

RESEARCH

Open Access



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

Ni Ni¹ and Gaetan Burgio^{1*}

Abstract

Background 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.

Methods 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.

Results 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 the error-prone activity of the MS2 replicase produced hybrid reads from the template-switching mechanism. We finally 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, from potentially the host RluA-mediated installation.

Conclusions Overall, our investigation revealed a more complex transcriptional dynamics of MS2 phages than anticipated within *E. coli* to maintain its growth and replication under host pressure.

Keywords RNA phage, MS2, Nanopore direct RNA sequencing, Transcription, Replication, Hybrid transcripts, RNA modification

*Correspondence:

Gaetan Burgio
Gaetan.burgio@anu.edu.au

¹Division of Genome Sciences and Cancer, The Shine Dalgarno Centre for RNA Innovation, John Curtin School of Medical Research, the Australian National University, Canberra, Australia



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

Viruses of bacteria (bacteriophages) are the most abundant viruses on Earth [1]. Phages play an important role in the regulation of biological processes and have been harnessed for biotechnology applications [2]. 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 [3]. It is thought that the ssRNA phages are the oldest lineage of RNA viruses still in existence [4]. These have proven to be useful models for comprehending basic molecular mechanisms such as RNA genome replication, translational control, and gene regulation [5, 6]. MS2 belongs to the family *Fiersviridae*, and is an icosahedral, non-enveloped, F-specific, positive-sense single-stranded RNA (+ ssRNA) lytic coliphage [7]. 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 (Fig. 1A) [8]. 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 (Fig. 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 [9]; *coat* encodes the coat protein (CP), which forms the viral capsid; *lys* encodes lysis protein (L), which allows for host lysis [10]; and *rep* encodes catalytic subunit of replicase (Rep), which is an RNA-dependent RNA polymerase [8] (Fig. 1B).

Previous research revealed RNA–protein interactions that regulate the precise timing and strength of viral protein production throughout the bacteriophage life cycle [6, 11]. 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 [8]. The replicase complex is repressed by the CP, which binds to specific sites on the positive-strand RNA [12]. While the interplay between different phage proteins resulting in entry, replication, assembly, and encapsidation has been relatively well studied in *in vitro* settings [6, 8], the dynamics in transcription throughout the MS2 life cycle remain elusive. Specifically, 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 [13], recombination [14], and RNA modification landscapes of various viruses [14].

Here, we employed Nanopore DRS to investigate the transcriptome and epitranscriptome landscape of the MS2 in infected *E. coli* throughout its life cycle. We discovered that MS2 phages sustain a high level of transcriptional activity within the bacterial cell. We found sub-genomic transcripts resulting from either RNA degradation or Nanopore DRS bias, adding difficulties in the analysis of the transcriptional dynamics. Interestingly, we found the presence of small sub-genomic transcripts encoding the full-length coat protein ORF. The MS2 replicase is error-prone, which frequently generates hybrid RNAs across time by a 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 modified sites. The conserved sequence and secondary structure (U-rich stem-loops) of Ψ installation sites were the most amenable to RNA modification and likely mediated by host pseudouridine synthase RluA. Using DRS RNA Nanopore sequencing, we have dissected the dynamics of MS2 transcriptional activity in bacterial cells across time.

Methods

Bacteria, phage strains, and culture conditions

The bacteria host *E. coli* DSM 5695 (also called *E. coli* W1485) (F⁺) and MS2 bacteriophages DSM 13,767 were obtained from the DSMZ-German Collection of Microorganisms and Cell Culture. The bacteria were maintained in NZCYM media (10.0 g/L Casein hydrolysate, 5.0 g/L NaCl, 1.0 g/L Casamino acids, 5.0 g/L Yeast extract, 2.0 g/L MgSO₄ × 7H₂O) with maltose (2 g/L) and incubated at 37 °C overnight. MS2 phages were maintained and propagated in SM buffer (NaCl 100 mM, MgSO₄ × 7H₂O 8 mM, Tris-HCl 50 mM, gelatin 0.01% (w/v), pH 7.5). MS2 plaque assays were performed in a double-layer Agar. Briefly, 100 μ l overnight culture of DSM 5695 bacteria and 100 μ l MS2 in SM buffer were mixed with 5 ml NZCYM soft agar with maltose (2 g/L). Plates were incubated at 37 °C overnight (~16 h). MS2 phages were harvested. The top layer was disrupted after

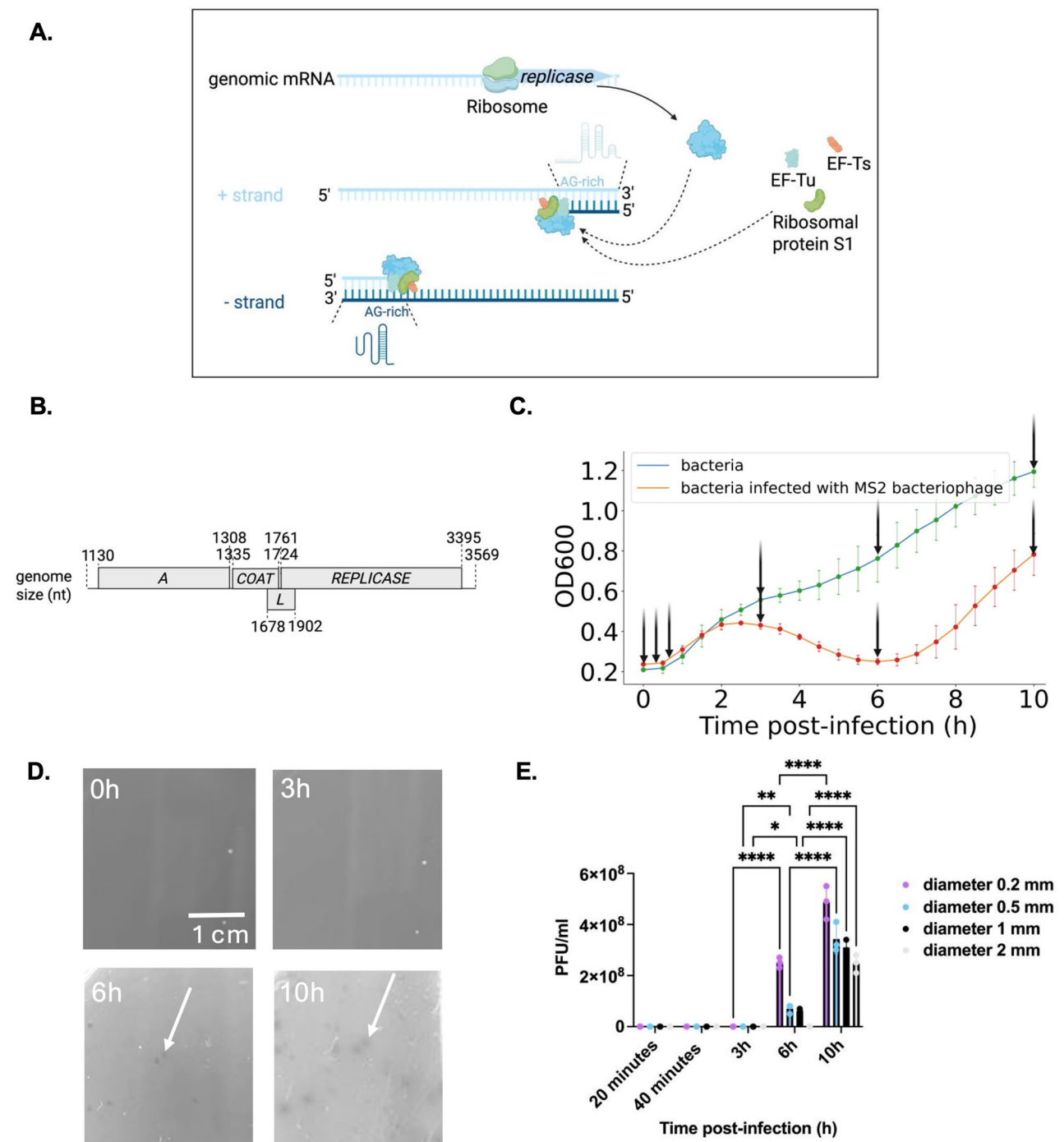


Fig. 1 Characteristics of MS2 phage. **A** MS2 replication pattern. The MS2 replicase complex contains host factors EF-Tu, EF-Ts, and S1 ribosomal protein. **B** Schematic presentation of the MS2 full-length genomic RNA coding regions organisation and composition. A, maturation protein-encoding gene; COAT, coat protein-encoding gene; L, lysis protein-encoding gene; REPLICASE, replicase protein-encoding gene. NTD, N-terminal domain; CTD, C-terminal domain. **C** Growth curves of bacterial cells (DSM5695 *E. coli*) infected with MS2 RNA phage at MOI=200 (red line) and bacterial cells without infection (blue). **D** MS2 plaques on the host (*E. coli* DSM5695). Plaques were pointed with white arrows. **E** Plaque PFU and size change. All pair-wise comparisons were performed after observing a significant Analysis of Variance (ANOVA) with * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$. Data are presented as means $\pm$ SD ($n = 3$)

pouring 5 ml SM buffer, and the lysate was centrifuged at 2,800 g for 30 min at room temperature. The lysate was filtered using a 3 ml syringe with a 0.22 μ M filter. The precise viral stock titer (PFU/ml) of the phage (MS2) was determined.

MS2 phage infection and RNA isolation

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

Reverse transcription quantitative PCR (RT-qPCR) of MS2 phage genes

The primer pairs for RT-qPCR were designed specifically to the MS2 gene regions using RefSeq Representative Genome Database (Organism limited to *Escherichia coli*) with the NCBI Primer-BLAST tool. The *E. coli* *cysG* gene (housekeeping gene) was selected as the reference gene. Briefly, reverse transcription for 100 ng whole RNA extracted was performed with SuperScript™ IV Reverse Transcriptase [Invitrogen] with random primers (hexamers) [Invitrogen] at 50 °C, then qPCR was performed with PowerTrack™ SYBR Green [Invitrogen] using specific sets of primers (Suppl. Table S12) for specific phage genes (400 nM primers with annealing temperature 60 °C) with input cDNA (serially diluted to determine the qPCR primer efficiency) according to the manufacturer's protocol and visualized by QuantStudio™ 12 K Flex Real-Time PCR System. Data analysis using the $\Delta\Delta$ CT method was performed to quantify the relative expression of MS2 phage genes.

In vitro transcription (IVT) of MS2 RNA

One μ g of total RNA was purified from MS2-infected *E. coli* cells. Reverse transcription (SuperScript™ IV Reverse Transcriptase [Invitrogen]) was performed with each phage-specific RT primer (RTprimer1-6) (Suppl. Table S12). Templates for in vitro transcription (IVT) were prepared by PCR with each phage-specific PCR primer pair (i.e., IVT-frag1-F & IVT-frag1-R primers with 5' terminal

T7 RNA polymerase promoter region) followed by agarose gel purification (Wizard SV Gel and PCR Clean-Up System kit [Promega]), in vitro transcription using T7 polymerase (MEGAscript™ T7 Transcription Kit [Thermo Fisher]) and RNA purification (Monarch RNA Cleanup Kit [NEW ENGLAND Biolabs]) with re-suspension in ddH₂O. The oligonucleotides used in this study are listed in the Suppl. Table S12.

Poly(A) tailing of RNA and beads purification

Total RNAs and IVT RNAs were incubated at 65 °C for three min and cooled on ice for one min before polyadenylating RNAs using the *E. coli* poly(A) polymerase (NEB M0276S [NEW ENGLAND Biolabs]). Briefly, 10 μ g RNA, 5 units poly(A) polymerase, 2 μ L 10X reaction buffer, and 1 mM ATP in a total reaction volume of 20 μ L were incubated for 45 min at 37 °C. To stop and clean up the reaction, poly(A)-tailed RNAs were purified following the Agencourt AMPure XP magnetic beads (1:1.8 RNA to beads ratio) according to the manufacturer's protocol and resuspended in 10 μ L of ddH₂O. The integrity of total RNA from *E. coli* was assessed via a Bioanalyzer (Agilent).

Direct RNA nanopore sequencing

The Direct RNA Sequencing Kit (SQK-RNA004 by Oxford Nanopore) was used to prepare the libraries and sequence the native poly(A)-tailed RNA library according to the manufacturer's protocol and including the reverse transcription step to generate RNA: cDNA hybrids. Sequencing runs were performed using PromethION flow cells (FLO-PRO004RA). The Flow Cell Wash Kit (EXP-WSH004) was used to remove the library between flow cell washes and library reloading, following the manufacturer's instructions.

Basecalling and read mapping

The Computational workflow for processing the Oxford RNA-direct Nanopore sequencing datasets was shown (Suppl. Figure S1). Basecalling was performed using MinKNOW software (Nanopore) using the Super accurate model. The Fast5 files were filtered. Reads with a minimum PHRED score of 10 were retained. Fast5 files were then converted to pod5 using pod5_file_convert script using the `–one-to-one` option. Basecalling on pod5 files was performed with Dorado (v 0.9.1) using the rna004_130bps_sup@v5.1.0 basecalling model for m5C, m6A_DRACH, Inosine_m6A, PseuU. Reads were mapped to the reference sequence NC_001417.2 using minimap2 [15] using the following parameters: `–ax map-ont –y –secondary = no –k 4 –w 10`. Samtools [16] was used to retain the mapped reads either on the positive strand (Samtools view `–F2324 –b`) or negative strand (Samtools view `–f 16 –b`) and reverse complemented.

Reads were sorted and indexed. Mapped read lengths were counted using `genomecov` function from `Bedtools` [17]. Other metrics, such as alignment identity, were determined using `MarginStats` [18] and based on a similar workflow previously described [19].

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` (options `-ax map-ont -y --secondary = no -k 4 -w 10`). Mapped reads were counted with `Salmon` [20] using the options: `--ont -P 48 -l U`. Gene expression analysis was performed with `Deseq2` [21] 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 reads at these specific window positions was quantified using `Samtools` and a custom bash script. Stop and start codons were determined using a motif search and overlaid 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 $e\text{-values} > 10^8$ 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.

Detection of single nucleotide and structural variants

To detect single nucleotide variants and indels, we performed the analysis using `Clair3` [22] and `Sniffles2` [23]. We included the following flags for `Clair3` analysis: `--platform="ont" --model_path=r941_prom_sup_g5014 --include_all_ctgs --no_phasing_for_fa`. For indels detection, we performed the analysis using `Sniffles2` with default parameters. The files were filtered and converted using `Bcftools` [24] in `vcf` format.

RNA modification analysis

Basecalled reads from Dorado were mapped to the NC_00141702.2 genome using `Minimap2`, sorted, and indexed using `Samtools`. Basecalling for modification was determined using `Modkit` `pileup` function using

`--max-depth 2,000,000` option and a modification threshold $> 80\%$. Differential methylation profile was determined using `modkit dmr` function. The Nanopore DRS technique has the limitation of high false-positive rates (FPR), and the IVT samples were prepared to reduce the FPR. Only `Stoichiometry (test) - Stoichiometry (IVT) > 0` was included in our study.

Statistical analysis

Statistical analysis was conducted. The mean $\pm$ standard deviation (SD) was shown. The differentially expressed genes were identified by assessing the statistical significance via the adjusted p -value after accounting for multiple hypothesis testing (Benjamini-Hochberg method). The threshold for statistical significance was set at P value < 0.05 , *, P value < 0.01 , **. The Benjamini-Hochberg method was performed to control the false discovery rate (FDR correction) for post-transcriptional modifications [25]. 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 h 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 (Fig. 1C). MS2 formed plaques from 6 h p.i., indicating its lytic nature (Fig. 1D). The size of MS2 plaques ranged from 0 to 2 mm in diameter within 10 h (Fig. 1D-E). Overall, the data indicated that the time points 20 min (before the first burst), 40 min (after the first burst), 3 h (maximum growth), and 6 h (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.

MS2 phage maintained a high transcriptional activity across time

To capture MS2 transcriptional changes across time, we performed long-read Nanopore direct RNA sequencing (DRS) runs on total RNA extracted from DSM5695 *E.*

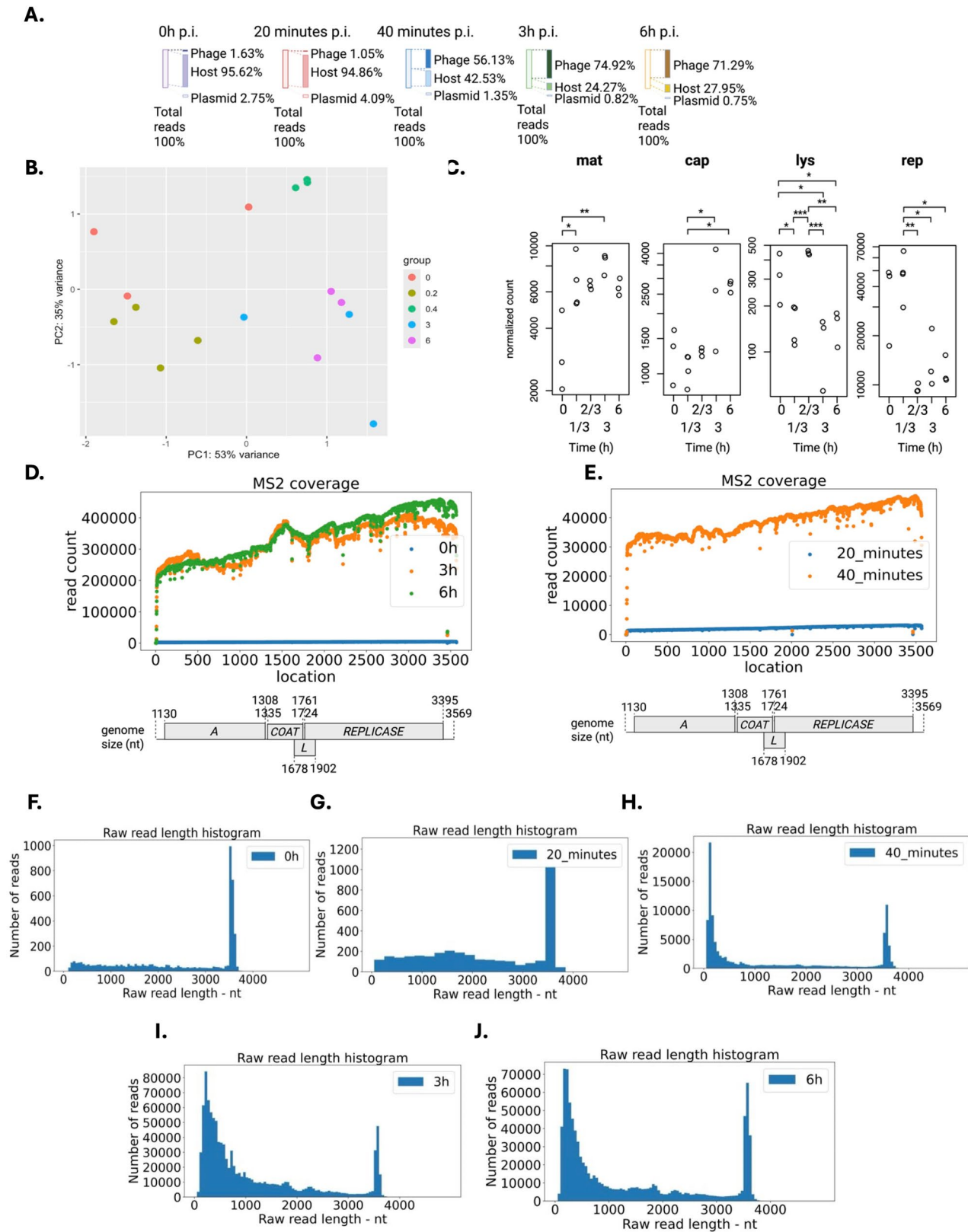


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Statistics of Oxford Nanopore Technologies (ONT) direct RNA sequencing data via computational approaches and transcription pattern of MS2. **A.** Read counts from Nanopore direct RNA sequencing of total RNA from DSM5695 *E. coli* cells infected with MS2 phage at different time points. **B.** Principal component analysis (PCA) plot of each MS2 sample. **C.** Differential expression patterns for four MS2 genes at different time points based on normalised read count. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **D.** Coverage map for MS2 total reads at 0 to 6 h p.i., **E.** 20 and 40 min p.i., **F.** MS2 raw read length distribution at 0 h p.i. **G.** At 20 min p.i. **H.** At 40 min p.i. **I.** At 3 h p.i. **J.** At 6 h p.i.

coli cells infected with MS2. We mapped over 17,6 million reads over a PHRED quality score threshold of 10. Out of 17,6 million reads, 7.5 million reads were mapped on the MS2 genome, and we found an average read length spanning from 668 to 954 nt (Suppl. Table S1). Reads over 90% of the MS2 transcript were included as full-length reads (Suppl. Table S1). Immediately after infection (0 h p.i.), only full-length RNAs were detected (which accounted for only 1.63% of the total viral reads) (Fig. 2A and Suppl. Table S2) and almost no observable negative-strand transcripts (Suppl. Figure S2C). A sudden increase in the ratio of viral RNAs to bacterial RNAs occurred from 20 min (1.05%) to 40 min (56.13%), suggesting the first cycle burst occurred at ~30 to 40 min post-inoculation (Fig. 2A and Suppl. Table S2). Full-length negative-strand RNA was observed at 40 min, 3 h, and 6 h p.i. (Suppl. Figure S2 G, I, K, and Suppl. Table S1). Gene expression analysis was conducted to assess variation between different time points. From the Principal Component Analysis (PCA), the observed variation on the first two PC axes accounted for 88% of the total variance (Fig. 2B). Three main groups were identified: the 0 h and 20 min p.i. on the PC1 negative axis and the infected group over 40 min p.i. on the positive PC1 axis (Fig. 2B). Among the infected group, the 40-minute time point group was on the PC2 positive axis, whereas the 3 and 6 h p.i. were clustered on the negative PC2 axis (Fig. 2B). The different time points exhibited distinct clusters on the correlation heatmap (Suppl. Figure S2A), confirming significant transcriptional changes of MS2 across time. Next, we analyzed the differential expression (DE) of the four MS2 ORFs across time (Fig. 2C). We found an upregulation of the *mat* and *cap* ORFs while the *rep* ORF was downregulated (Fig. 2C). Interestingly, we found a peak expression of the *lys* gene at 40 min p.i. We confirmed by RT-qPCR similar results to the Nanopore DRS (Suppl. Table S3). Compared with 0 h p.i., we found a significant fold increase of >400-fold at 3 and 6 h p.i. of the *coat* ORF (Suppl. Table S3). The results also indicated the relatively higher expression of phage RNA of the coat-protein ORF than the maturation-protein ORF at the same time point (Suppl. Table S3), confirming our observations from Nanopore DRS. Overall, our results suggest a significant up regulation of the *coat* gene and fine tuning of the ORFs expression to regulate MS2 transcription across time.

Read mapping revealed a bias in coverage towards the coat protein ORF

During replication, MS2 forms an intermediate negative transcript that serves as a template for transcription/translation of the ORFs [8]. We separated the positive and negative strands, and mapped reads to capture the intermediate state of viral replication. Next, we assessed the read coverage across time for all reads (Fig. 2D-E), positive and negative strands (Fig. 3A-B). We observed a coverage bias towards the *coat* ORF at 3 h and 6 h p.i. and the 3' end of the viral genome on the overall transcript (Fig. 2D), further confirming our results. As the nanopore DRS has a 3' bias due to directional sequencing from the 3' ends of RNAs [26], we deduced that coverage was biased only towards the *coat* ORF in the MS2 transcriptome.

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 (Fig. 2F-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 min p.i. (Fig. 2H-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 proportionally 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 min to 3 h 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. While the number of short bacterial transcripts did not increase across time, we found few small viral transcripts (100–1000 nt) at 20 min p.i. (Suppl. Figure S2 D) and a sudden surge in short viral RNA from 40 min p.i. after the first cycle burst had occurred. Together, the results suggested that the degradation due to lysis after phage burst plays a minor role in both the host and the MS2 RNA when analysing 40 min, 3 h and 6 h p.i. data. 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), quality score versus read length and the

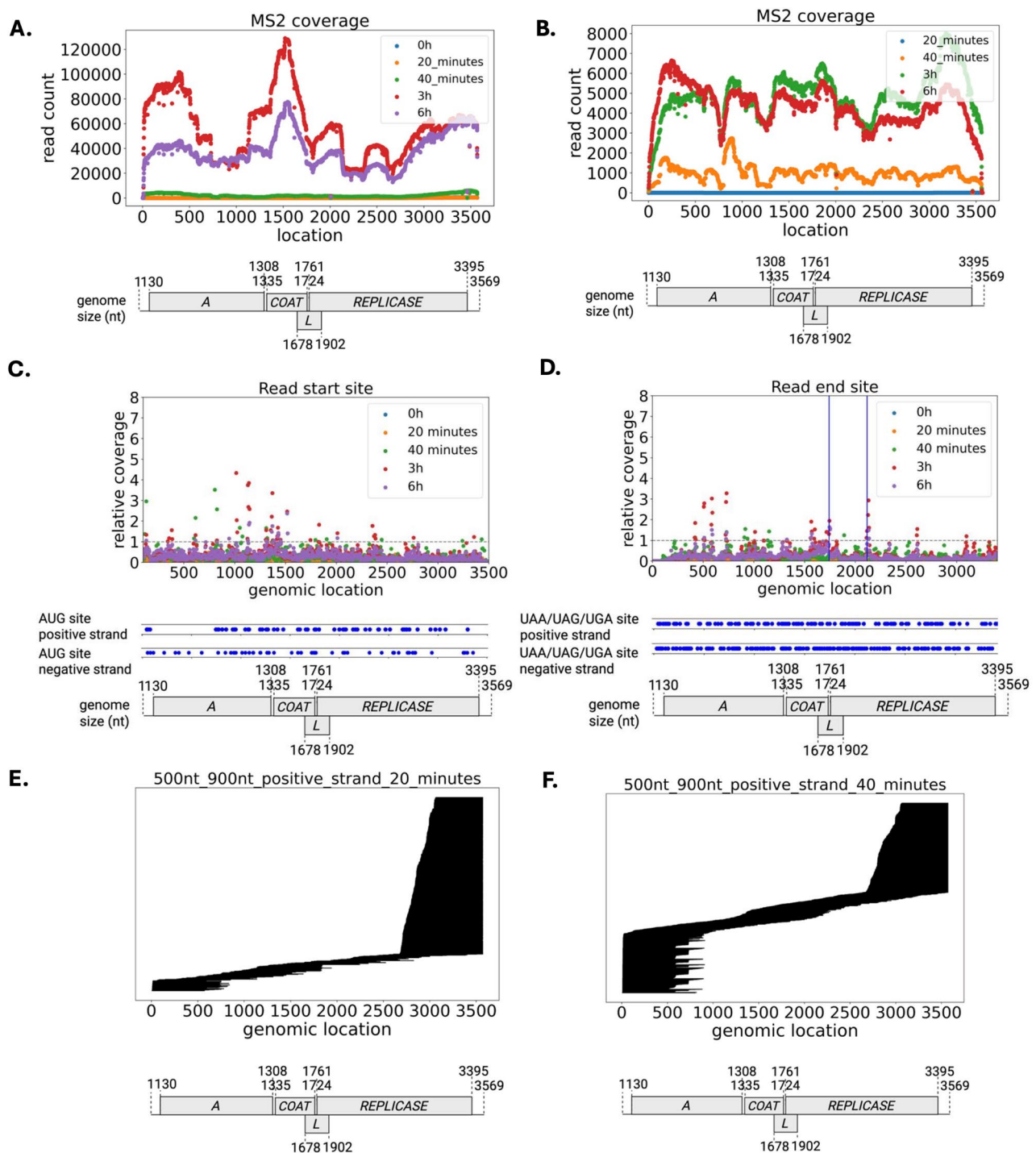


Fig. 3 Replication pattern of MS2. **A.** Coverage map for MS2 reads (<800 nt) aligned to the positive strand and **B.** the negative strand. **C.** Relative read start site (ranging 100–3500 nt) coverage profile and **D.** Relative read end site (ranging 0–3400 nt) coverage profile for MS2. The blue vertical line indicated two previously identified key coat protein binding sites. **E.** Positive-strand reads (shown as lines) aligned to the MS2 genome with read length between 500 nt and 900 nt at 20 min p.i., and **F.** Read length between 500 nt and 900 nt at 40 min p.i.

number of mismatches per read to the reference transcript. These measures are 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 minimise 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 h p.i. to 1,185 nt at 6 h p.i.

AG-rich regions were responsible for the initiation of small transcripts

Previous studies described the ability of the replicase to generate small spontaneous template-directed RNA species [8, 27]. We hypothesized that these small transcripts could be either RNA species resulting from Nanopore DRS 3' bias and library preparation, small transcripts required for replication, and/or products of RNA degradation from bacterial lysis.

To further investigate the nature of these small transcripts, we quantified their coverage across the MS2 genome (Fig. 3A-B). We found a >2 fold enrichment of the 5' end and the *coat* ORF (~18.8% of total positive-strand reads with read length <800 nt at 3 h) on the positive strands (Fig. 3A and Suppl. Table S7), confirming the initial mapping and RT-qPCR results (Suppl. Table S3). We also found enrichment of the 3' end of the full transcript, likely due to the DRS bias. Interestingly, transcripts mapped to the 3' end of the *rep* ORFs were enriched on the negative strand (Fig. 3B), whereas part and the 3' end of the *mat* ORF lacked coverage on the positive strand (Fig. 3A). To detail the mapping of these <800 nt transcripts, we performed a genome-wide mapping of the start and end of these transcripts within a 5 bp size window and calculated the fold changes relative to the overall coverage. We noticed peaks of significance on the positive and negative strands (Fig. 3C-D). We notably observed hotspots at ~1,000–1,300 bp within the *mat* ORF and hotspots of end sites at ~1740–1745 nt after the *coat* ORF and at 2125–2130 nt within *rep* ORF (Fig. 3C-D).

We next postulated that specific RNA signals (ribosome binding sites/codon start) are the key determinants for MS2 transcript initiation of these < 800 nt transcripts. We investigated several start site peaks by assessing the genomic composition and secondary structure of their corresponding genomic locations. We found AG-rich regions near these hotspots on the negative strand. Using Clustal alignment and MEME, we found that the MEME motif (Suppl. Figure S5 A-E) of these regions superimposed to known ribosomal binding sites (RBS) [28] (Suppl. Table S8). Together, our results suggest the production of these small transcripts occurs during the phage life cycles.

A population of subgenomic reads is produced during the MS2 life cycles

To replicate its own transcript, MS2 generates a negative-strand transcript, which initiates the replication from the 3' to the 5' end [8]. Prior work on Nanopore sequencing technology has demonstrated that degraded RNAs/

random cleavages during library preparation exhibit a trimmed 3' end [29]. We postulated that transcripts mapped starting from the 3' end resulted from Nanopore DRS 3' end bias since the sequence starts from the 3' end, whereas transcripts mapped from the 5' end were degraded transcripts. We aligned these short reads to the MS2 transcript. We found three types of aligned reads (Fig. 3E-F). The first block aligned from 3' to 5' end accounted for 78.42% of the read count (length ranging 500nt – 900 nt) at 20 min p.i. (Suppl. Table S9) and reduced in proportion after the first cycle. These reads corresponded to Nanopore 3' end bias. The second block aligned from 5' to 3' end accounted for degraded RNA and increased over time, likely due to active lysis in bacteria and potentially DRS sample preparation. Previous studies identified that the coat protein subunit binds to the RNA template to mediate and regulate the viral assembly and packaging during replication [30]. Interestingly, a third block of aligned transcripts aligned to the *coat* and the *lys* ORFs and increased across time (Fig. 3F and Suppl. Figure S6B). This observation was in accordance with our previous results on the gene expression analysis backed by RT-qPCR (Fig. 2D). Within one life cycle, we observed an increased count of transcripts containing the coat protein encoding region, from 1.92% for 0 h, 2.73% for 20 min, and 3.38% for 40 min (Suppl. Table S9). For transcripts containing both the coat and lysis protein encoding region, we observed an increase of 0.14% at 20 min to 1.23% at 40 min (Suppl. Table S9). We found no clear alignment patterns with MS2 on the negative strand, suggesting that, unlike the full-length transcripts, these small transcripts did not produce a negative strand (Suppl. Figure S6 D-F, Suppl. Table S10). Overall, our results suggest that these small transcripts throughout the life cycle(s) were a combination of degraded RNA, small transcripts required for phage replication, and Nanopore DRS 3' bias across time, adding difficulties in the analysis of the transcriptional dynamics.

Error-prone replicase frequently generates hybrid RNAs

Previous studies have described the generation of by-products from the error-prone replicase [31]. While it was previously reported that MS2 phage was capable of forming hybrid RNAs, it was not found in an in vivo setting [32]. Nanopore DRS enables the characterization of recombination events and template switching in viral RNA [33]. We interrogated the landscape of hybrid RNA transcripts from DRS. We quantified the presence of hybrid reads by blasting these reads to MS2 genomes and retaining the reads mapping to at least two different MS2 locations in single reads (Fig. 4A). Reads of recombination (20–60 nt deletions and/or duplications) within the MS2 RNAs were detected (Suppl. Table S11). We found a 0.2% of duplicated or deletion hybrid reads

to total reads ratio on the positive strand at 6 h p.i. We found a few RNA joining events of the negative strand to negative strand and sporadic negative strand to positive strand, and very rare (+/-) foldback RNAs (also known as copyback or snapback RNAs - Suppl. Figure S7 A-B). We noticed internal tandem duplication events and deletion events between the negative strand and the negative strand (Suppl. Figure S7 A). While we discerned recombination events within phage RNA, we did not observe any recombination events between viral genome segments and host RNAs. We wondered whether new variants have emerged over time. We assessed the mutational landscape of MS2 phage (SNV and structural variation) across time. We did not observe the emergence of any new mutations up to 6 h p.i. Together, our results indicate that the error-prone replicase generates hybrid transcripts and foldback RNAs as by-products.

Static and dynamic RNA modification on MS2 phage during replication

RNA modification plays an important role in modulating RNA function [34]. Previous studies have identified the presence of ψ modification on MS2 phages using either mass spectrometry and/or by DRS [35]. We wondered whether MS2 acquired new modifications during a single life cycle (from 0 h to 40 min p.i.) and across time (3 and 6 h p.i.). We predicted and mapped m⁵C and ψ modifications throughout a single MS2 life cycle on the RNA positive strands (Fig. 5A-D). We found m⁵C, and ψ modifications were widely distributed (Fig. 5A-D). On full-length transcripts (> 3,000 nt), we found no evidence of newly acquired modifications within a single cycle (Fig. 5A and C). Instead, we identified commonly modified static sites with high stoichiometry throughout the transcripts, varying from 23.4% to 78.0% (Fig. 5A and C). Interestingly, we found shorter transcripts increase in m⁵C stoichiometry across time (Fig. 5B and Suppl. Figure S9B). Consistent to a previous study [35], we also identified a conserved ψ modification site at 924 nt (Fig. 5C). We found additional m⁵C (at location 634, 927, 2636, 3378, 3413, and 3521 nt) and ψ installation sites (at location 2145 and 2407 nt) for small transcripts with length < 800 nt at 3 h and 6 h 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 Ψ UACA RluA canonical recognition motif [36]. 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 Ψ UACA 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.

Discussion

RNA bacteriophages such as Q β and MS2 have been extensively studied for their ability to rapidly replicate [8, 37, 38]. 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 h 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. Our results suggest that, despite carrying a minimal genome, MS2 efficiently enhanced its transcriptional activity and harnessed its error-prone replicase to rapidly replicate and lyse the bacterial cell.

High transcriptional/replication level of MS2 within bacterial cells

Previous studies assessed RNA phages [38] and RNA viruses [39] replication patterns (including host factors). The archetypical MS2 and Q β replicase complexes have been extensively investigated since the 1960s, including the host factors responsible for forming the replicase

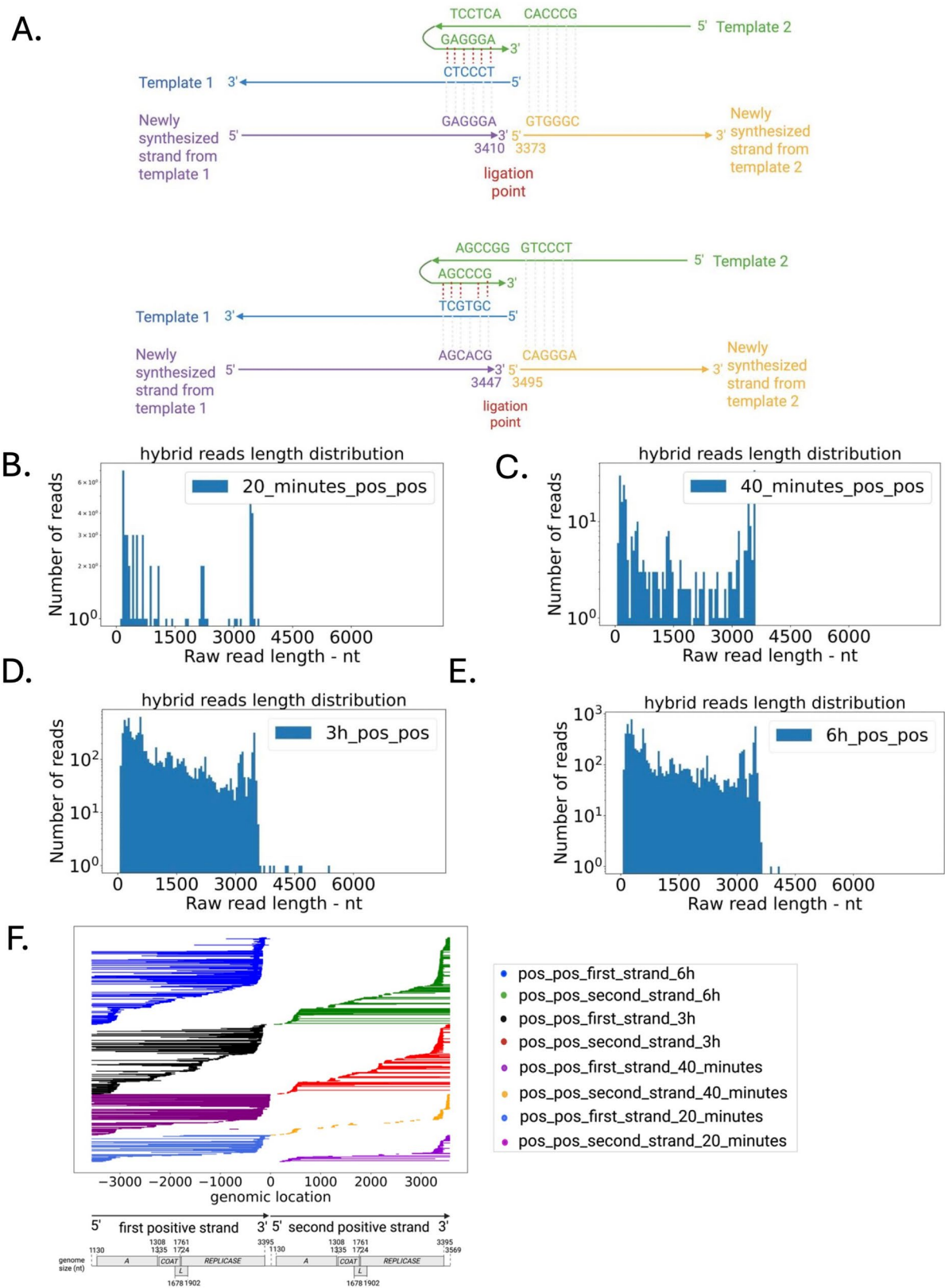


Fig. 4 (See legend on next page.)

(See figure on previous page.)

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

complex [8] and their underlying roles [38]. However, little is known about the dynamics of the transcriptional activity during the MS2 phage life cycles. A key reason is that most previous studies were performed in an in vitro or a transcription-translation system for MS2 [8] and Q β [38, 40]. Using Nanopore RNA DRS, we reported significant changes in MS2 transcript levels across multiple life cycles. Strikingly, we noticed an over 50-fold increase in full-length MS2 transcripts in only three hours post-inoculation and over 400-fold increase in coat protein expression, indicative of high activity of the phage for replication and assembly to maintain the infection. To a lesser extent, other RNA viruses, such as the Human Coronavirus 229E (HCoV-229E), produce their own transcripts (~ 33% reads mapped to the virus genome) at high levels [14] for viral replication. Another study also showed a large proportion of reads (~ 65% reads) mapped to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [13]. As such, our observation is in line with previous reports on other viruses, suggesting the ability of a virus to rapidly transcribe within hours post-inoculation. Together, our study reports that RNA phages such as MS2 show significant upregulation and fine-tuning of the expression of their ORFs across time.

Formation of subgenomic transcripts during the MS2 life cycle

Previous studies have reported that MS2 utilizes its negative strand for replication via the replicase complex [8, 41]. While the Q β replicase complex has been extensively studied for the formation of spontaneous short amplifying RNA and additional by-products, MS2 replicase has been recently described as less promiscuous towards the formation of spontaneous replicative products [8]. The generation of subgenomic RNA has been reported for various single-stranded positive viruses, and it is a common strategy for viruses to control their viral expression [42].

During our investigation, we identified a large population of short subgenomic reads appearing after the latent period of phage infection and increasing over time. We hypothesised that this population could be a Nanopore sequencing bias due to degraded RNA from bacterial lysis after MS2 infection, and 3' end enrichment is the Nanopore DRS 3' end bias [26]. These are inherent limitations of Nanopore RNA DRS and bacterial lysis, rendering the analysis more complex [43]. Using stringent

quality control measures such as high-quality score filtering, we nonetheless identified subgenomic viral RNA. Previous study using Nanopore DRS on single positive-strand RNA viruses such as the Yellow citrus vein virus (CYVCV) reported to form subgenomic RNA [44], has found similar findings [33] validating the use of the Nanopore DRS approach to characterise subgenomic viral RNA. The viral transcripts within bacterial cells are more vulnerable/sensitive to nuclease after lysis, and the capsid-protecting full-length positive strand viral RNAs are less likely to be degraded compared to the full-length negative strand viral RNA and any RNAs within the bacterial cells after lysis. We further hypothesised that these short transcripts could be single ORF transcripts. Interestingly, we identified hotspots of mapped transcripts corresponding to the full-length coat protein ORF. Specifically, the top RNA virion contact sites were previously identified at positions surrounding 1740–1750 nt and 2110–2120 nt [6] and overlaid with the top termination hits (Fig. 3D), suggesting the assembly and packaging process could potentially inhibit the replication process by prematurely ending the transcription at specific RNA contact sites (Fig. 3E–F). Additionally, we reported that this transcription is potentially regulated by the occupancy of the coat protein on the positive strand on specific RNA binding sites, potentially aborting the transcription/translation from the replicase complex [6]. Together, we suggest that MS2 phage produces a high level of coat protein transcripts to potentially repress the replicase complex, and is critical for virion assembly.

Error-prone replicase complex generates hybrid transcripts

An important requirement for some viruses to replicate and evolve is to harbour either proofreading [45] or non-proofreading replicase [46]. A mechanism for viruses to acquire novel traits and evolve is to form hybrids with other viruses or within the same virus [47]. We wondered whether the error-prone replicase would favour the generation of *de novo* mutations and such hybrid RNAs to form new virions. Using Nanopore DRS, we reported long intraspecies hybrid transcripts (> 3569 nt), including fold-back RNAs accounted for 0.2% of the total transcript count. We also found that these hybrid transcripts resulted from template switching recombination. The presence of hybrid RNA or recombination events has been reported for other eukaryotic viruses such as SARS-CoV-2 [48] or plant viruses [33]. Hybrid RNAs have also

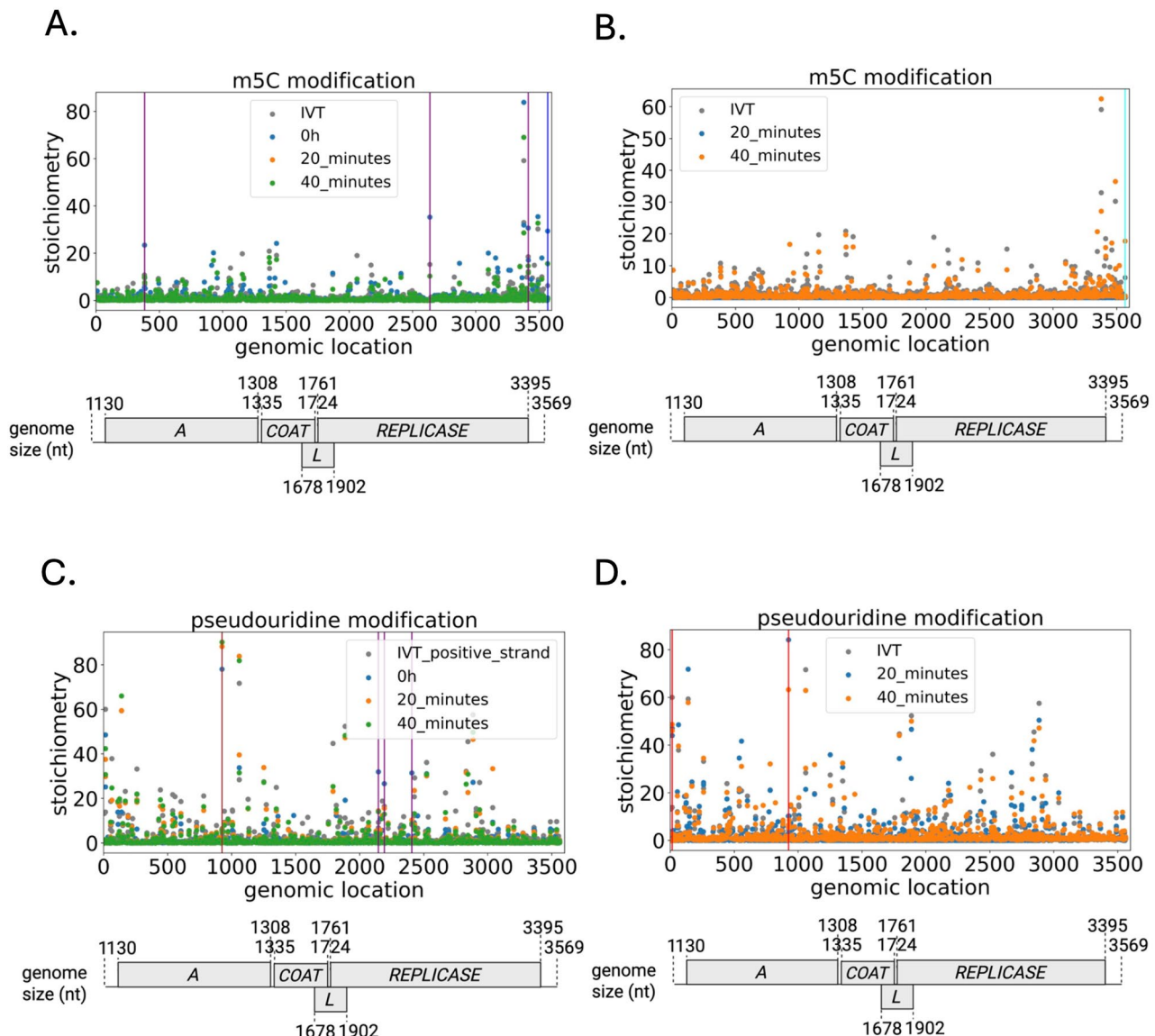


Fig. 5 Predicted sites of ψ and m^5C RNA modification dynamics on MS2 phage positive RNA (> 3000 nt length and < 800 nt length) within one life cycle (at 0 h, 20 min, and 40 min p.i.). Synthetic modification-free RNA as the negative control. **(A)** m^5C modification (> 3000 nt) landscapes according to the counts of modified and unmodified bases. **(B)** m^5C 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, 0 h vs. 20 min decrease, 20 min vs. 40 min increase (vertical blue line); 0 h vs. 20 min decrease, 20 min vs. 40 min no difference (vertical purple line); 0 h vs. 20 min increase, 20 min vs. 40 min no difference (vertical brown line); For the positive strand with read length < 800 nt, 20 min vs. 40 min increase (vertical cyan line), 20 min vs. 40 min 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

been reported for MS2 in plasmid systems [32]. Consistent with previous observations, our findings indicate the ability of MS2 to recombine using template switching mechanisms into new RNA transcripts.

U-rich hairpin of Ψ installation sites on native MS2 RNA

Previous studies have evidenced the crucial role of RNA modification in viral replication [49]. To achieve high translation efficiency and evade host immunity, some

viruses encode their own methyltransferase or other modifying enzymes [50]. However, the MS2 genome doesn't harbor any modifying enzymes and requires hijacking the host to modify its own RNA. Using Nanopore DRS, we aimed to assess whether the modification landscape changes within a single life cycle and across time. We evidenced the presence of the Ψ installation site and secondary structure (U-rich hairpin) amenable to Ψ modification with a variation of stoichiometry within a

single cycle and across time. Notably, we found evidence of canonical and non-canonical Ψ UACA RluA-mediated installation, providing evidence that host pseudouridine synthase modifies MS2 RNA during its life cycle. Interestingly, we found a dynamic of loss and gain in methylation of the small transcripts during the MS2 across time. Our results suggest that, unlike the full transcripts, small transcripts are more 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 [35]. These modifications are either transient, predominantly on short transcripts, or permanent with various stoichiometries on canonical Ψ UACA RluA-mediated installation sites. However, the functionality of these modifications on MS2 ORF translation remains unknown and will require further investigations.

Limitations and further investigations

Using an in-vivo setting in bacteria, degradation is unavoidable after cell lysis in the bacteria after the first cycle burst. To potentially overcome this issue, an in vitro cell-free synthesis system using bacterial cell lysate [8], supplemented with an RNase inhibitor, might be a potential avenue for further study of the MS2 phage replication/transcription process and the formation of subgenomic RNA across replication. Furthermore, additional time points in the first life-cycle could be useful to better capture the formation of subgenomic RNA and the dynamics in the control of viral transcription, additional time points in the first life-cycle. RNA degradation due to unavoidable bacterial cell lysis, and strand biases, inherent of the Nanopore DRS technology, impacted the interpretation of the read length/coverage results. Additionally, it has added a layer of complexity in the analysis of the Nanopore DRS data to distinguish the degraded RNAs and the real transcripts with high integrity.

Conclusions

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. Our results showed the numerous generations of byproducts during transcription, as either products of the error-prone replicase (hybrid reads) or coat protein transcripts required for assembly, attesting the ability of the phage to rapidly adapt to its host. Lastly, our results demonstrate 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 RluA-mediated.

These analyses of the transcriptional dynamics of MS2 bacteriophages were however complicated by the combination of factors (transcript degradation, transcripts generated due to error-prone replicase, and Nanopore DRS bias). As such, it has added another layer of complexity to untangle the inherent bias in the sequencing technology versus the byproducts of viral replication. Nonetheless, together, the direct RNA sequencing approach has revealed new insights into the complex transcriptional dynamics of MS2 bacteriophage.

Abbreviations

ANOVA	Analysis of Variance
CP	Coat protein
CTD	C-terminal domain
DE	Differential expression
DRS	Direct RNA sequencing
EF-Ts	Elongation factor thermos stable
EF-Tu	Elongation factor thermo unstable
FDR	False Discovery Rate
FPR	False-positive rates
gRNA	Genomic RNA
HCoV-229E	Human Coronavirus 229E
IVT	In vitro transcription
m ⁵ C	5-methylcytidine
MOI	Multiplicity of infection
MP	Maturation protein
NTD	N-terminal domain
ONT	Oxford Nanopore Technologies
PCA	Principal Component Analysis
p.i.	Post-infection
RBS	Ribosomal binding sites
Rdrp	RNA-dependent RNA polymerase
RIN	RNA integrity number
RT-qPCR	Reverse transcription quantitative PCR
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SD	Standard deviation
SRA	Sequence Read Archive
ssRNA	Single-stranded RNA

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12985-026-03121-5>.

Supplementary Material 1.

Supplementary Material 2

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.

Author contributions

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.

Funding

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.

Data 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](https://github.com/gburgio/MS2_Nanopore_DRS)(https://github.com/gburgio/MS2_Nanopore_DRS%20).

Declarations

Ethics approval and consent to participate

This study involved any human or animal subjects. All protocols were approved from the ANU institutional biosafety committee approval N°20.05.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 22 September 2025 / Accepted: 23 February 2026

Published online: 06 March 2026

References

- Clokier MR, et al. Phages Nat Bacteriophage. 2011;1(1):31–45.
- Zhang W, Wu Q. Applications of phage-derived RNA-based technologies in synthetic biology. Synth Syst Biotechnol. 2020;5(4):343–60.
- Bollback JP, Huelsenbeck JP. Phylogeny, genome evolution, and host specificity of single-stranded RNA bacteriophage (family Leviviridae). J Mol Evol. 2001;52(2):117–28.
- Wolf YI, et al. Origins and evolution of the global RNA virome. MBio. 2018;9(6):02329–18. <https://doi.org/10.1128/mbio.2018.02329>
- Poot RA, et al. RNA folding kinetics regulates translation of phage MS2 maturation gene. Proc Natl Acad Sci. 1997;94(19):10110–5.
- Rolfsson Ó, et al. Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. J Mol Biol. 2016;428(2):431–48.
- Salas M, de Vega M. Replication of bacterial viruses. Desk Encyclopedia of Animal and Bacterial Virology. 2008. p. 553.
- Wagner A, Weise LI, Mutschler H. Vitro characterisation of the MS2 RNA polymerase complex reveals host factors that modulate emesviral replicase activity. Commun Biol. 2022;5(1):264.
- Meng R, et al. Structural basis for the adsorption of a single-stranded RNA bacteriophage. Nat Commun. 2019;10(1):3130.
- Chamakura KR, Tran JS, Young R. MS2 lysis of *Escherichia coli* depends on host chaperone DnaJ. J Bacteriol. 2017;199(12):101128.
- Dykeman EC. Design of a self-regulating mRNA gene circuit. Sci Rep. 2024;14(1):19421.
- Peabody DS. Translational repression by bacteriophage MS2 coat protein expressed from a plasmid. A system for genetic analysis of a protein-RNA interaction. J Biol Chem. 1990;265(10):5684–9.
- Kim D, et al. The architecture of SARS-CoV-2 transcriptome. Cell. 2020;181(4):914–21. e10.
- Viehweg A, et al. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. Genome Res. 2019;29(9):1545–54.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.
- Li H, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.
- Jain M, et al. Improved data analysis for the MinION nanopore sequencer. Nat Methods. 2015;12(4):351–6.
- Workman RE, et al. Nanopore native RNA sequencing of a human poly (A) transcriptome. Nat Methods. 2019;16(12):1297–305.
- Patro R, et al. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417–9.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
- Zheng Z, et al. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. Nat Comput Sci. 2022;2(12):797–803.
- Smolka M, et al. Detection of mosaic and population-level structural variants with Sniffles2. Nat Biotechnol. 2024;42(10):1571–80.
- Danecek P, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc: Ser B (Methodol). 1995;57(1):289–300.
- Jain M, et al. Advances in nanopore direct RNA sequencing. Nat Methods. 2022;19(10):1160–4.
- Chetverin AB, Chetverina HV, Munishkin AV. On the nature of spontaneous RNA synthesis by Q β replicase. J Mol Biol. 1991;222(1):3–9.
- Ng W. Database of ribosome binding site of all genes in *Escherichia coli* K-12 MG1655. figshare. Dataset. 2019. <https://doi.org/10.6084/m9.figshare.10282685.v1> Accessed 1 Jun 2025.
- Prawer YD, et al. Pervasive effects of RNA degradation on Nanopore direct RNA sequencing. NAR Genomics Bioinf. 2023;5(2):lqad060.
- Valegård K, et al. Crystal structure of an RNA bacteriophage coat protein-operator complex. Nature. 1994;371(6498):623–6.
- Kopsidas G, et al. RNA mutagenesis yields highly diverse mRNA libraries for in vitro protein evolution. BMC Biotechnol. 2007;7(1):18.
- van Meerten D, et al. In vivo generation of hybrids between different species of RNA phages. J Gen Virol. 2002;83(5):1223–35.
- Johnson PZ, et al. 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 Bioinf. 2024;6(3):lqae104.
- Pozhydaeva N, et al. The enigmatic epitranscriptome of bacteriophages: putative RNA modifications in viral infections. Curr Opin Microbiol. 2024;77:102417.
- Jones J. *Unveiling the RNA Modification Landscape Through Bioanalytical and Biochemical Analyses*. 2023.
- Schaeninger-Burgos C, et al. RluA is the major mRNA pseudouridine synthase in *Escherichia coli*. PLoS Genet. 2024;20(9):e1011100.
- Chang J-Y, et al. Structural assembly of Q β virion and its diverse forms of virus-like particles. Viruses. 2022;14(2):225.
- Vasilyev NN, et al. Ribosomal protein S1 functions as a termination factor in RNA synthesis by Q β phage replicase. Nat Commun. 2013;4(1):1781.
- Hillen HS, et al. Structure of replicating SARS-CoV-2 polymerase. Nature. 2020;584(7819):154–6.
- Tomita K, Ichihashi N, Yomo T. Replication of partial double-stranded RNAs by Q β replicase. Biochem Biophys Res Commun. 2015;467(2):293–6.
- Dykeman EC. Modelling ribosome kinetics and translational control on dynamic mRNA. PLoS Comput Biol. 2023;19(1):e1010870.
- Koev G, Miller WA. A positive-strand RNA virus with three very different subgenomic RNA promoters. J Virol. 2000;74(13):5988–96.
- Liu-Wei W, et al. Sequencing accuracy and systematic errors of nanopore direct RNA sequencing. BMC Genomics. 2024;25(1):528.
- Renovell A, et al. Mapping the subgenomic RNA promoter of the Citrus leaf blotch virus coat protein gene by Agrobacterium-mediated inoculation. Virology. 2010;406(2):360–9.
- Robson F, et al. Coronavirus RNA proofreading: molecular basis and therapeutic targeting. Mol Cell. 2020;79(5):710–27.
- Whitfield ZJ, et al. Species-specific evolution of Ebola virus during replication in human and bat cells. Cell Rep. 2020;32(7):108028. <https://doi.org/10.1016/j.celrep.2020.108028>.
- Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? Nat Rev Microbiol. 2011;9(8):617–26.
- Garushyants SK, Rogozin IB, Koonin EV. Template switching and duplications in SARS-CoV-2 genomes give rise to insertion variants that merit monitoring. Commun biology. 2021;4(1):1343.

49. Bhattacharya T, et al. Differential viral RNA methylation contributes to pathogen blocking in *Wolbachia*-colonized arthropods. *PLoS Pathog.* 2022;18(3):e1010393.
50. Li N, Rana TM. Regulation of antiviral innate immunity by chemical modification of viral RNA. Volume 13. *Wiley Interdisciplinary Reviews: RNA*; 2022. p. e1720. 6.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.