, small transcripts required for phage replication,
375 and Nanopore DRS 3' bias across time.
376

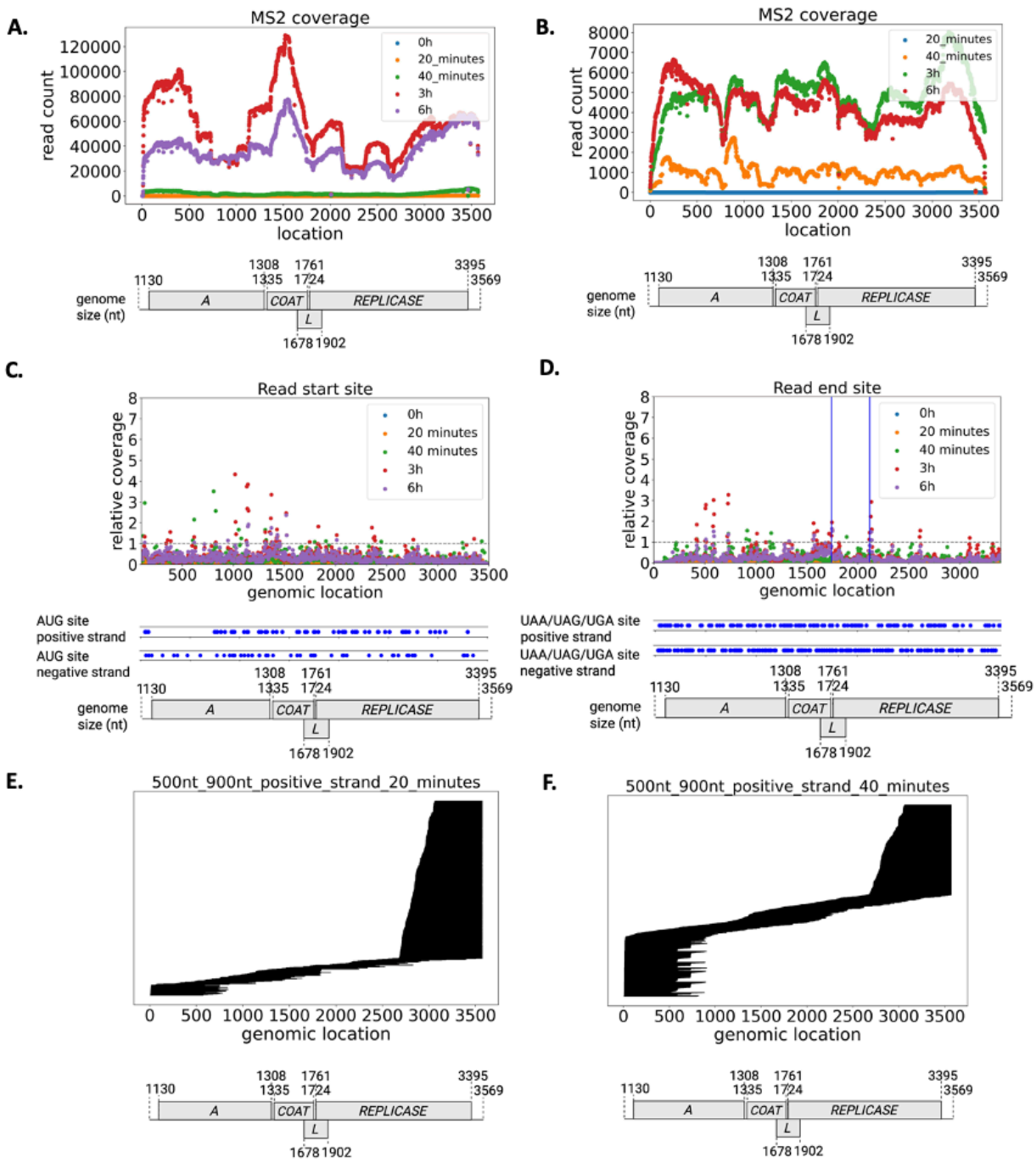


Figure 3. Replication pattern of MS2. A. Coverage map for MS2 reads (< 800 nt) aligned to the positive strand and **B.** the negative strand. **C.** Relative read start site (ranging 100-3500 nt) coverage profile and **D.** Relative read end site (ranging 0-3400 nt) coverage profile for MS2. The blue vertical line indicated two previously identified key coat protein binding sites. **E.**

384 Positive-strand reads (shown as lines) aligned to the MS2 genome with read length between
385 500 nt and 900 nt at 20 minutes p.i., and **F.** Read length between 500 nt and 900 nt at 40
386 minutes p.i.
387

388 **Error-prone replicase frequently generates hybrid RNAs**

389 Previous works have described the generation of by-products from the error-prone replicase
390 (Kopsidas et al., 2007). While it was previously reported that MS2 phage was capable of
391 forming hybrid RNAs, it was not found in an *in vivo* setting (van Meerten et al., 2002).
392 Nanopore DRS enables the characterization of recombination events and template switching in
393 viral RNA (Johnson et al., 2024). We interrogated the landscape of hybrid RNA transcripts
394 from DRS. We quantified the presence of hybrid reads by blasting these reads to MS2 genomes
395 and retaining the reads mapping to at least two different MS2 locations in single reads with
396 evidence of a recombination event (**Figure 4A**). Reads of recombination (20-60 nt deletions
397 and/or duplications) within the MS2 RNAs were detected (**Suppl. Table S11**). We found a
398 0.2% of duplicated or deletion hybrid reads to total reads ratio on the positive strand at 6h p.i.
399 We found a few RNA joining events of the negative strand to negative strand and sporadic
400 negative strand to positive strand, and very rare (+-) foldback RNAs (also known as copyback
401 or snapback RNAs - **Suppl. Figure S7 A-B**). We noticed internal tandem duplication events
402 and deletion events between the negative strand and the negative strand (**Suppl. Figure S7 A)**.
403 While we discerned recombination events within phage RNA, we did not observe any
404 recombination events between viral genome segments and host RNAs. We wondered whether
405 new variants have emerged over time. We assessed the mutational landscape of MS2 phage
406 (SNV and structural variation) across time. We have not observed the emergence of any new
407 mutations up to 6 hours p.i. Together, it indicates the error-prone replicase generates hybrid
408 transcripts and foldback RNAs as by-products.

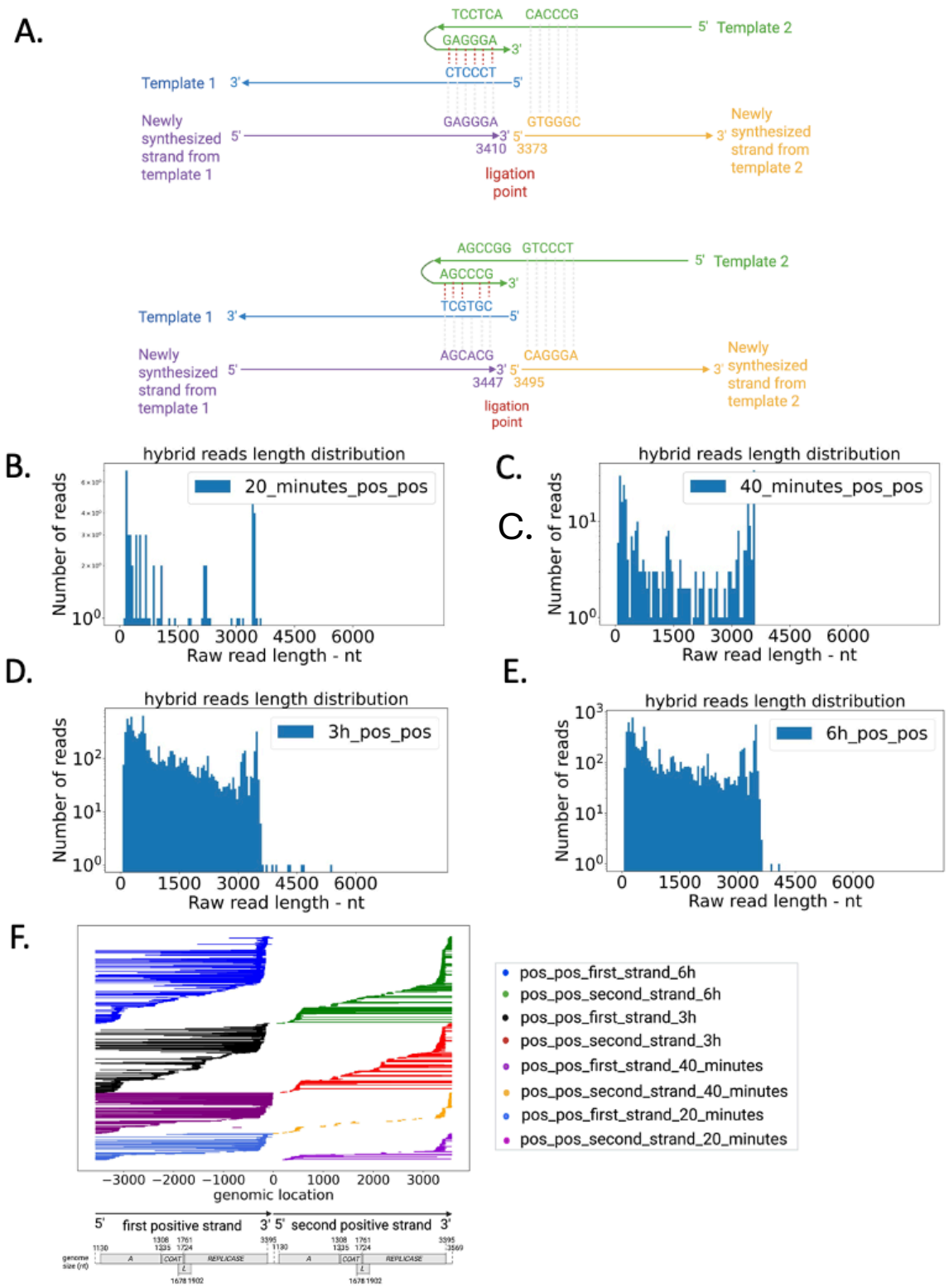


Fig 4. Characteristics summary of MS2 hybrid reads. A. Examples of template switching events occurring at multiple time points, ligation at location 3410 and 3373, and ligation at

413 location 3447 and 3495. Template 1 and Template 2 have a complementary region (red dashed
414 line). Replicase first uses template 1 as the template, then switches to template 2 as the template
415 for the following RNA synthesis. **B.** Read length of hybrid reads (positive to positive strand
416 ligation - pos_pos) histogram at 20 mins, **C.** 40 mins, **D.** 3h, and **E.** 6h post inoculation. **F.**
417 Features of hybrid reads (pos_pos) of MS2. Ligation events occurred at the 3' end of the first
418 strand and the 5' end of the second strand. All pos_pos hybrid (20 minutes) reads were
419 visualized. 30% of total pos_pos hybrid (40 minutes) reads were visualized. 2% of total
420 pos_pos hybrid (3h and 6h) reads were visualized.
421

422 **Static and dynamic RNA modification on MS2 phage during replication**

423 RNA modification plays an important role in modulating RNA function (Pozhydaieva et al.,
424 2024). Previous works have identified the presence of ψ modification on MS2 phages using
425 either mass spectrometry and/or by DRS (Jones, 2023). We wondered whether MS2 acquired
426 new modifications during a single life cycle (from 0h to 40 minutes p.i.) and across time (3 and
427 6 hours p.i.). We predicted and mapped m5C and ψ modifications throughout a single MS2 life
428 cycle on the RNA positive strands (**Figure 5 A-D**). We found m^5C, and ψ modifications were
429 widely distributed (**Figure 5 A-D**). On full-length transcripts (> 3,000 nt), we found no
430 evidence of newly acquired modifications within a single cycle (**Figure 5 A and C**). We
431 identified instead commonly modified static sites with high stoichiometry throughout the
432 transcripts, varying from 23.4% to 78.0% (**Figure 5 A and C**). Interestingly, we found shorter
433 transcripts increase in m5C stoichiometry across time (**Figure 5B and Suppl. Figure S9B**).
434 Consistent to a previous study (Jones, 2023), we also identified a conserved ψ modification
435 site 924 nt (**Figure 5C**) We found additional m5C (at location 634, 927, 2636, 3378, 3413, and
436 3521 nt) and ψ installation sites (at location 2145 and 2407 nt) for small transcripts with length
437 < 800 nt at 3h and 6h p.i. (**Suppl. Figure S9 B-D**). These additional modifications were located

on the *mat* and *rep* ORFs. We assessed m6A modification, and we observed no modification pattern (**Suppl. Figure S8 A-D**). Overall, it suggests that RNA modifications in modified sites vary in stoichiometry during a single life cycle. Additionally, we demonstrated that *de novo* and stable modifications were acquired across time.

We next assessed the secondary structures at these differentially methylated sites and the potential motifs that have served as substrates for known modification enzymes. We found no specific modification sites or secondary structures for m5C modification. We observed conserved loops or at the stem-loop junctions ψ installation motif on the MS2 RNA (**Suppl Figure S10 A-L**). MEME motif pattern showed the conserved sequence among these sites (U-rich hairpin) (**Suppl Figure S10 M-N**). On a conserved position at 924 nt (with the highest stoichiometry: >80% among all the modification sites), we found a conserved recognition structure (installation at stem-loop junctions) and a highly conserved sequence compatible with ΨUACA RluA canonical recognition motif (Schaening-Burgos et al., 2024). Other pseudouridine installation sites exhibited relatively low stoichiometry (<30%) compared to location 924 (>80%). Analysis of their secondary structures showed either a fold compatible with ΨUACA RluA canonical recognition motif or a non-canonical motif substrate (**Suppl Figure. 10**). Together, these results suggest that RNA modification was present in the initial transcripts and throughout the life cycle without the acquisition of new modified sites on full transcripts, whereas small transcripts do acquire and remove modified sites during replication. These newly modified sites, amenable to modifications, are U-rich hairpins for Ψ modification.

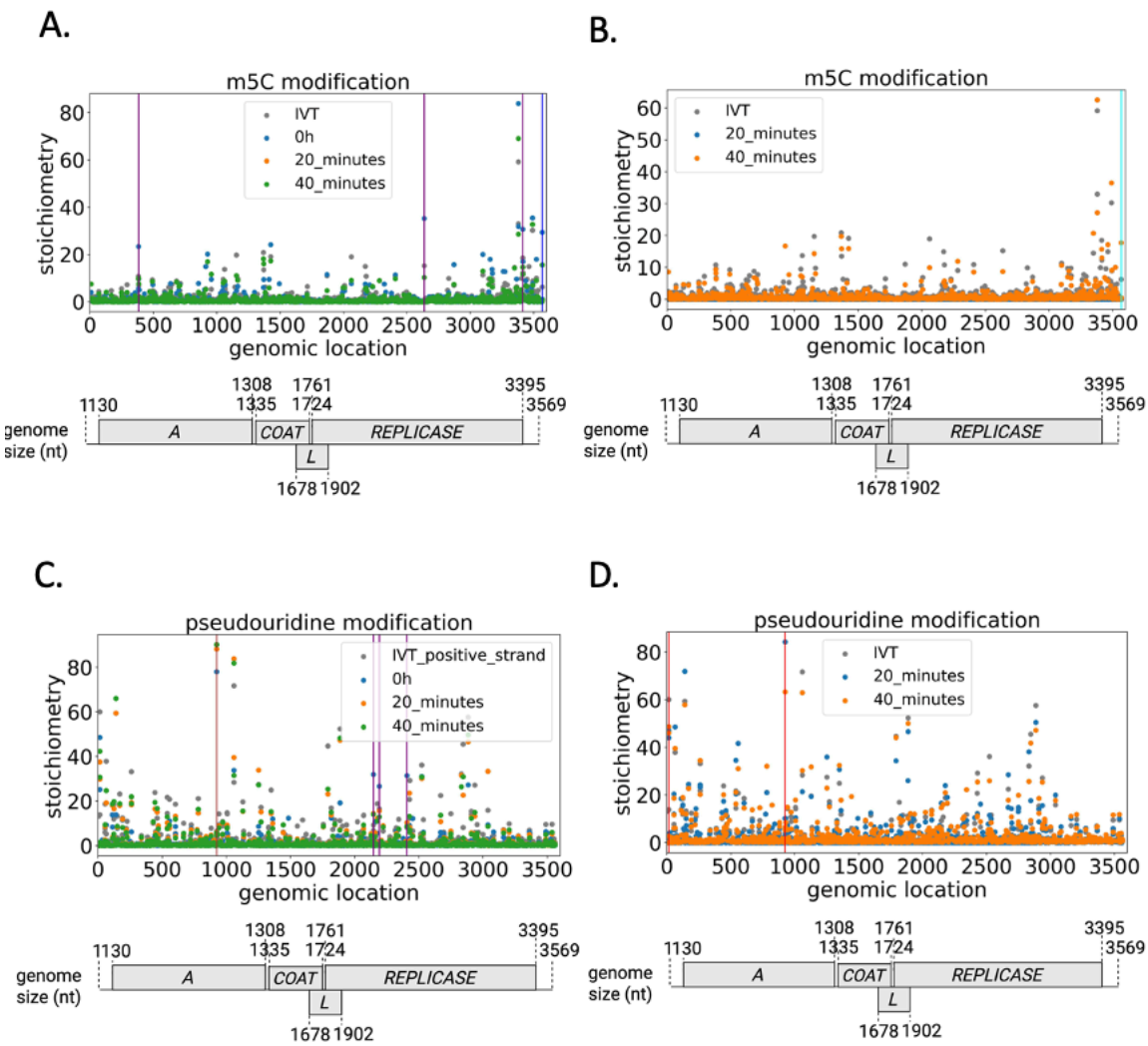


Figure 5. Predicted sites of ψ and m^{5}C RNA modification dynamics on MS2 phage positive RNA (>3000 nt length and <800 nt length) within one life cycle (at 0h, 20 minutes, and 40 minutes p.i.). Synthetic modification-free RNA as the negative control. A. m5C modification (>3000 nt) landscapes according to the counts of modified and unmodified bases. **B.** m5C modification (<800 nt) landscapes. **C.** ψ modification (>3000 nt) landscapes. **D.** ψ modification (<800 nt) landscapes. With selection criteria based on single-base analysis (those that meet all the requirements are highlighted in vertical lines): For the positive strand with read length > 3000 nt, 0h vs. 20 minutes decrease, 20 minutes vs. 40 minutes increase (vertical blue line); 0h vs. 20 minutes decrease, 20 minutes vs. 40 minutes no difference (vertical purple

line); 0h vs. 20 minutes increase, 20 minutes vs. 40 minutes no difference (vertical brown line); For the positive strand with read length <800 nt. 20 minutes vs. 40 minutes increase (vertical cyan line), 20 minutes vs. 40 minutes no difference (vertical red line). Stoichiometry (test samples) - stoichiometry (IVT samples) > 0, balanced MAP-based p-value < 0.05. All p-values pass the significance q value for reducing the False Discovery Rate (FDR). 5-methylcytidine, m^5C.

Discussion:

RNA bacteriophages such as Qβ and MS2 have been extensively studied for their ability to rapidly replicate (Chang et al., 2022; Vasilyev et al., 2013; Wagner, Weise, and Mutschler, 2022). However, the dynamics of their transcriptional activities across a single and multiple life cycles remain relatively unexplored. This study aimed to detail the dynamics of MS2 transcriptional activity using the Nanopore direct RNA sequencing approach across 6 hours post-inoculation. We found that MS2 phage maintained a high transcriptional activity, especially expressing high amounts of small transcripts containing the coat protein-encoding region. We also found hybrid transcripts and foldback RNA that have arisen from template switching recombination. By investigating the RNA modification landscapes, we found a conserved U-rich hairpin compatible with RluA-mediated Ψ modification. Together, direct RNA Nanopore sequencing has allowed the dissection of the detailed dynamics of MS2 transcriptional activity across time. This has led to our main finding that, despite carrying a minimal genome, MS2 efficiently enhanced its transcriptional activity and harnessed its error-prone replicase to rapidly attach, replicate, and lyse the bacterial cell.

High transcriptional/replication level of MS2 within bacterial cells

498 Previous works assessed in depth RNA phages (Vasilyev et al., 2013) and RNA viruses (Hillen
499 et al., 2020) replication patterns (including host factors). The archetypical MS2 and Qβ
500 replicase complexes have been extensively investigated since the 1960s, including the host
501 factors responsible for forming the complex (Wagner et al., 2022) and their underlying roles
502 (Vasilyev et al., 2013). However, little is known about the dynamics of the transcriptional
503 activity during the MS2 phage life cycles. A key reason is that most previous studies were
504 performed in an *in vitro* or a transcription-translation system for MS2 (Wagner et al., 2022) and
505 Qβ (Vasilyev et al., 2013; Tomita, Ichihashi, and Yomo, 2015). Using Nanopore RNA DRS,
506 we reported significant changes in MS2 transcript levels across multiple life cycles. Strikingly,
507 we noticed an over 50-fold increase in full-length MS2 transcripts in only three hours post-
508 inoculation and over 400-fold increase in coat protein expression, indicative of high activity of
509 the phage for replication and assembly to maintain the infection. To a lesser extent, other RNA
510 viruses, such as the Human Coronavirus 229E (HCoV-229E), produce their own transcripts
511 (~33% reads mapped to the virus genome) at high levels (Viehweger et al., 2019) for viral
512 replication. Another study also showed a large proportion of reads (~65% reads) mapped to the
513 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Kim et al., 2020). As such,
514 our observation is in line with previous reports on other viruses, suggesting the ability of a
515 virus to rapidly transcribe within hours post-inoculation. Together, our study reports that RNA
516 phages such as MS2 require significant upregulation and fine-tuning of the expression of their
517 ORFs across time.
518
519 **Formation of subgenomic transcripts during the MS2 life cycle.**
520 Previous works have reported that MS2 utilizes its negative strand for replication via the
521 replicase complex (Dykeman, 2023; Wagner, Weise & Mutschler, 2022). While the Qβ
522 replicase complex has been extensively studied for the formation of spontaneous short

523 amplifying RNA and additional by-products, MS2 replicase has been recently described as less
524 promiscuous towards the formation of spontaneous replicative products (Wagner, Weise &
525 Mutschler, 2022).
526
527 During our investigation, we identified a large population of short subgenomic reads appearing
528 after the latent period of phage infection and increasing over time. We hypothesised that this
529 population could be a Nanopore sequencing bias due to degraded RNA from bacterial lysis
530 after MS2 infection, and 3' end enrichment is the nanopore DRS 3' end bias (Jain et al., 2022).
531 We also hypothesised that these short transcripts could be single ORF transcripts. We found
532 indeed evidence of RNA degradation and Nanopore DRS sequencing bias. Interestingly, we
533 identified hotspots of mapped transcripts corresponding to the full-length coat protein ORF.
534 Specifically, the top RNA virion contact sites were previously identified at positions
535 surrounding 1740-1750 nt and 2110-2120 nt (Rolfsson et al., 2016) and overlayed with the top
536 termination hits **(Figure 3D),** suggesting the assembly and packaging process could potentially
537 inhibit the replication process by prematurely ending the transcription at specific RNA contact
538 sites (**Figure 3E-F**). Additionally, we reported that this transcription is potentially regulated by
539 the occupancy of the coat protein on the positive strand on specific RNA binding sites,
540 potentially aborting the transcription/translation from the replicase complex (Rolfsson et al.,
541 2016). Together, we suggest that MS2 phage produces a high level of coat protein transcripts
542 to potentially repress the replicase complex, and critical for virion assembly.
543
544 **Error-prone replicase complex generates hybrid transcripts**
545 An important requirement for some viruses to replicate and evolve is to harbour either
546 proofreading (Robson et al., 2020) or non-proofreading replicase (Whitfield et al., 2020). A
547 mechanism for viruses to acquire novel traits and evolve is to form hybrids with other viruses

548 or within the same virus (Simon-Loriere and Holmes, 2011). We wondered whether the error-
549 prone replicase would favour the generation of *de novo* mutations and such hybrid RNAs to
550 form new virions. Using Nanopore DRS, we reported long intraspecies hybrid transcripts
551 (>3569 nt), including fold-back RNAs accounted for 0.2% of the total transcript count. We also
552 found that these hybrid transcripts resulted from template switching recombination. The
553 presence of hybrid RNA or recombination events has been reported for other eukaryotic viruses
554 such as SARS-CoV-2 (Garushyants, Rogozin, and Koonin, 2021) or plant viruses (Johnson et
555 al., 2024). Hybrid RNAs have also been reported for MS2 in plasmid systems (van Meerten et
556 al., 2002). Consistent with previous observations, our findings indicate the ability of MS2 to
557 recombine using template switching mechanisms into new RNA transcripts.
558
559 **U-rich hairpin of Ψ installation sites on native MS2 RNA**
560 Previous works have evidenced the crucial role of RNA modification in viral replication
561 (Bhattacharya et al., 2022). To achieve high translation efficiency and evade host immunity,
562 some viruses encode their own methyltransferase or other modifying enzymes (Li and Rana,
563 2022). However, the MS2 genome doesn't harbor any modifying enzymes and requires
564 hijacking the host to modify its own RNA. Using Nanopore DRS, we aimed to assess whether
565 the modification landscape within a single life cycle and across time. We evidenced the
566 presence of the Ψ installation site and secondary structure (U-rich hairpin) amenable to Ψ
567 modification with a variation of stoichiometry within a single cycle and across time. We notably
568 found evidence of canonical and non-canonical ΨUACA RluA-mediated installation,
569 providing evidence that host pseudouridine synthase modifies MS2 RNA during its life cycle.
570 Interestingly, we found a dynamic of loss and gain in methylation of the small transcripts during
571 the MS2 across time. It suggests that, unlike the full transcripts, small transcripts are more
572 amenable to RNA modification, indicating the distinct secondary structure of small transcripts

compared to full-length transcripts, rendering them more amenable to Ψ installation. Together, it confirms previous work that the MS2 transcript is amenable to RNA modification (Jones, 2023). These modifications are either transient, predominantly on short transcripts, or permanent with various stoichiometries on canonical ΨUACA RluA-mediated installation sites. However, the functionality of these modifications on MS2 ORF translation remains unknown and will require further investigations.

Conclusion

Our investigation, capturing the dynamics of MS2 transcriptional activity across time using the Nanopore direct RNA sequencing approach, has revealed a complex interplay between the replicase, the coat protein, and the negative RNA strand to control and fine-tune its transcription. It has revealed the numerous generations of byproducts during transcription, as either products of the error-prone replicase (hybrid reads) or coat protein transcripts required for assembly. We finally reported the dynamics of modified RNA stoichiometry changes during a single cycle and the addition of new modification sites with low stoichiometry of the canonical and non-canonical substrate motif RulA-mediated. Together, the direct RNA sequencing approach has revealed new insights into the complex transcriptional dynamics of MS2 bacteriophage.

Acknowledgements.

The authors would like to acknowledge the Biomolecular Resource Facility (Dr Carolina Correa-Ospina and Mr Ziyan Zhang) for assistance in the Nanopore sequencing. The authors would like to acknowledge Mr Agin Ravindran for helping with the magnetic beads purification. The authors would like to acknowledge the Burgio laboratory for comments on the early version of the manuscript. This work was undertaken with the assistance of resources

from the National Computational Infrastructure (ANUMAS and NCMAS schemes), an NCRIS enabled capability supported by the Australian Government. GB work is supported from the National Health and Medical Research Council, the Medical Research Future Funds and the Australian Research Council and the Medical Research Future Fund.

Competing interests

The authors declare no competing interests

Authors contribution

GB conceived the study. GB supervised the work and provided funding. NN performed the experiments. NN and GB performed the analysis. NN and GB wrote the paper.

References:

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), pp.289-300.

Bhattacharya, T., Yan, L., Crawford, J.M., Zaher, H., Newton, I.L. and Hardy, R.W., 2022. Differential viral RNA methylation contributes to pathogen blocking in Wolbachia-colonized arthropods. *PLoS pathogens*, *18*(3), p.e1010393.

Bollback, J.P. and Huelsenbeck, J.P., 2001. Phylogeny, genome evolution, and host specificity of single-stranded RNA bacteriophage (family Leviviridae). Journal of molecular evolution, 52, pp.117-128.

Chamakura, K.R., Tran, J.S. and Young, R., 2017. MS2 lysis of Escherichia coli depends on host chaperone DnaJ. *Journal of Bacteriology*, *199*(12), pp.10-1128.

Chang, J.Y., Gorzelnik, K.V., Thongchol, J. and Zhang, J., 2022. Structural assembly of Qβ virion and its diverse forms of virus-like particles. *Viruses*, *14*(2), p.225.

Chetverin, A.B., Chetverina, H.V. and Munishkin, A.V., 1991. On the nature of spontaneous RNA synthesis by Qβ replicase. *Journal of molecular biology*, *222*(1), pp.3-9.

Clokie, M.R., Millard, A.D., Letarov, A.V. and Heaphy, S., 2011. Phages in nature. *Bacteriophage*, *1*(1), pp.31-45.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M. and Li, H., 2021. Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2), p.giab008.

Dykeman, E.C., 2023. Modelling ribosome kinetics and translational control on dynamic mRNA. *PLoS Computational Biology*, *19*(1), p.e1010870.

Dykeman, E.C., 2024. Design of a self-regulating mRNA gene circuit. *Scientific Reports*, *14*(1), p.19421.

Garushyants, S.K., Rogozin, I.B. and Koonin, E.V., 2021. Template switching and duplications in SARS-CoV-2 genomes give rise to insertion variants that merit monitoring. *Communications biology*, *4*(1), p.1343.

Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D. and Cramer, P., 2020. Structure of replicating SARS-CoV-2 polymerase. *Nature*, *584*(7819), pp.154-156.

Jain, M., Abu-Shumays, R., Olsen, H.E. and Akeson, M., 2022. Advances in nanopore direct RNA sequencing. *Nature methods*, *19*(10), pp.1160-1164.

Jain, M., Fiddes, I.T., Miga, K.H., Olsen, H.E., Paten, B. and Akeson, M., 2015. Improved data analysis for the MinION nanopore sequencer. *Nature methods*, *12*(4), pp.351-356.

Johnson, P.Z., Needham, J.M., Lim, N.K. and Simon, A.E., 2024. Direct nanopore RNA sequencing of umbra-like virus-infected plants reveals long non-coding RNAs, specific cleavage sites, D-RNAs, foldback RNAs, and temporal-and tissue-specific profiles. *NAR Genomics and Bioinformatics*, *6*(3), p.lqae104.

Jones, J., 2023. Unveiling the RNA Modification Landscape Through Bioanalytical and Biochemical Analyses (Doctoral dissertation).

Kim, D., Lee, J.Y., Yang, J.S., Kim, J.W., Kim, V.N. and Chang, H., 2020. The architecture of SARS-CoV-2 transcriptome. *Cell*, *181*(4), pp.914-921.

Kopsidas, G., Carman, R.K., Stutt, E.L., Raicevic, A., Roberts, A.S., Siomos, M.A.V., Dobric, N., Pontes-Braz, L. and Coia, G., 2007. RNA mutagenesis yields highly diverse mRNA libraries for in vitro protein evolution. *BMC biotechnology*, *7*, pp.1-12.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. *bioinformatics*, *25*(16), pp.2078-2079.

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), pp.3094-3100.

Li, N. and Rana, T.M., 2022. Regulation of antiviral innate immunity by chemical modification of viral RNA. *Wiley Interdisciplinary Reviews: RNA*, *13*(6), p.e1720.

Love, M.I., Huber, W. and Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*, pp.1-21.

Meng, R., Jiang, M., Cui, Z., Chang, J.Y., Yang, K., Jakana, J., Yu, X., Wang, Z., Hu, B. and Zhang, J., 2019. Structural basis for the adsorption of a single-stranded RNA bacteriophage. *Nature communications*, *10*(1), p.3130.

Ng, Wenfa (2019). Database of ribosome binding site of all genes in Escherichia coli K-12 MG1655. figshare. Dataset. https://doi.org/10.6084/m9.figshare.10282685.v1

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*, *14*(4), pp.417-419.

Peabody, D.S., 1990. Translational repression by bacteriophage MS2 coat protein expressed from a plasmid. A system for genetic analysis of a protein-RNA interaction. *Journal of Biological Chemistry*, *265*(10), pp.5684-5689.

Poot, R.A., Tsareva, N.V., Boni, I.V. and Van Duin, J., 1997. RNA folding kinetics regulates translation of phage MS2 maturation gene. *Proceedings of the National Academy of Sciences*, *94*(19), pp.10110-10115.

Pozhydaieva, N., Wolfram-Schauerte, M., Keuthen, H. and Höfer, K., 2024. The enigmatic epitranscriptome of bacteriophages: putative RNA modifications in viral infections. *Current Opinion in Microbiology*, *77*, p.102417.

Prawer, Y.D., Gleeson, J., De Paoli-Iseppi, R. and Clark, M.B., 2023. Pervasive effects of RNA degradation on Nanopore direct RNA sequencing. *NAR Genomics and Bioinformatics*, *5*(2), p.lqad060.

Quinlan, A.R. and Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), pp.841-842.

Robson, F., Khan, K.S., Le, T.K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P. and Ng, W.L., 2020. Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Molecular cell*, *79*(5), pp.710-727.

Rolfsson, Ó., Middleton, S., Manfield, I.W., White, S.J., Fan, B., Vaughan, R., Ranson, N.A., Dykeman, E., Twarock, R., Ford, J. and Kao, C.C., 2016. Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. Journal of molecular biology, 428(2), pp.431-448.

Salas, M. and de Vega, M., 2008. Replication of bacterial viruses. Encyclopedia of Virology, third ed. Academic Press, Oxford, pp. 399-544.

Schaening-Burgos, C., LeBlanc, H., Fagre, C., Li, G.W. and Gilbert, W.V., 2024. RluA is the major mRNA pseudouridine synthase in Escherichia coli. *PLoS Genetics*, *20*(9), p.e1011100.

Simon-Loriere, E. and Holmes, E.C., 2011. Why do RNA viruses recombine?. *Nature Reviews Microbiology*, *9*(8), pp.617-626.

Smolka, M., Paulin, L.F., Grochowski, C.M., Horner, D.W., Mahmoud, M., Behera, S., Kalef-Ezra, E., Gandhi, M., Hong, K., Pehlivan, D. and Scholz, S.W., 2024. Detection of mosaic and population-level structural variants with Sniffles2. *Nature biotechnology*, *42*(10), pp.1571-1580.

Tomita, K., Ichihashi, N. and Yomo, T., 2015. Replication of partial double-stranded RNAs by Qβ replicase. *Biochemical and biophysical research communications*, *467*(2), pp.293-296.

Twarock, R., Towers, G.J. and Stockley, P.G., 2024. Molecular frustration: a hypothesis for regulation of viral infections. *Trends in Microbiology*, *32*(1), pp.17-26.

Vasilyev, N.N., Kutlubaeva, Z.S., Ugarov, V.I., Chetverina, H.V. and Chetverin, A.B., 2013. Ribosomal protein S1 functions as a termination factor in RNA synthesis by Qβ phage replicase. *Nature communications*, *4*(1), p.1781.

Valegård, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. and Liljas, L., 1994. Crystal structure of an RNA bacteriophage coat protein–operator complex. *Nature*, *371*(6498), pp.623-626.

van Meerten, D., Groeneveld, H., Miller, D.M., Marechal, G.B., Tsareva, N.V., Olsthoorn, R.C., de la Peña, M. and van Duin, J., 2002. In vivo generation of hybrids between different species of RNA phages. *Journal of general virology*, *83*(5), pp.1223-1235.

Viehweger, A., Krautwurst, S., Lamkiewicz, K., Madhugiri, R., Ziebuhr, J., Hölzer, M. and Marz, M., 2019. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome research*, *29*(9), pp.1545-1554.

Wagner, A., Weise, L.I. and Mutschler, H., 2022. In vitro characterisation of the MS2 RNA polymerase complex reveals host factors that modulate emesviral replicase activity. Communications Biology, 5(1), p.264.

Whitfield, Z.J., Prasad, A.N., Ronk, A.J., Kuzmin, I.V., Ilinykh, P.A., Andino, R. and Bukreyev, A., 2020. Species-specific evolution of Ebola virus during replication in human and bat cells. *Cell reports*, *32*(7).

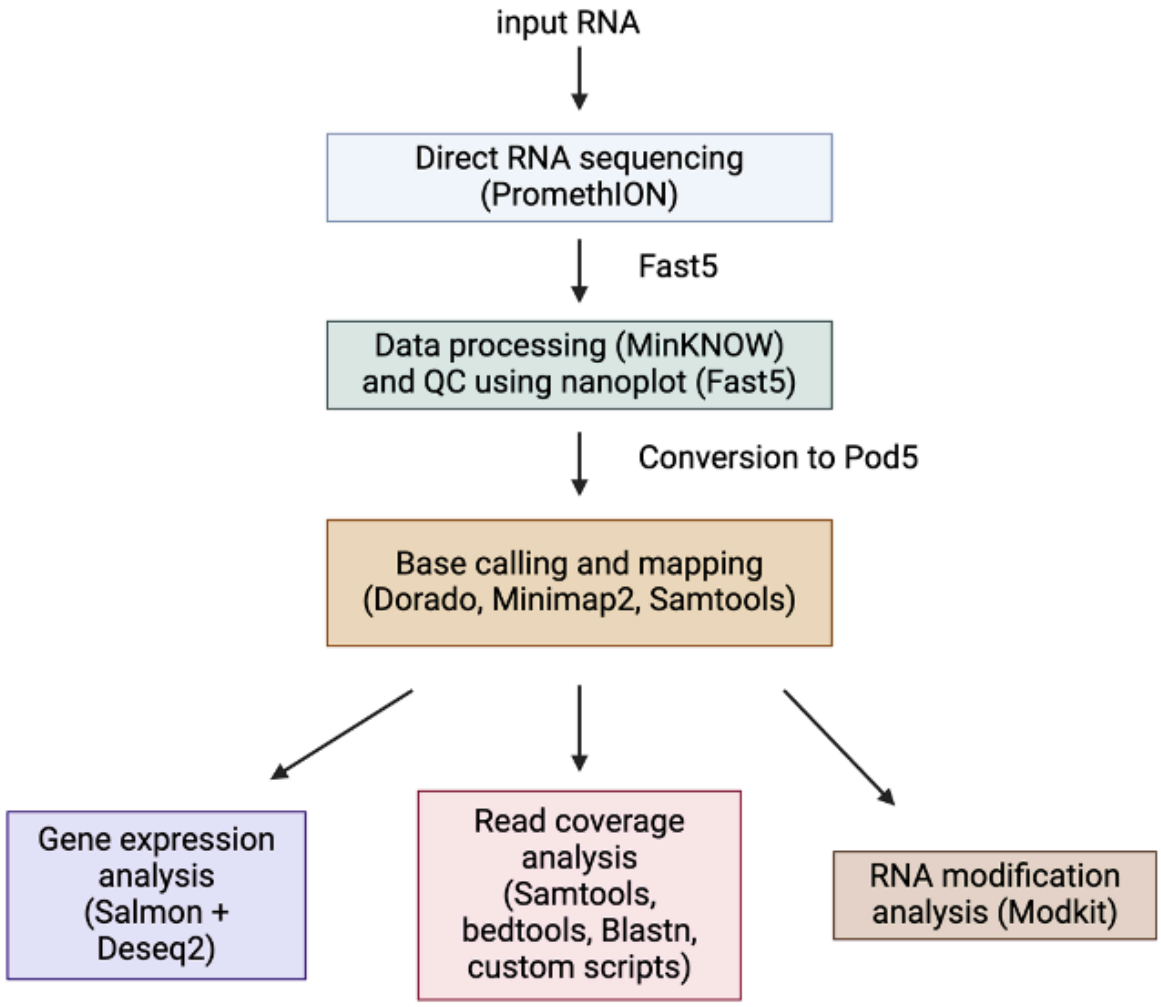
Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J.H., Krupovic, M., Dolja, V.V. and Koonin, E.V., 2018. Origins and evolution of the global RNA virome. *MBio*, *9*(6), pp.10-1128.

Workman, R.E., Tang, A.D., Tang, P.S., Jain, M., Tyson, J.R., Razaghi, R., Zuzarte, P.C., Gilpatrick, T., Payne, A., Quick, J. and Sadowski, N., 2019. Nanopore native RNA sequencing of a human poly (A) transcriptome. *Nature methods*, *16*(12), pp.1297-1305.

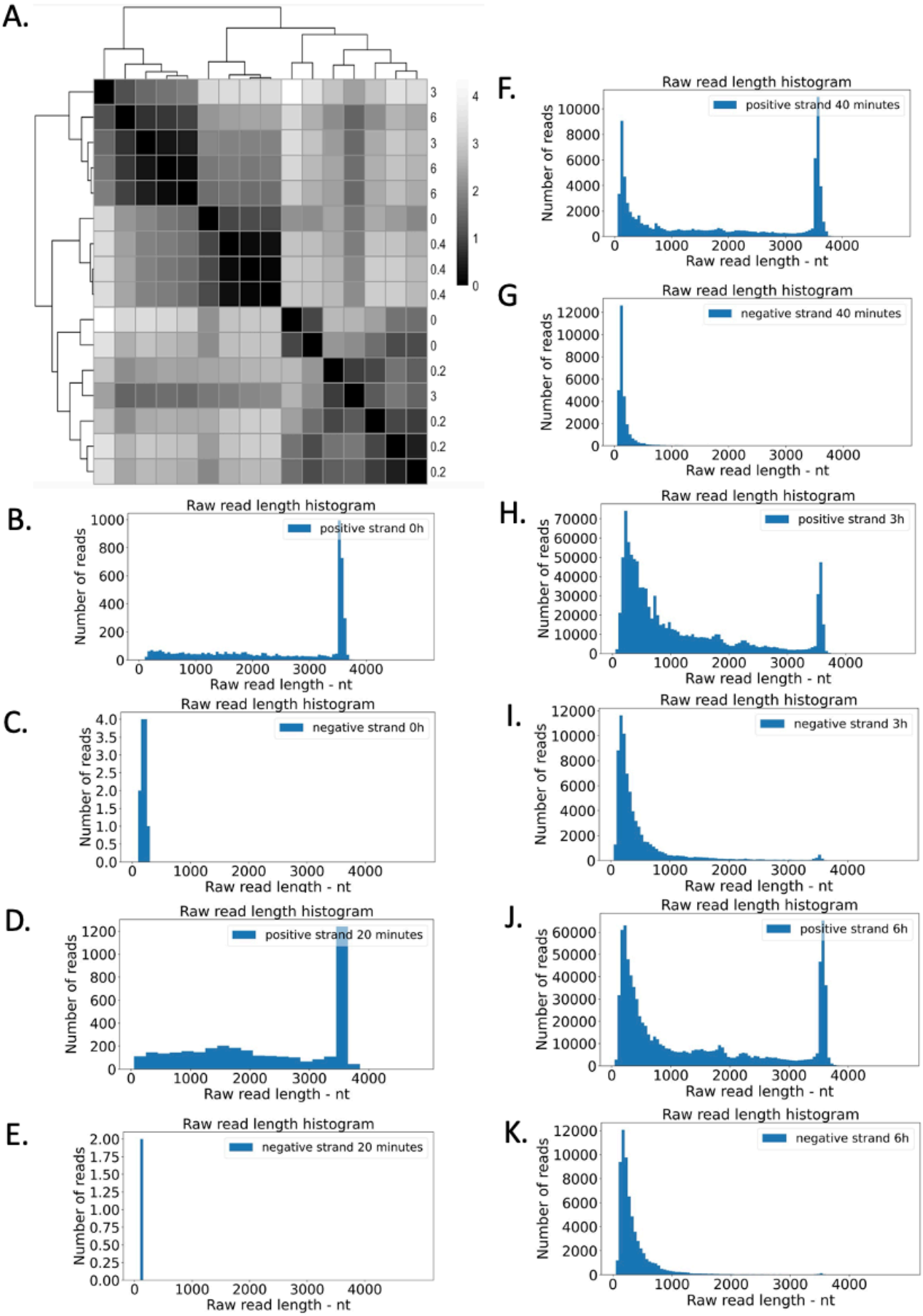
Zhang, W. and Wu, Q., 2020. Applications of phage-derived RNA-based technologies in synthetic biology. *Synthetic and Systems Biotechnology*, *5*(4), pp.343-360.

Zheng, Z., Li, S., Su, J., Leung, A.W.S., Lam, T.W. and Luo, R., 2022. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nature Computational Science*, *2*(12), pp.797-803.

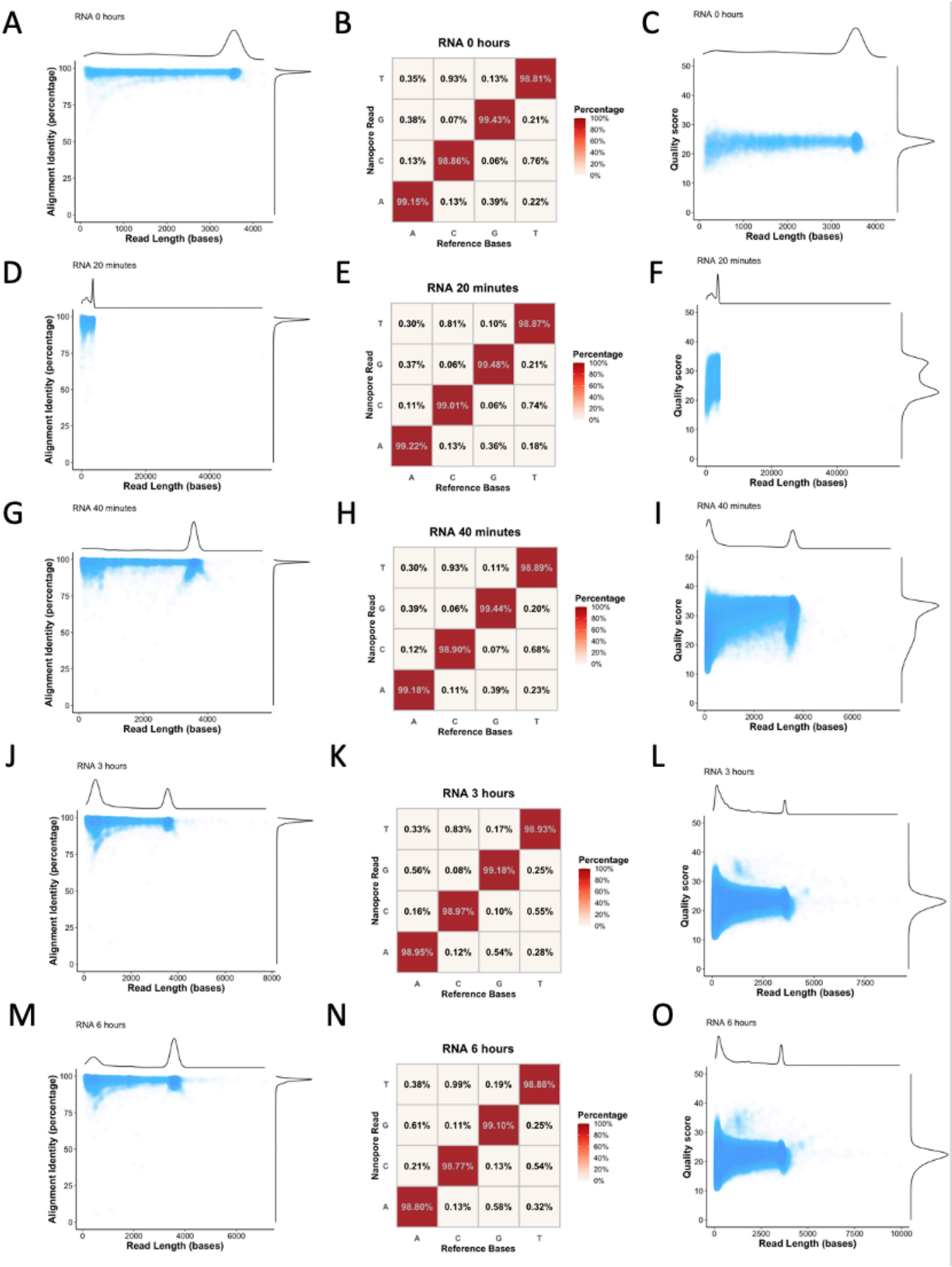
Supplementary figures:



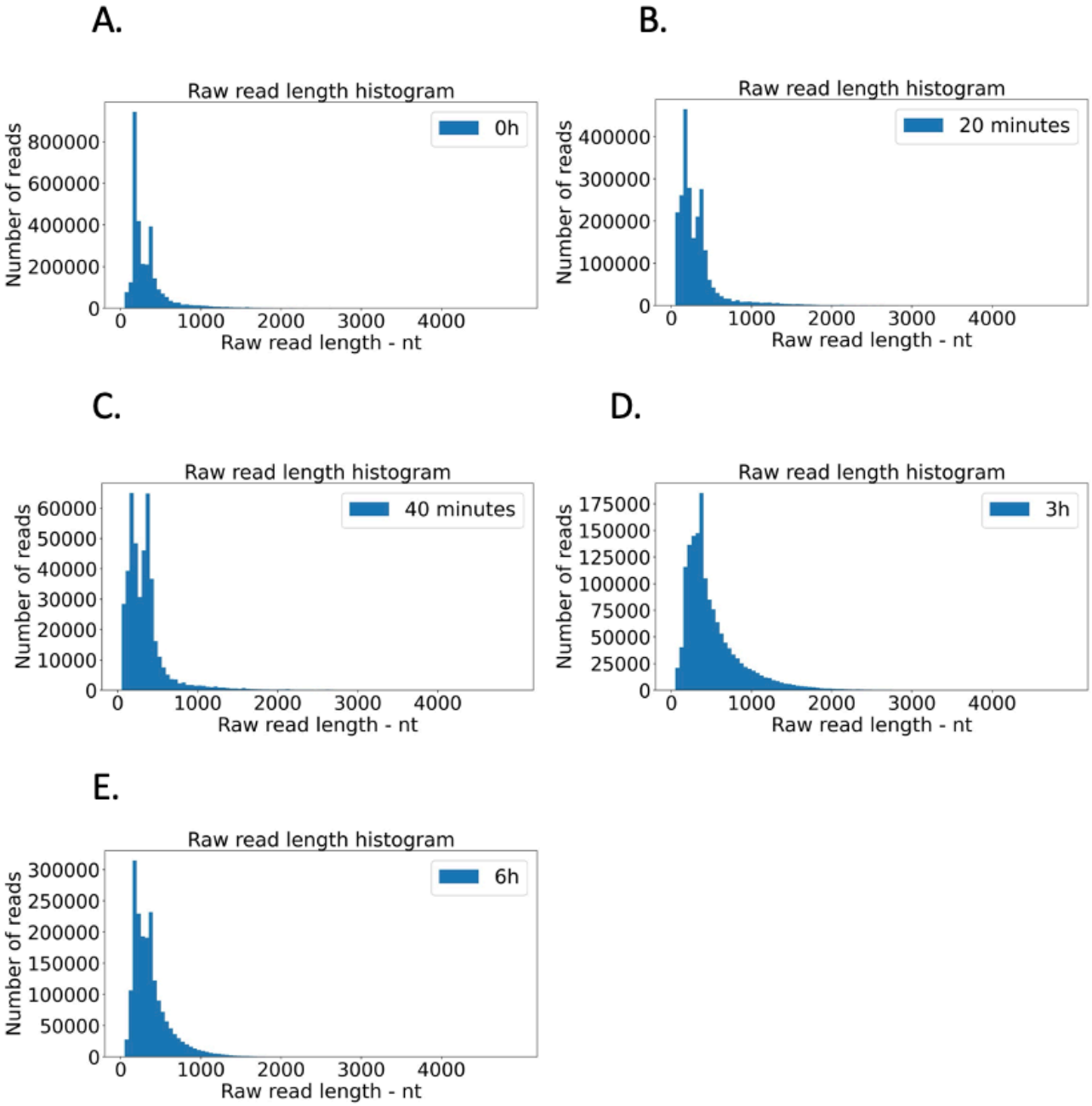
Supplementary Fig. S1. The schematic computational analysis workflow for processing the Oxford RNA-direct Nanopore sequencing datasets for MS2 RNAs.


A.
B.
Raw read length histogram
positive strand 0h
Number of reads
Raw read length - nt
C.
Raw read length histogram
negative strand 0h
Number of reads
Raw read length - nt
D.
Raw read length histogram
positive strand 20 minutes
Number of reads
Raw read length - nt
E.
Raw read length histogram
negative strand 20 minutes
Number of reads
Raw read length - nt
F.
Raw read length histogram
positive strand 40 minutes
Number of reads
Raw read length - nt
G
Raw read length histogram
negative strand 40 minutes
Number of reads
Raw read length - nt
H.
Raw read length histogram
positive strand 3h
Number of reads
Raw read length - nt
I.
Raw read length histogram
negative strand 3h
Number of reads
Raw read length - nt
J.
Raw read length histogram
positive strand 6h
Number of reads
Raw read length - nt
K.
Raw read length histogram
negative strand 6h
Number of reads
Raw read length - nt

Supplementary Fig. S2. A. Heatmap of MS2 sample-to-sample distances. **B-I.** Raw read length distribution histogram of sequencing basecalled reads of MS2 samples **B.** positive strand at 0 h **C**. negative strand at 0h, **D**. positive strand at 20 mins **E**. negative strand at 20 mins **F**. positive strand at 40 mins **G**. negative strand at 40 mins. **H**. positive strand at 3h **I**. negative strand at 3h **J**. positive strand at 6h **K**. negative strand at 6h.


A
RNA 0 hours
Alignment Identity (percentage)
Read Length (bases)
B
RNA 0 hours
Nanopore Read
Reference Bases
Percentage
T 0.35% 0.93% 0.13% 98.81%
G 0.38% 0.07% 99.43% 0.21%
C 0.13% 98.86% 0.06% 0.76%
A 99.15% 0.13% 0.39% 0.22%
C
RNA 0 hours
Quality score
Read Length (bases)
D
RNA 20 minutes
Alignment Identity (percentage)
Read Length (bases)
E
RNA 20 minutes
Nanopore Read
Reference Bases
Percentage
T 0.30% 0.81% 0.10% 98.87%
G 0.37% 0.06% 99.48% 0.21%
C 0.11% 99.01% 0.06% 0.74%
A 99.22% 0.13% 0.36% 0.18%
F
RNA 20 minutes
Quality score
Read Length (bases)
G
RNA 40 minutes
Alignment Identity (percentage)
Read Length (bases)
H
RNA 40 minutes
Nanopore Read
Reference Bases
Percentage
T 0.30% 0.93% 0.11% 98.89%
G 0.39% 0.06% 99.44% 0.20%
C 0.12% 98.90% 0.07% 0.68%
A 99.18% 0.11% 0.39% 0.23%
I
RNA 40 minutes
Quality score
Read Length (bases)
J
RNA 3 hours
Alignment Identity (percentage)
Read Length (bases)
K
RNA 3 hours
Nanopore Read
Reference Bases
Percentage
T 0.33% 0.83% 0.17% 98.93%
G 0.56% 0.08% 99.18% 0.25%
C 0.16% 98.97% 0.10% 0.55%
A 98.95% 0.12% 0.54% 0.28%
L
RNA 3 hours
Quality score
Read Length (bases)
M
RNA 6 hours
Alignment Identity (percentage)
Read Length (bases)
N
RNA 6 hours
Nanopore Read
Reference Bases
Percentage
T 0.38% 0.99% 0.19% 98.88%
G 0.61% 0.11% 99.10% 0.25%
C 0.21% 98.77% 0.13% 0.54%
A 98.80% 0.13% 0.58% 0.32%
O
RNA 6 hours
Quality score
Read Length (bases)

Supplementary Fig. S3. Performance metrics for Nanopore native RNA sequencing. Nanopore phage native RNA read alignment identity versus read length at different time points (0h, 20 minutes, 40 minutes, 3h, and 6h) (**Left**). Native RNA read base substitution matrix (**Middle**). The MS2 transcriptome's known base identity is represented on the x-axis. For Nanopore reads, the y-axis is the base identity at the same position. Matches or mismatches are represented by the colour intensity in the boxes, with dark red standing for a high match of correct calls. Quality score versus read length at different time points (0h, 20 minutes, 40 minutes, 3h, and 6h) (**Right**). **A-C**. At 0h p.i. **D-F**. At 20 minutes p.i. **G-I**. At 40 minutes p.i. **J-L**. At 3h p.i. **M-O**. At 6h p.i.



Supplementary Fig. S4. Raw read length distribution histogram of sequencing basecalled reads of host bacterial RNA (**A**) at 0h p.i., (**B**) 20 mins p.i., (**C**) 40 mins p.i., (**D**) 3h p.i., (**E**) 6h p.i.

A.

bits
GAAAGGGGA
1 2 3 4 5 6 7 8 9

B.

bits
CAA GAGAAAGG
1 2 3 4 5 6 7 8 9 10 11 12

C.

bits
UAGGGA
1 2 3 4 5 6

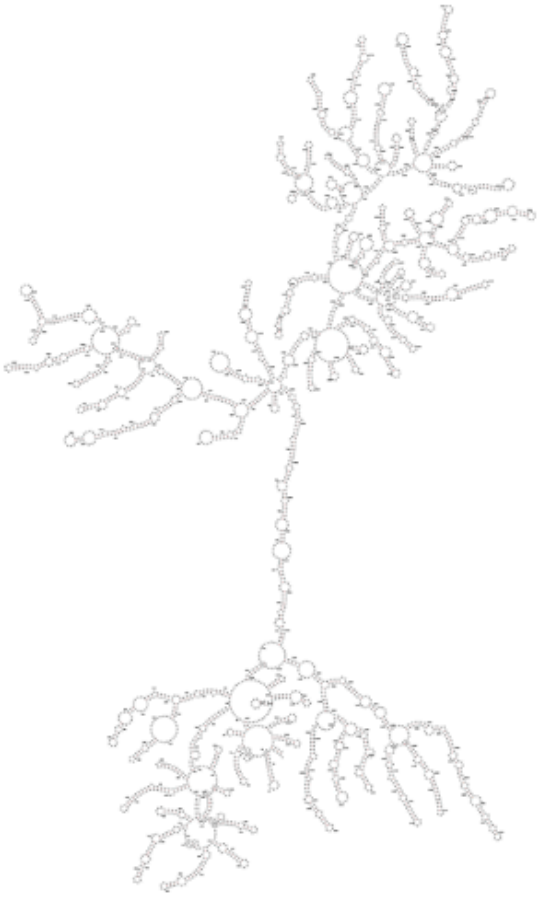
D.

bits
GAGAACG
1 2 3 4 5 6 7

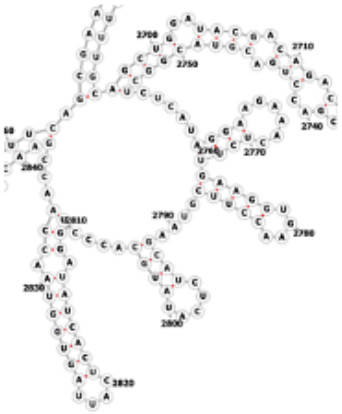
E.

bits
GGUGUA
1 2 3 4 5 6

F.

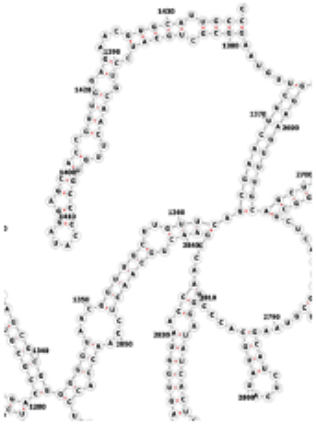


G.



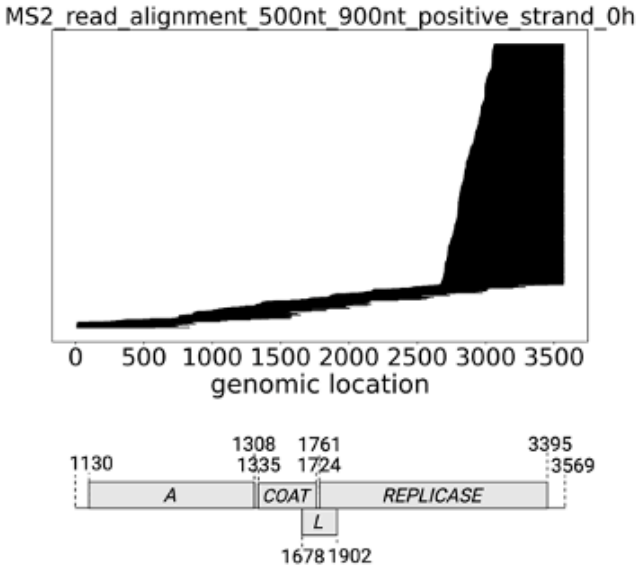
H.

I.

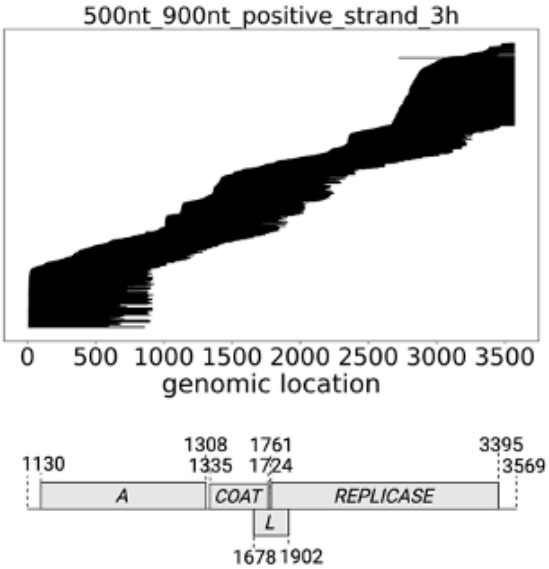


Supplementary Fig. S5 MEME alignment between putative replicase binding sites on MS2 negative strand and identified RBS sequence of genes in *Escherichia coli* K-12 MG1655 and secondary structure of template strand (negative strand) corresponding to the potential TES sites (positive-strand transcript). A. 8-16 nt, **B**. 1003-1010 nt, **C**. 1140-1145 nt, **D**. 1356-1364 nt, **E**. 1509-1514 nt are corresponding genomic locations on MS2 positive-strand, the conserved sequence showed the negative strand sequence from 5' end to 3' end on those locations. **F.** Secondary structure of native MS2 full-length negative strand, generated by RNAfold. The secondary structure showed the location of the negative strand at **G**. 2740-2840 nt, **H**. 1720-1820 nt, **I**. 1335-1435 nt, corresponding to the read end site, generated by RNAfold.

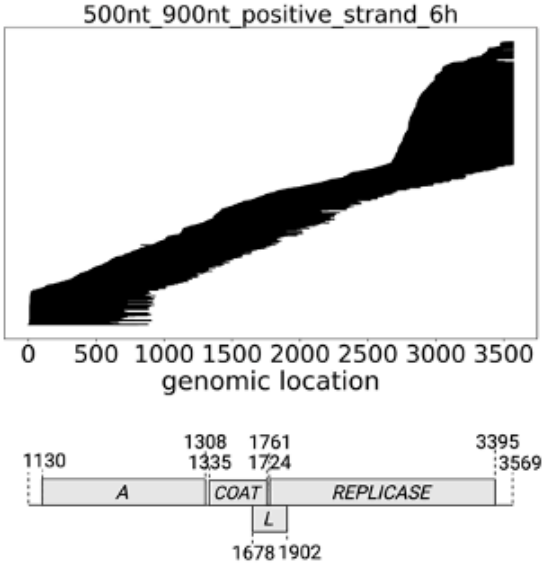
A.


MS2_read_alignment_500nt_900nt_positive_strand_0h
0 500 1000 1500 2000 2500 3000 3500
genomic location
1130
1308
1335
1761
1724
3395
3569
A
COAT
REPLICASE
L
1678 1902

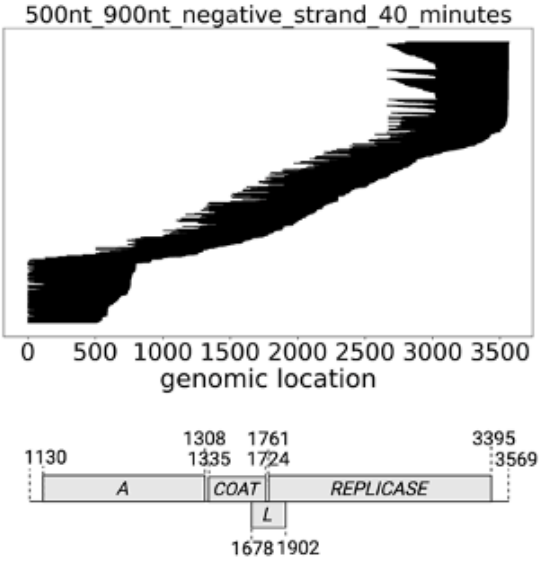
B.


500nt_900nt_positive_strand_3h
0 500 1000 1500 2000 2500 3000 3500
genomic location
1130
1308
1335
1761
1724
3395
3569
A
COAT
REPLICASE
L
1678 1902

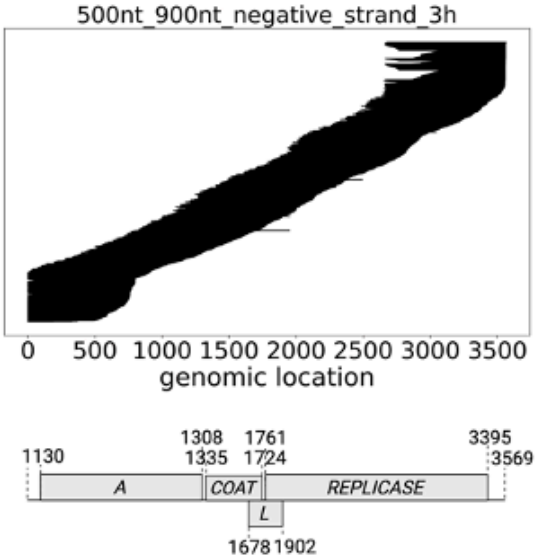
C.


500nt_900nt_positive_strand_6h
0 500 1000 1500 2000 2500 3000 3500
genomic location
1130
1308
1335
1761
1724
3395
3569
A
COAT
REPLICASE
L
1678 1902

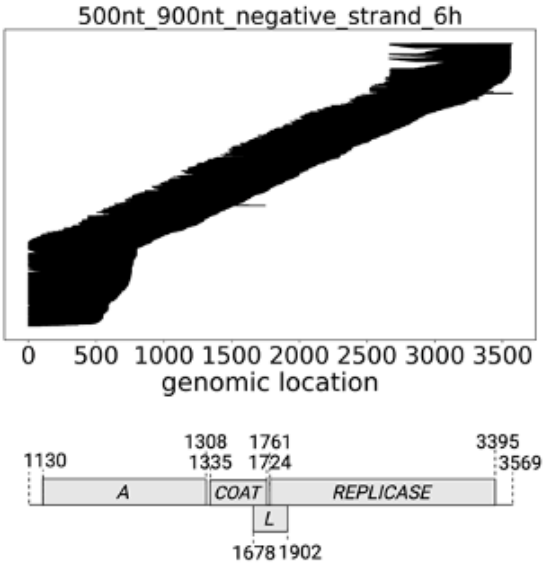
D.


500nt_900nt_negative_strand_40_minutes
0 500 1000 1500 2000 2500 3000 3500
genomic location
1130
1308
1335
1761
1724
3395
3569
A
COAT
REPLICASE
L
1678 1902

E.

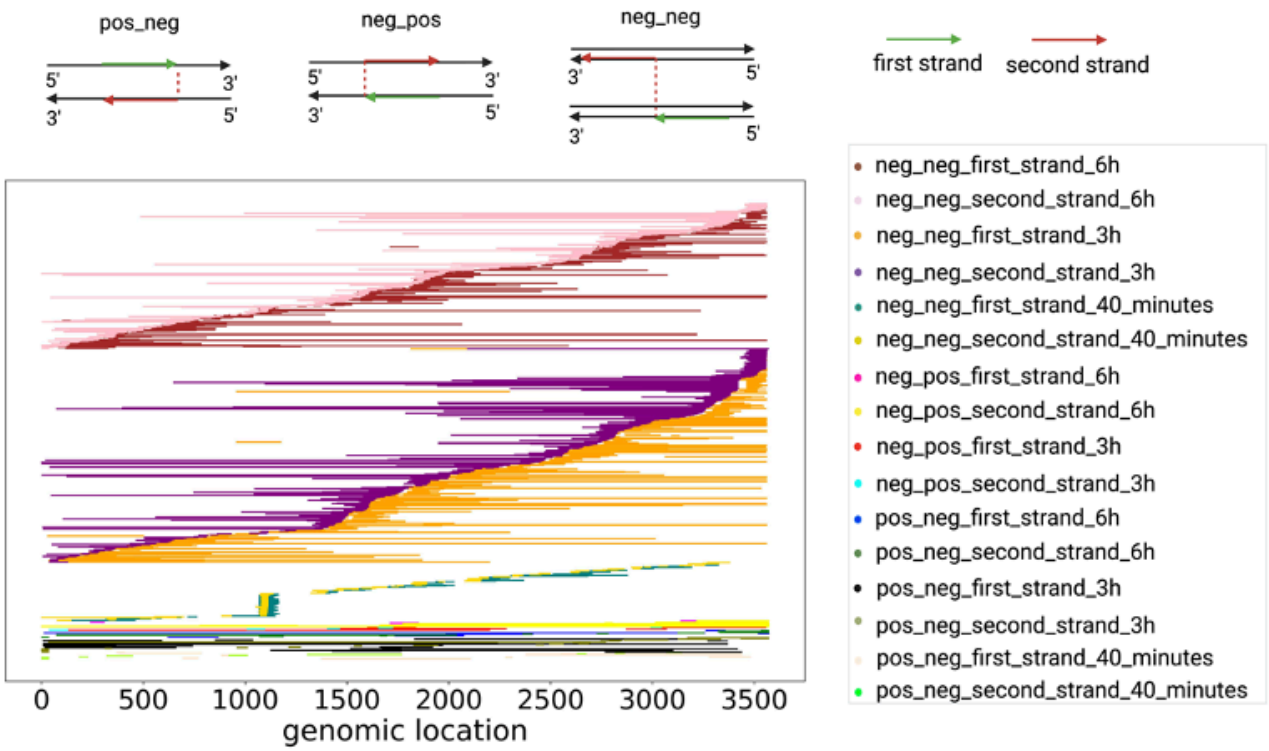

500nt_900nt_negative_strand_3h
0 500 1000 1500 2000 2500 3000 3500
genomic location
1130
1308
1335
1761
1724
3395
3569
A
COAT
REPLICASE
L
1678 1902

F.

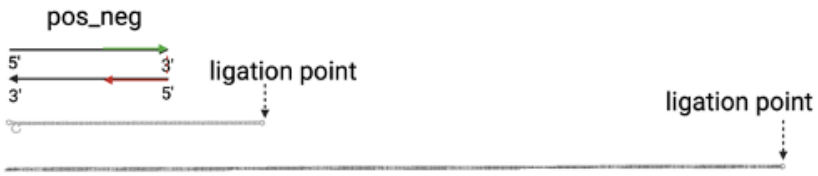

500nt_900nt_negative_strand_6h
0 500 1000 1500 2000 2500 3000 3500
genomic location
1130
1308
1335
1761
1724
3395
3569
A
COAT
REPLICASE
L
1678 1902

Supplementary Fig. S6. Reads aligned to MS2 genome (shown in lines), positive-strand or negative-strand, separated into different fractions of similar read length (500-900 nt) at 0h, 20 minutes, 40 minutes, 3h, and 6h p.i. A. Read aligned to positive-strand at 0h p.i., **B.** Read aligned to positive-strand at 3h p.i., **C.** Read aligned to positive-strand at 6h p.i., **D.** Read aligned to negative-strand at 40 minutes p.i., **E.** Read aligned to negative-strand at 3h p.i., **F.** Read aligned to negative-strand at 6h p.i.

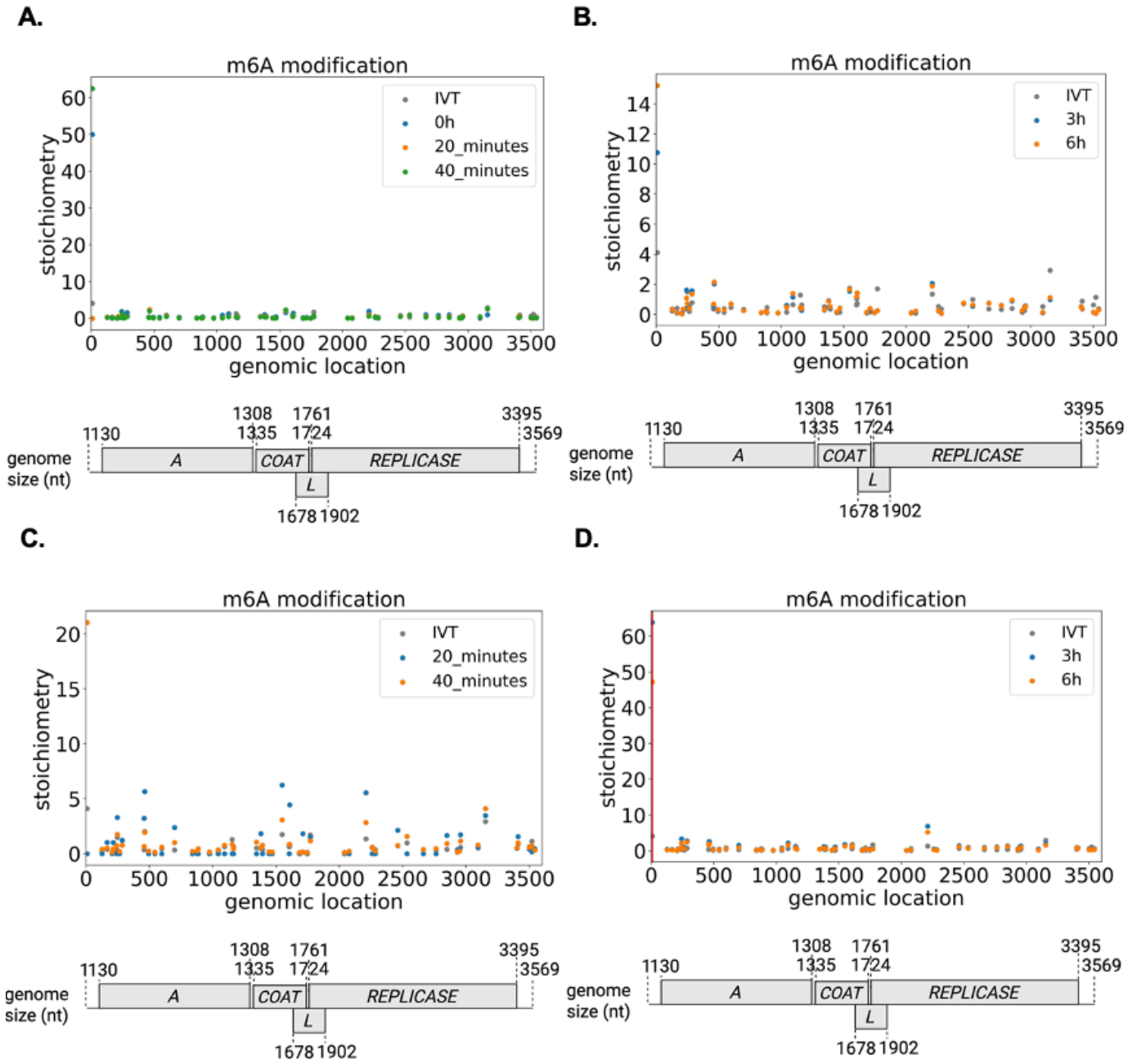
A.



B.

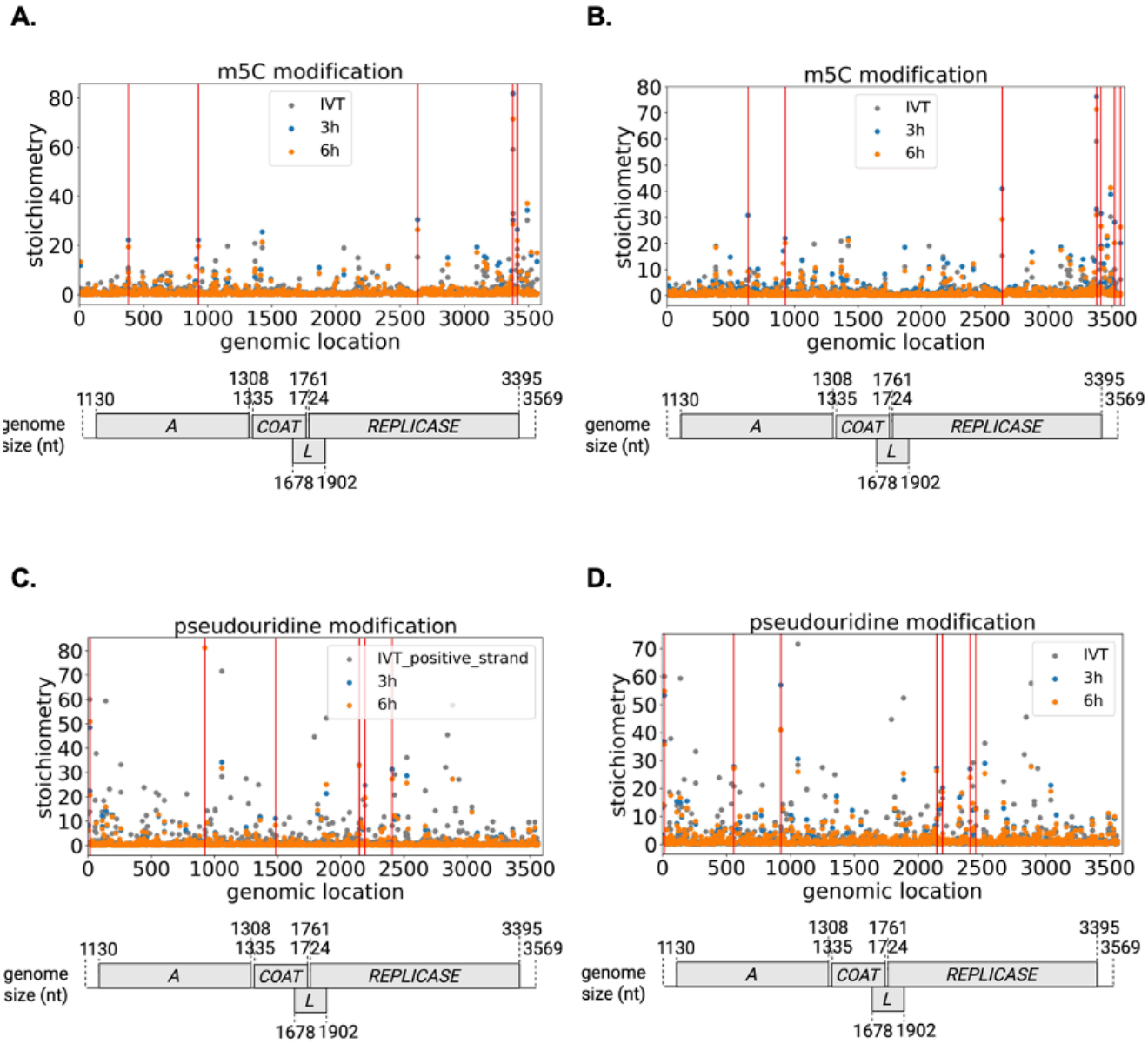


Supplementary Figure S7. Systematically summarised the characteristics of MS2 hybrid reads. A. Features of MS2 hybrid reads of neg_neg, pos_neg, and neg_pos. Total hybrid reads were visualised. **B.** Schematic diagram and secondary structure of fold-back RNAs (3' end of the positive strand ligates with 5' end of the negative strand start) by RNAfold. Positive-strand 3451-3568 ligated with negative-strand 2-135 (positive-strand location 3568-3435), and positive-strand 3173-3568 ligated with negative-strand 2-395 (positive-strand location 3568-3175).

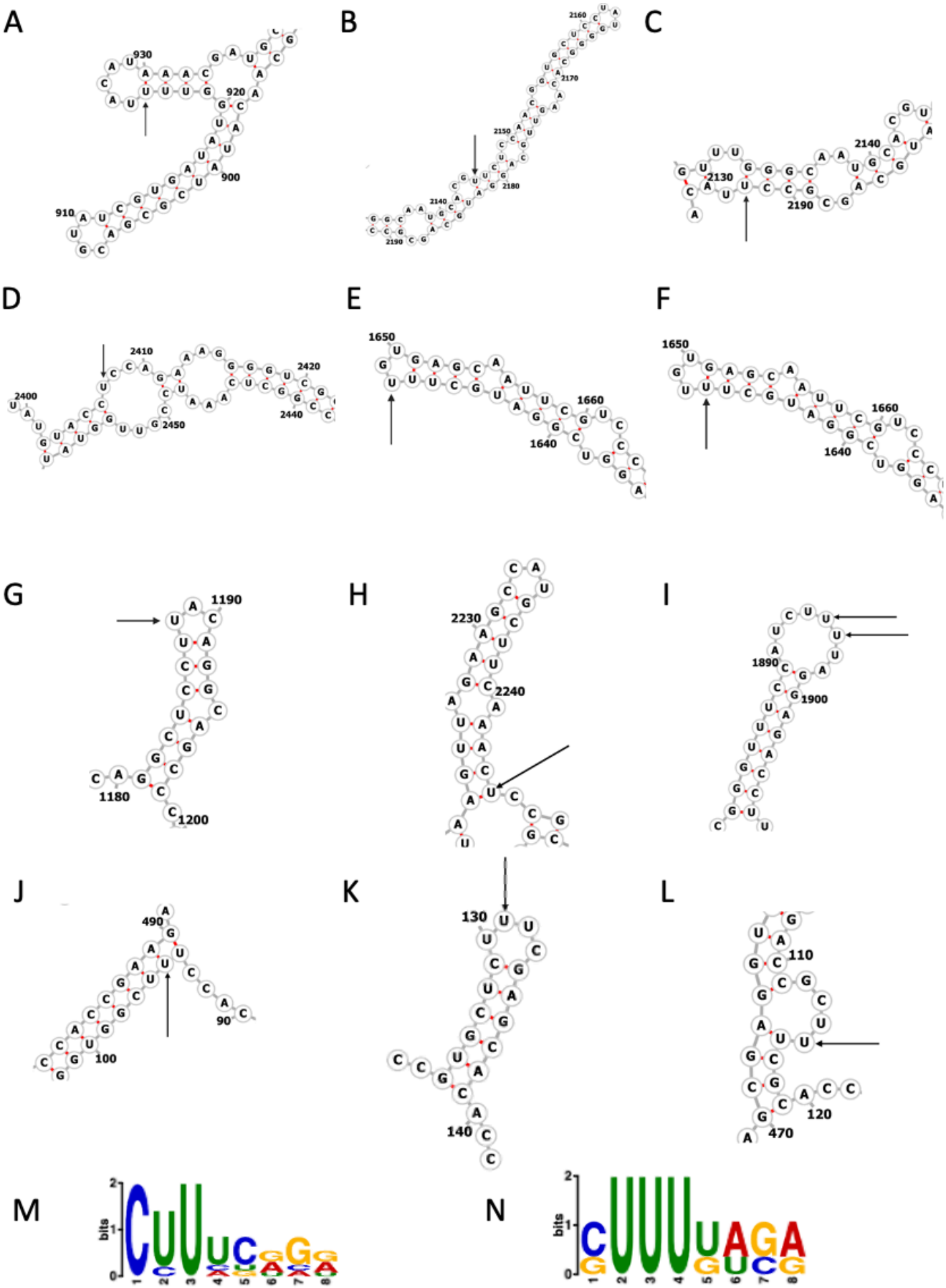


Supplementary figure S8. Frequent potential sites of m^{6}A (based on DRACH motifs) RNA modification dynamics on MS2 phage positive RNA (>3000 nt transcript length and <800 nt transcript length) via Dorado and Modkit software packages. Synthetic modification-free RNA (IVT RNA) as the negative control. A. m6A (DRACH) modifications (>3000 nt) at 0h, 20 minutes, and 40 minutes **B.** m6A (DRACH) modifications (>3000 nt) at 3h and 6h. **C.** m6A (DRACH) modifications (<800 nt) at 0h, 20 minutes, and 40 minutes **D.** m6A

(DRACH) modifications (<800 nt) at 3h and 6h. With selection criteria based on single-base analysis (those that meet all the requirements are highlighted in vertical lines): Presence of modification at 3h and 6h (vertical red line). stoichiometry (test samples) - stoichiometry (IVT samples) > 0, balanced MAP-based *p*-value < 0.05. All *p*-values pass the significance *q* value for reducing the False Discovery Rate (FDR).



Supplementary figure S9. Frequent potential sites of ψ and m^{5}C RNA modification dynamics on MS2 phage positive RNA (>3000 nt transcript length and <800 nt transcript length) via Dorado and Modkit software packages. Synthetic modification-free RNA (IVT RNA) as the negative control. A. m5C modification (>3000 nt) landscapes. **B.** m5C modification (<800 nt) landscapes **C.** ψ modification (>3000 nt) landscapes **D.** ψ modification (<800 nt) landscapes. With selection criteria based on single-base analysis (those that meet all the requirements are highlighted in vertical lines): presence of modifications at 3h or 6h (vertical red line). Stoichiometry (test samples) - stoichiometry (IVT samples) > 0, balanced MAP-based *p*-value < 0.05. All *p*-values pass the significance *q* value for reducing the False Discovery Rate (FDR).


A
930
920
900
910
B
2160
2170
2150
2180
2140
2190
C
2140
2130
2190
D
2410
2400
2420
2440
2450
E
1650
1660
1640
F
1650
1660
1640
G
1190
1180
1200
H
2230
2240
I
1890
1900
J
490
90
100
K
130
140
L
110
120
470
M
bits
N
bits

Supplementary Figure S10. Conserved secondary structure of the installation of potential ψ modifications. Positive strand **A**. at location 924, **B**. at 2145, **C**. at 2192, **D**. at 2407. Negative strand at location **E**. 1922 (1648), **F.** 1923 (1647), **G**. 2382 (1188), **H**. 1325 (2245), **I**. 1674 (1896), 1675 (1895), **J**. 3475 (95), **K**. 3439 (131), **L**. 3455 (115). **M.** Conserved recognition sequence for D, F, G, I, J, K, and L, shown in MEME motif, position 3 is ψ the site. **N**. Conserved recognition sequence for A, E, I, position 4 is ψ the site. Position 1, 5' end, position 8, 3' end.

Supplementary Tables:

parameters	0h	20_minutes	40_minutes	3h	6h	IVT
number of alignments	11038	5798	171914	1592073	1478148	3622044
median length	997.67	1142.00	400.33	468.33	453.33	391.00
mean length	930.00	954.00	733.33	668.33	745.67	474.00
median identity (%)	98.31	98.62	98.72	98.37	98.31	98.77
mean identity (%)	98.27	98.42	98.49	98.26	98.15	98.30
N50 length	1432.33	1386.25	1185.00	1070.67	1185.33	588.00
Full-length reads (positive strand + negative strand	2215	1039	24637	108380	169731	-
Full-length read ratio (%)	20.06	17.92	14.33	6.81	11.48	-

Table S1: Oxford Nanopore Technologies (ONT) DRS statistical metric for read quality control

Number of alignments

Mapped to	MS2	Bacterial genome	Bacterial palsmid	Total	MS2/Total*100%	Bacterial genome/Total*100%	Bacterial plasmid/Total*100%	MS2/Bacterial genome
0h	33114	1939187	55796	2028097	1.63%	95.62%	2.75%	1.71%
20 minutes	21414	1938841	83573	2043828	1.05%	94.86%	4.09%	1.10%
40 minutes	515742	390761	12380	918883	56.13%	42.53%	1.35%	131.98%
3h	4776219	1547128	52017	6375364	74.92%	24.27%	0.82%	308.72%
6h	4434444	1738813	46870	6220127	71.29%	27.95%	0.75%	255.03%

Table S2: Read counts from Nanopore direct RNA sequencing of total RNA from DSM5695 *E. coli* cells infected with MS2 phage at different time points.

RT-qPCR	calibrated using cysG-encoding region	fold-change (test/control)	coat/mat ratio within the same time point
3h (0h as control)	coat-encoding region as target, cysG-encoding region as reference	448.6534331	1.434321007
	mat-encoding region as target, cysG-encoding region as reference	312.7984815	
6h (0h as control)	coat-encoding region as target, cysG-encoding region as reference	468.2840244	2.200153495
	mat-encoding region as target, cysG-encoding region as reference	212.8415247	

Table S3: RT-qPCR of MS2 phage genes at different time points

Number of transcripts	0h_positive_strand	20minutes_positive_	40minutes_positive_s	3h_positive_strand	6h_positive_strand
number of transcripts <800 nt	2309	1536	86729	1645584	1267559
total number of transcripts	14874	10867	234903	2965528	2563239
number of transcripts <800 nt / total number of transcripts	0.16	0.14	0.37	0.55	0.49

Table S4: Number of transcripts (positive strand) with read length <800 nt

Number of transcripts	0h	20 mins	40 mins	3h	6h
Number of transcripts with length ranging 100 nt -800 nt	2765093	1979368	370390	1270874	1744140
Number of total transcripts	3031068	2352732	434632	1543057	1891854
proportion (%) (Number of transcripts with length ranging 100 nt -800 nt/Number of total transcripts)	91.23	84.13	85.22	82.36	92.19

Table S5: Number of transcripts with read length ranging from 100 nt to 800 nt

The RNA Integrity Number (RIN) estimated by Tapestation

samples	0h p.i. rep1	0h p.i. rep2	0h p.i. rep3	
RIN	7.8	6.9	7.7	
samples	20 mins p.i. rep1	20 mins p.i. rep2	20 mins p.i. rep3	20 mins p.i. rep4
RIN	8.4	8.6	8.2	8.1
samples	40 mins p.i. rep1	40 mins p.i. rep2	40 mins p.i. rep3	
RIN	6.5	6.4	7.2	
samples	3h p.i. rep1	3h p.i. rep2	3h p.i. rep3	
RIN	8.0	8.4	7.8	
samples	6h p.i. rep1	6h p.i. rep2	6h p.i. rep3	
RIN	8.1	7.8	7.7	

Table S6: The RNA Integrity Number (RIN) per sample estimated by tapestation

Number of transcripts

	Reads (*coat* ORF)	Reads (total)	Percentage (%)
Transcripts from positive-strand with read length <800 nt at 3h p.i.	620092480	116629530	18.81%

Table S7: Number of transcripts covering the coat protein-encoding region with read length <800 nt at the 3h time point

peaks of initiation site of transcripts	AG rich region on negative strand RNA		RBS_sequence_of_genes_in_Escherichia coli K-12 MG1655	
positive strand 5-25	3561-3554 negative strand (equivalent to positive strand 9 – 16)			
	3'GGGGAAAG5'	(negative strand 5' GAAAGGGG 3')	yfaZ	ttaattgcttgtgggaaaggggata
			iroK	actgatggcgtggaagaaaggggag
			potE	gcaccctgttgaaaggggaaaaatt
positive strand 1010-1025	2567-2560 negative strand (equivalent to positive strand 1003 – 1010)			
	3'GGAAAGAG5'	(negative strand 5' GAGAAAGG 3')	yfjT	aaaaacaacgagaaaggagacataa
			pnp	tctaagaaagagaaaggatattaca
			yajI	aattgcgacgagaaagggacgacct
			uspG	tcatgattaacagggagaaagggtt
			rhtA	agtctagatatttgtgggagaaagg
			hokB	ccgccatccggcaaggagaaaggct
positive strand 1125-1145	2431-2424 negative strand (equivalent to positive strand 1139 – 1146)			
	3'GAGGGAUG5'	(negative strand 5' GUAGGGAG 3')	yeeL	aaccagtagggaggcgaggttaggt
positive strand 1360-1375	2214-2206 negative strand (equivalent to positive strand 1356 – 1364)			
	3'AAGCAAGAG5'	(negative strand 5' GAGAACGAA 3')	yhaJ	ttcaaatttcctgaatgagaacgaa
positive strand 1510-1520	2061-2055 negative strand (equivalent to positive strand 1509 – 1515)			
	3'AUGUGGU5'	(negative strand 5' UGGUGUA 3')	dgcT	acaacctacagattggtgtagcttt

Table S8: Putative replicase binding sites on MS2 negative strand and identified RBS sequence of genes in Escherichia coli K-12 MG1655

read length ranging from 500 nt to 900 nt (For positive strand)	0h	20 mins	40 mins	3h	6h
number of alignment (total)	1040	732	15107	15901	20471
number of alignment (mapped to the 3' end)(enrichment due to Nanopore sequencing 3' bias)	866	574	6424	3172	5996
proportion (%) (number of alignment (mapped to the 3' end)/number of alignment (total)*100%)	83.27	78.42	42.52	19.95	29.29
number of alignment (mapped to the 5' end)(degraded RNA)	21	39	4583	3075	2290
proportion (%) (number of alignment (mapped to the 5' end)/number of alignment (total)*100%)	2.02	5.33	30.34	19.34	11.19
number of alignment (mapped to the coat protein encoding region)	20	20	511	389	842
proportion (%) (number of alignment (mapped to the coat protein encoding region)/number of alignment (total)*100%)	1.92	2.73	3.38	2.45	4.11
number of alignment (mapped to the lysis protein encoding region)	22	11	828	1773	1630
proportion (%) (number of alignment (mapped to the lysis protein encoding region)/number of alignment (total)*100%)	2.12	1.50	5.48	11.15	7.96
number of alignment (mapped to the coat and lysis protein encoding region)	6	1	186	311	216
proportion (%) (number of alignment (mapped to the coat and lysis protein encoding region)/number of alignment (total)*100%)	0.58	0.14	1.23	1.96	1.06

Table S9: Number of transcripts with read length ranging from 500 nt to 900 nt aligned to different regions on the MS2 genome (positive strand) at different time points

Read length ranging from 500 nt to 900 nt (For negative strand)	40 mins	3h	6h
number of alignment (total)	1708	18074	23389
number of alignment (mapped to the 5' end of negative strand)(degraded RNA)	465	2851	2459
proportion (%) (number of alignment (mapped to the 5' end)/number of alignment (total)*100%)	27.22	15.77	10.51
number of alignment (mapped to the 3' end of negative strand)(enrichment due to Nanopore sequencing 3' bias)	156	1078	2362
proportion (%) (number of alignment (mapped to the 3' end)/number of alignment (total)*100%)	9.13	5.96	10.10

Table S10: Number of transcripts with read length ranging from 500 nt to 900 nt aligned to different regions on the MS2 genome (negative strand) at different time points

Number of hybrid reads	20 minutes	40 minutes	3h	6h
deletions	22	68	3335	3848
duplications	42	262	5417	5585

Table S11: Number of deletion and duplication events

RT-qPCR primers		Amplicon size
Mat_Fwd_RTqPCR	TGTGGAGAGACAGGGCACTG	77 bp
Mat_Rev_ RTqPCR	CAGTTGTTGGCCATACGGATT	
Coat_Fwd_RTqPCR	TCGCGTTCACAGGCTTACAA	134 bp
Coat_Rev_RTqPCR	GCGGCTACAGGAAGCTCTAC	
cysG_Fwd_RTqPCR	GAAGCGCAAAAAGGCAAACG	92 bp
cysG_Rev_RTqPCR	TTGCACAGTGTTTCCAGCTC	
RT primers for in vitro transcription		
RTprimer-1	GGGTGGGACCCCTTTCGG	
RTprimer-2	TCTATCTAGAGAGCCGTTGCCT	
RTprimer-3	CATAAGCGTTGACGCTCC	
RTprimer-4	GACTCGTTATAGCGGACC	
RTprimer-5	GGCACAAGTTGCAGGATGCA	
RTprimer-6	TGGGTGGTAACTAGCCAAGCAG	
PCR primers for in vitro transcription		
IVT-frag1-F	GCATTAATACGACTCACTATAGGGAGAGGGTGGGACCCCTTTCGG	
IVT-frag1-R	TCTATCTAGAGAGCCGTTGCCT	
IVT-frag2-F	GGGTGGGACCCCTTTCGG	
IVT-frag2-R	GCATTAATACGACTCACTATAGGGAGATCTATCTAGAGAGCCGTTGCCT	
IVT-frag3-F	GCATTAATACGACTCACTATAGGGAGACATAAGCGTTGACGCTCC	
IVT-frag3-R	GACTCGTTATAGCGGACC	
IVT-frag4-F	CATAAGCGTTGACGCTCC	
IVT-frag4-R	GCATTAATACGACTCACTATAGGGAGAGACTCGTTATAGCGGACC	
IVT-frag5-F	GCATTAATACGACTCACTATAGGGAGAGGCACAAGTTGCAGGATGCA	
IVT-frag5-R	TGGGTGGTAACTAGCCAAGCAG	
IVT-frag6-F	GGCACAAGTTGCAGGATGCA	
IVT-frag6-R	GCATTAATACGACTCACTATAGGGAGATGGGTGGTAACTAGCCAAGCAG	

Table S12: Primers used in the study