Nanopore direct RNA sequencing (DRS) of MS2 bacteriophages in *E.coli* throughout its life cycle 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 a positive single stranded RNA phage that infects the bacterium *E. coli* and is one of the most studied and representative model organisms in molecular biology and microbiology. Previous research revealed translational control and feedback inhibitions in the MS2 life cycle. However, the dynamics of its transcriptional activity and replication during the life cycle within the bacteria is still relatively unknown. Here, we employed Nanopore Direct RNA sequencing (DRS) to investigate the transcriptome and epitranscriptome landscape of the MS2 in infected *E. coli* throughout its life cycle and detailed the dynamics of MS2 transcriptional activity in real-time. We discovered that MS2 phages sustain a high level of transcriptional activity within the bacterial cell. We found large amounts of subgenomic small RNA transcripts containing the *coat*-encoding region. We found hybrid reads and mutations due to the error-prone activity of the MS2 replicase complex. 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 and remove modified sites. The conserved sequence (5'-ΨU-3') and secondary structure (stem-loops) of Ψ installation sites were the most amenable to RNA modification. Our investigation reveals a more complex transcriptional dynamics of MS2 phages within *E.coli* to maintain its growth and replicate under host pressure.

Introduction:

There are an estimated 10^{31} bacteriophage particles in the biosphere, and they are found ubiquitously in nature (Comeau et al., 2008). Bacteriophages are submicroscopic and obligate intracellular organisms that play important roles in the evolution of bacterial host cells. To guide the replication of their genome and generate the necessary viral-encoded proteins, phages hijack the host's metabolic machinery. RNA bacteriophages (commonly known as RNA phages), such as MS2, R17, PP7, Q-beta, etc., are viruses that infect a variety of bacteria (Bollback and Huelsenbeck, 2001) and have exceptionally high mutation rates. It is thought that the ssRNA phages are the oldest lineage of RNA viruses still in existence. RNA phages 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).

Bacteriophage MS2 belongs to the family *Leviviridae* (*Levivirus*), and is an icosahedral, non-enveloped, F-specific, positive-sense single-stranded RNA (+ssRNA) lytic coliphage that infects the bacterium *E. coli* (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 for RNA replication (**Figure 1A**) (Wagner, Weise and Mutschler, 2022). Ribosomal protein S1 might play a role in replication initiation and termination (Vasilyev et al., 2013), and EF-Tu and EF-Ts play roles in RNA elongation. Full-length negative RNA intermediates are generated to serve

as the template for the synthesis of positive genomic RNA (gRNA) by replicase complex (**Figure 1A**). MS2 RNA secondary structure and AG-rich region on the negative strand might regulate the replicase's function in initiating replication of the positive strand (**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 that replicates the ssRNA viral genome (Wagner, Weise, and Mutschler, 2022). **(Figure 1B)**. 89 CP dimers and one copy of a genome-bound MP, in replacement of another CP dimer, make up the entire virion.

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). The packaging signal (19nt-TR) and other RNA sites (putative stem-loops), known as the assembly initiation sites of MS2, help in mediating the assembly of MS2 by RNA-coat protein interactions (Rolfsson et al., 2016). Despite years of research on MS2 replication and assembly, little is known about the dynamics of transcription of MS2 during its life cycle, leading to rapid phage replication. Like many other viruses, MS2 is a polycistronic unit encoding an error-prone replicase, which is critical for replication. The replicase complex is repressed from the CP, which binds to specific sites on the positive strand RNA. While the interplay between different phage proteins resulting in entry, replication, assembly, and encapsidation has been relatively well studied in *in-vitro* settings, the dynamics in transcription throughout the MS2 life cycle remain unexplored. Specifically, the unanswered question

remains 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 dynamics of transcription and measure each transcript, 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) using Nanopore technologies.

Here, we employed Nanopore DRS to investigate the transcriptome and epitranscriptome landscape of the MS2 in infected *E. coli* throughout its life cycle and detailed the dynamics of MS2 replication in real-time. We discovered that MS2 phages sustain a high level of transcriptional activity within the bacterial cell. We found large amounts of small transcripts corresponding to the full-length coat protein ORF. The MS2 replicase is error-prone, which frequently generates hybrid RNAs and leads to the rapid emergence of *de novo* synonymous mutations during the MS2 life cycle. 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 and remove modified sites. The conserved sequence (5'-ΨU-3') and secondary structure (stem-loops) of Ψ installation sites were the most amenable to RNA modification. Together, using DRS RNA Nanopore sequencing, we have detailed the dynamics of MS2 transcriptional activity in bacterial cells during its life cycle.

Methods:

Bacteria, phage strains, and culture conditions

The bacteria host *E. coli* DSM 5695 (also called *E. coli* W1485) (F^+) and MS2 bacteriophages DSM 13767 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, 5.0g/L Yeast extract, 2.0 g/L MgSO4 X 7H2O) with maltose (2g/L) and incubated at 37ºC overnight. MS2 phages were maintained and propagated in SM buffer (NaCl 100 mM, MgSO4 X 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 µl overnight culture of DSM 5695 bacteria and 100 µl MS2 in SM buffer were mixed with 5 ml NZCYM soft agar with maltose (2g/L). Plates were incubated at 37ºC overnight (~16 h). MS2 phages were harvested. The top layer was disrupted after pouring 5 ml SM buffer, and the lysate was centrifuged at 2,800g for 30 minutes at room temperature. The lysate was filtered using a 3 ml syringe with a 0.22 µM filter. The precise viral stock titer (PFU/ml) of the phage (MS2) was determined.

MS2 phage infection and RNA isolation

The bacteria *E. coli* (DSM 5695) liquid culture growth infected with bacteriophage MS2 was generated 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 (~10^10

bacterial cells were prepared as the initial material) at 0h, 3h and 6h post-infection (p.i.). Total RNAs (at 0h, 3h, and 6h p.i.) were extracted from infected *E. coli* samples (TRIzol [Invitrogen]) following the manufacturer's instructions.

In vitro transcription of MS2 RNA

One ug of total RNA purified from MS2-infected *E. coli* cells. Reverse transcription (SuperScript™ IV Reverse Transcriptase [Invitrogen]) was performed with each phage-specific RT primer (RTprimer1-6) (Supplementary table S1). Templates for *in vitro* transcription 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 (MEGAshortscript™ T7 Transcription Kit [Thermo Fisher]) and RNA purification (Monarch RNA Cleanup Kit [NEW ENGLAND Biolabs]) with resuspension in ddH2O. The oligonucleotides used in this study are listed in Table S1.

Poly(A) tailing of RNA and beads purification

Total RNAs and IVT RNAs were heat 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 µg RNA, 5 units poly(A) polymerase, 2 µL 10X reaction buffer, and 1 mM ATP in a total reaction volume of 20 µL 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 µL 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 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 A**). Basecalling was performed using MinKNOW software (Nanopore) using the Super accurate model. The Fast5 files were filtered. Reads with a minimum Q score of 10 were removed. 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 (Li, 2018) using the following parameters: -ax map-ont -y --secondary=no -k 4 -w 10. Samtools (Li et al., 2009) 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 (Quinlan and Hall, 2010).

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 in 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. Read coverage within 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. Chimeric reads were mapped to NC_00141702.2 genome using blastn. The top hits e.values > $10e^8$ 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 (Zheng et al., 2022) and Sniffles2 (Smolka et al., 2024). 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 filteres and converted using Bcftools (Danecek et al., 2021) in vcf format.

RNA modification analysis:

Basecalled reads from Dorado were mapped to NC_00141702.2 genome using Minimap2, sorted and indexed using Samtools. Basecalling for modification was determined using Modkit pileup function using –max-depth 2000000 option and a modification threshold > 80%. Differential methylation profile was determined using modkit dmr function.

Statistical analysis

Statistical analysis was conducted. The mean ± 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 (Benjamini and Hochberg, 1995).

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 6 hours in *E.coli*

To determine the course of infection between phage replication and lysis, we infected the bacterial cells with MS2 phage and monitored *E. coli* growth and lysis over its life cycle in liquid culture at the multiplicity of infection (m.o.i) of 200 (**Figure 1C**). MS2 phages actively replicated (3 h post-infection), assembled in *E. coli*, and mature progenies burst out of the cells and lysed the culture (6h post-infection), enabling new cycles of infection (10h post-infection) (**Figure 1C**). MS2 formed clear plaques, 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**). Over the MS2 life cycle, the size and the number of plaques increased gradually (**Figure 1 D-E**). Overall, the data indicated the time points 3h and 6 h captured, respectively, phage growth and lysis, and were suitable time points to investigate the dynamics of phage replication in real-time.

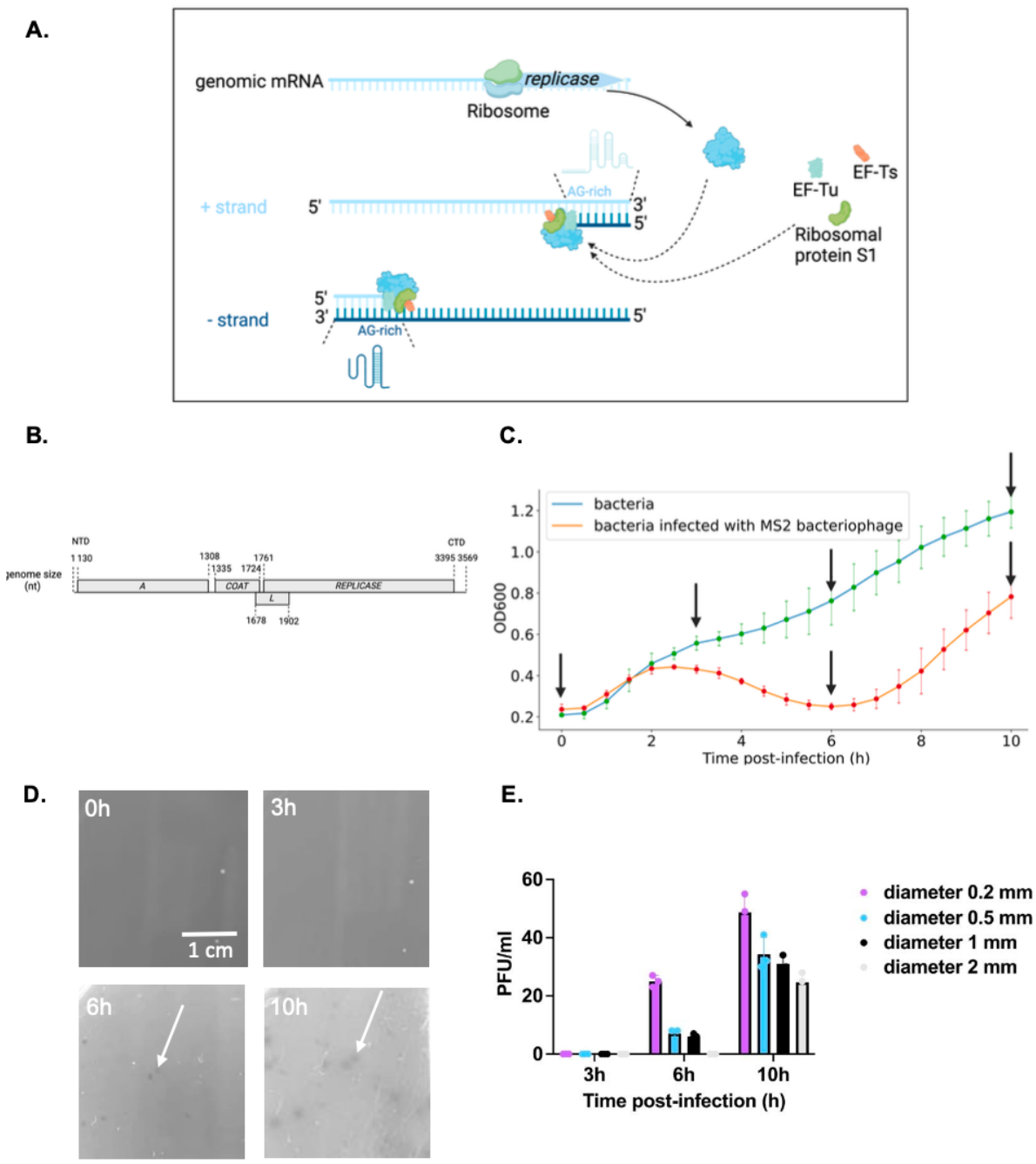

Figure 1. Characteristics of MS2 phage A. MS2 replication pattern. The MS2 replicase complex contains host factors EF-Tu, EF-Ts and S1 ribosomal protein. **B.** Schematic presentation of the MS2 full-length genomic RNA coding regions organisation and composition. NTD, N-terminal domain, CTD, C-terminal domain. The 5′ UTR of MS2 contains ~130 nt, and the 3′ UTR of MS2 contains 171 nt. **C.** Growth curves of bacterial cells (DSM5695 *E. coli*) infected with MS2 RNA phage at MOI = 200 (red line) and bacteria cells without infection (blue). Phages were added at t = 0. Cells were harvested at 0h, 3h, and 6h post-infection (p.i.), and RNAs were extracted. Each growth curve represents the mean values and standard deviations obtained from biological triplicates. **D.** MS2 plaques on the host *E. coli* DSM5695 at 0h, 3h, 6h, 10h p.i. Plaques were pointed with white arrows. **E.** Plaque PFU and size change at 0h, 3h, 6h, 10h p.i. All pair-wise comparisons (3h vs. 6h vs. 10h at each diameter) were done after observing a significant Analysis of Variance (ANOVA) test result and showed statistically significant results (*P* value <0.05 after accounting for multiple comparisons),

except the 3h vs. 6h (1 mm diameter) and the 3h vs. 6h (2 mm diameter) one. Data are presented as means ± standard deviation (n = 3).

Table 1. Read statistics of ONT DRS reads aligned to the MS2 genome at different time points 0h, 3h, 6h post-infection (p.i.). The table statistics are based on the average of triplicates at each time point p.i.

parameters	MS2_0h	MS2_3h	MS2_6h	MS2_IVT
number of alignments	5,068.33	1,097,886.33	1,002,011.33	2906655.00
median length	2,209.17	715.67	589.00	554.00
mean length	2,119.33	1,101.67	1,243.00	669.00
N50 length	3,518.33	1,988.00	3,070.33	1134.00
Full-length reads	1,979.33	98,028.33	137,317.67	-
Full-length read ratio (%)	39.05	8.93	13.70	-

IVT, *in vitro* transcribed.
Full-length reads were defined as those covering 90% of the reference genome.

MS2 phage maintained a high transcriptional activity for replication.

To capture the dynamics of MS2 replication in real-time, we performed long-read Nanopore direct RNA sequencing (DRS) runs on total RNA extracted from DSM5695 *E. coli* cells infected with MS2 at 0h, 3h and 6 h post-infection (p.i.). There was an average read length spanning from 669 to 2119 nt mapped reads for infected cells (**Table 1)**. Reads over 90% of the MS2 transcript were qualified as full-length reads (**Table 1**). Immediately after infection (0h time point), the MS2 virus just disassembled and injected its genomic RNA (gRNA) into

the cytoplasm of the bacteria cells prior to replication; only a few full-length RNAs were detected (only accounted for 0.09% of the total reads) and almost no observable negative strand transcripts (**Figure 2A**). Compared with 0h, an increased ratio of viral RNAs to bacterial RNAs at 3 h (~500x fold change) and 6 h (~200x fold change) suggests the active replication at 3 h and replication and lysis at 6h (**Figure 2A**). Unmapped reads were identified with blast and mapped to *E.coli*. Gene expression analysis was conducted to assess variation between different time points across the open reading frames (ORFs) during replication. A Principal Component Analysis (PCA) on differential gene expression was conducted across the different time points of the MS2 life cycle. Two main groups were visible: the uninfected on the PC1 negative axis and the infected group on the positive PC1 axis. Among the infected group, the 3h time point group was on the PC2 positive axis **(Figure 2B)**. The different time points exhibited distinct clusters on the correlation heatmap (**Suppl. Figure S2A**), suggesting profound transcription changes across time. Next, we analyzed the differential expression (DE) of the four MS2 transcripts to determine the change in expression over time (**Figure 2C**). We found an upregulation of the *mat* and *cap* ORFs while the *rep* and the *lys* ORFs were downregulated (**Figure 2C**), confirming an active transcription state of the virus at 3- and 6-hours post-inoculation inside the bacterial cell. To better capture the dynamics of MS2 replication in real-time, we separated the positive and negative strands mapped reads. Next, we assessed the read coverage across time for all reads (**Figure 2D**), positive and negative strands (**Suppl. Figure S3. A-B**). The negative strand read count accounted for less than 5% of the positive-strand reads **(Suppl. Figure S3. A-B)**. We observed a coverage bias towards the *coat* ORF and the 3' end of the viral genome on the overall transcript and positive strand (**Figure 2D, Suppl Figure S3 A**), whereas the 5' non-coding region and the 3' end of MS2 transcript was enriched in the negative strand at 6 h post-inoculation (**Suppl Figure S3 B**). As the nanopore DRS has the 3' bias due to directional sequencing from the 3' ends of RNAs (Jain et

al., 2022), we deduced that coverage was biased only towards the *coat* ORF and 5' end in MS2 transcriptome. It suggests an enrichment of the attachment and synthesis of the positive strand from the complex replicase•EF-Tu•EF-Ts•ribosomal S1 protein on the negative strand at 6 h post-inoculation (Wagner, Weise, and Mutschler, 2022). While we were quantifying read counts, we noticed a 60-fold increase of the full-length transcripts (> 3000 nt) (**Figure 2 E-G**) across time, suggesting the rapid MS2 replication leading to the formation of matured new progenies. Interestingly, we noticed the dominance of a large population of small transcripts (< 800 nt) across the life cycle of the phage (**Figure 2 F-G**) on the positive and negative strands (**Suppl. Figure S2 B-F)** ranging from 53% to 93 % of the read count (**Suppl. Table S2**). Overall, the data suggest during the viral life cycle, MS2 phage enhanced the expression of key proteins involved in phage adhesion and maturation. Intriguingly, a high frequency of small transcripts < 800 nt were observed during replication with unknown function.

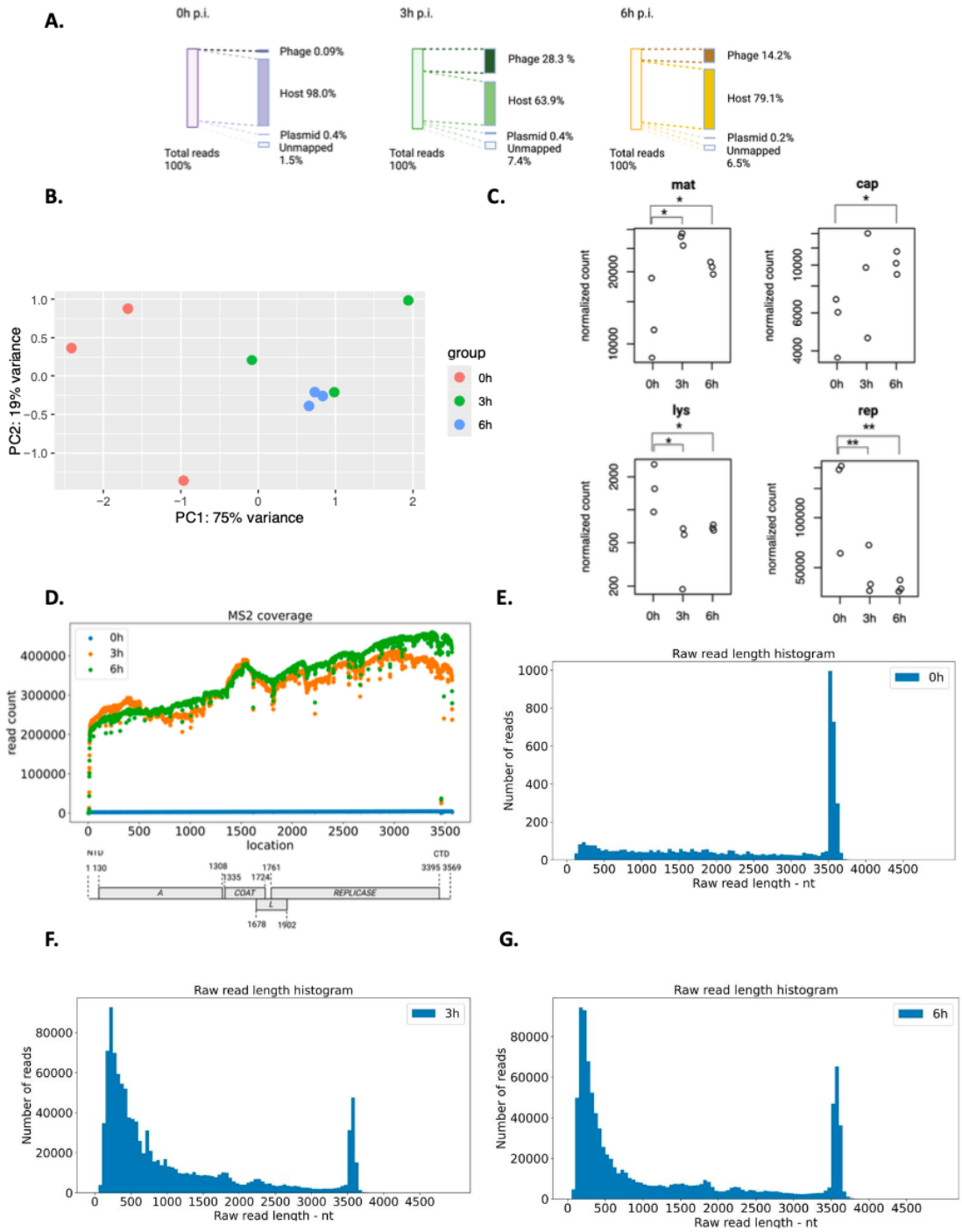

Figure 2. Statistics of Oxford Nanopore Technologies (ONT) direct RNA sequencing data via computational approaches and transcription pattern of MS2. A. Read counts from nanopore direct RNA sequencing of total RNA from DSM5695 *E. coli* cells infected with MS2 phage at different time points (0h, 3h, 6h p.i.). **B.** Principal component analysis (PCA) plot of MS2 samples at 0h, 3h, 6h p.i., each with three biological replicates to visualise the batch effects. **C.** Differential expression patterns for four MS2 genes at different time points (0h, 3h and 6h p.i.), analysed by DESeq2 with 'median of ratios' to normalise the read count. Adjusted *P*-value (after accounting for multiple hypothesis testing using the Benjamini-Hochberg method) <0.05, *; <0.01, **. **D.** Coverage map for MS2 total reads at 0h, 3h, 6h p.i. **E-G.** MS2 raw read length distribution at 0h, 3h and 6h p.i.

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

Previous works described the ability of the replicase to generate small spontaneous template-directed RNA species known as "RNA parasitic" species (Chetverin, Chetverina, and Munishkin, 1991; Wagner, Weise, and Mutschler, 2022). We hypothesized these 800 nt transcripts could be indeed either RNA parasitic species, RNA species resulting from defective replications, and/or small transcripts required for replication. To further investigate the nature of these < 800 nt transcripts, we quantified their coverage across the MS2 genome (**Figure 3 A-B**). We found an enrichment of the 5' end and the *coat* ORF (~19% of total positive-strand reads with read length <800 nt at 3h) on the positive and negative strands (**Figure 3 A-B**). We also found enrichment of the 3' end of the full transcript, likely due to the DRS bias. Interestingly, transcripts mapped to the *lys* ORF (7% of total negative strand reads with read length < 800 nt) and 3' end of the *rep* ORFs were enriched on the negative strand (**Figure 3B**), whereas part and the 3' end of the *mat* ORF lacked coverage on the positive strand (**Figure 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 small and large peaks of relative coverage on the positive and negative strands (**Figure 3 C-D**). We observed hotspots of transcripts start at ~1000-1300 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 at 3h and 6h post-inoculation (**Figure 3 C-D**). We performed a similar analysis on the full-length and < 800 nt transcripts (**Suppl. Figure S3 C-J**). We found similar peaks to those previously described in **Figure 3 C-D**, explained by the high proportion of these reads from the total read count (start at location 1010-1015 nt: 23 (relative coverage) of total positive strand reads with read length <800 nt at 3h, end of the transcript at location 1740-1745: 8 (relative coverage) of total positive strand reads with read length <800 nt at 3h, end of the transcript at location 2125-2130: 18 (relative coverage) of total

positive strand reads with read length <800 nt at 3h). Previous works have demonstrated that the positive strand generates an intermediate negative strand, which is used as the template for replication, and further, the holomeric replicase complex binds to the negative strand to produce the positive strand (Wagner, Weise & Mutschler, 2022). To avoid the formation of a double-stranded RNA, which would impair phage replication, the ribosome binds to translate positive strands to avoid annealing of the positive and the negative strands (Wagner, Weise & Mutschler, 2022). We 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 showed the MEME motif (**Suppl. Figure S4 A-E**) of these regions superimposed to known ribosomal binding sites (RBS), i.e., share the characteristics of putative RBS (Ng, Wenfa (2019) (**Suppl. Table S3 & Suppl. File 1**). Together, they suggest the presence of cryptic RBS throughout the MS2 genome, initiating the transcription of small-length transcripts (< 800 nt). A remaining question is how these transcripts terminate.

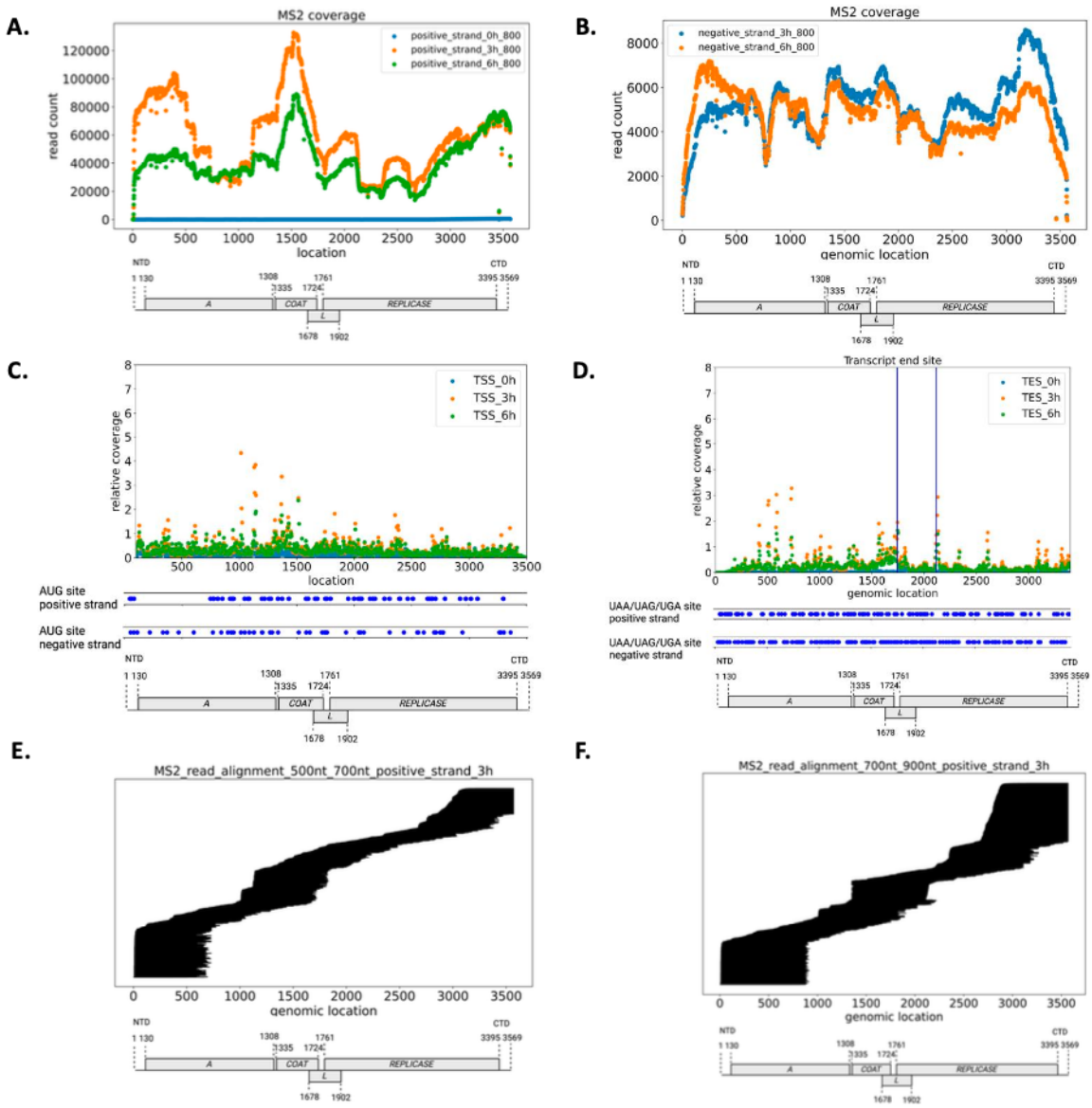

Figure 3. Replication pattern of MS2. A-B. Coverage map for MS2 reads (Reads aligned to the positive strand or the negative strand) at 0h, 3h, 6h p.i. with length of transcript <800 nt. **C.** Relative transcript start site (TSS) (ranging 100-3500 nt) coverage profile and **D.** Relative transcript end site (TES) (ranging 0-3400 nt) coverage profile for MS2 at 0h, 3h, and 6h p.i. blue vertical line indicated two previously identified key coat protein binding sites (Rolfsson et al., 2016). **E**. Positive-strand reads (shown as lines) aligned to MS2 genome with read length between 500 nt and 700 nt, and **F.** Read length between 700 nt and 900 nt. A schematic diagram of the MS2 genome is shown below.

A population of small Coat ORF transcripts and by-products of the replication are produced during the MS2 life cycle.

Unknowns are the factors on the phage leading to the transcript termination of these small transcripts. We suspected that the transcript end site (TES) is related to the stop codon (UAA, UAG, and UGA sites), the secondary structure of MS2 RNA, or an MS2 protein binding blocking the polymerase complex (**Figure 3D**). We investigated the presence of stop codons under these peaks and the changes in secondary structures from the template. We found no evidences that either the presence of stop codons or changes in secondary structures across replication were causative for these transcripts to end (**Figure 3D, Suppl. Figure S4 F-I**). Previous studies identified the coat protein subunit binds alongside the RNA template to mediate and regulate the viral assembly and packaging during replication (Rolfsson et al., 2016). The RNA-*coat* and –*mat* protein contacts were previously identified by pull-down and Clip-seq (Rolfsson et al., 2016). Specifically, the top RNA virion contact sites were previously identified at positions surrounding 1740-1750 nt and 2110-2120 nt (Rolfsson et al., 2016) and overlayed with the top termination hits **(Figure 3D)** suggesting that the assembly and packaging process could potentially inhibit the replication process by prematurely ending the transcription at specific RNA contact sites (**Figure 3 E-F**). We next sought to map these transcripts to MS2 to determine what these byproducts are. We separated the transcripts by size and aligned these transcripts to the MS2 genome. We found transcripts aligned across the entire positive and negative strand, suggesting the presence of "parasitic" or transcription-defective RNA species. Interestingly on the positive strand, read length between 500 nt and 700 nt containing a *coat* ORF (~5% of total transcripts ranging 500-700 nt) and read length between 700 nt and 900 nt containing a *lysis*-encoding region (~15% of total transcripts ranging 700-900 nt) (**Figure 3 E-F**), and the start site and end site of these small transcripts verify the spikes observed in **Figure 3 C-D**. Interestingly, 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. Overall, it suggests that during replication, the production of a population of small transcripts comprising coat protein transcripts, is required for assembly and packaging and RNA by-product from the replicase during replication.

Error-prone replicase frequently generates hybrid RNAs and synonymous mutations

Previous works have described the generation of by-products from the error-prone replicase (Kopsidas et al., 2007). The formation of interspecies and/or intraspecies hybrids RNA is a mechanism for the virus to adapt and evolve (Wells et al., 2023). We interrogated the landscape of hybrid RNA transcripts from DRS in the form of duplication or recombination. We quantified the presence of duplicated and hybrid reads by blasting these reads to MS2 genomes and retaining the reads mapping to at least two different MS2 locations from a single transcript. Reads of recombination (20-60 nt deletions and 20-60 nt duplications) within the MS2 RNAs were detected. We found high rates of duplicated reads on the positive strand (**Figure 4 A-C**). 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) were detected (only two pairs of reads; the schematic diagram and secondary structure was shown) (**Suppl. Figure S6 A-H**). We found internal tandem duplication events and deletion events between the negative strand and the negative strand (**Suppl. Figure S6 G)**. We only observed recombination events within phage RNA and did not observe any recombination events between viral genome segments and host RNAs. Together, it indicates the error-prone replicase generates not only small RNA but also a series of hybrid transcripts and foldback RNAs as by-products. The emergence of *de novo* mutation is an additional mechanism of virus adaptation and evolution in the host (Betancourt, 2009). We monitored for mutated nucleic acid sequences, and we found the emergence of synonymous

single nucleotide variants, but no structural variants across the life cycle in the replicase at position 2000 nt (U→C) and 2408 nt (C→U) and in the 3' UTR a 3455 nt (A→G), 3537 nt (U→C) without affecting the RNA secondary structure and at position 3451 (G→C) affecting the secondary structure (**Suppl. Table S4, Suppl Figure S6 I**). Together, it suggests the error-prone replicase generates various RNA species, hybrid RNA, and a series of rapidly emerging mutations from its interaction within the host.

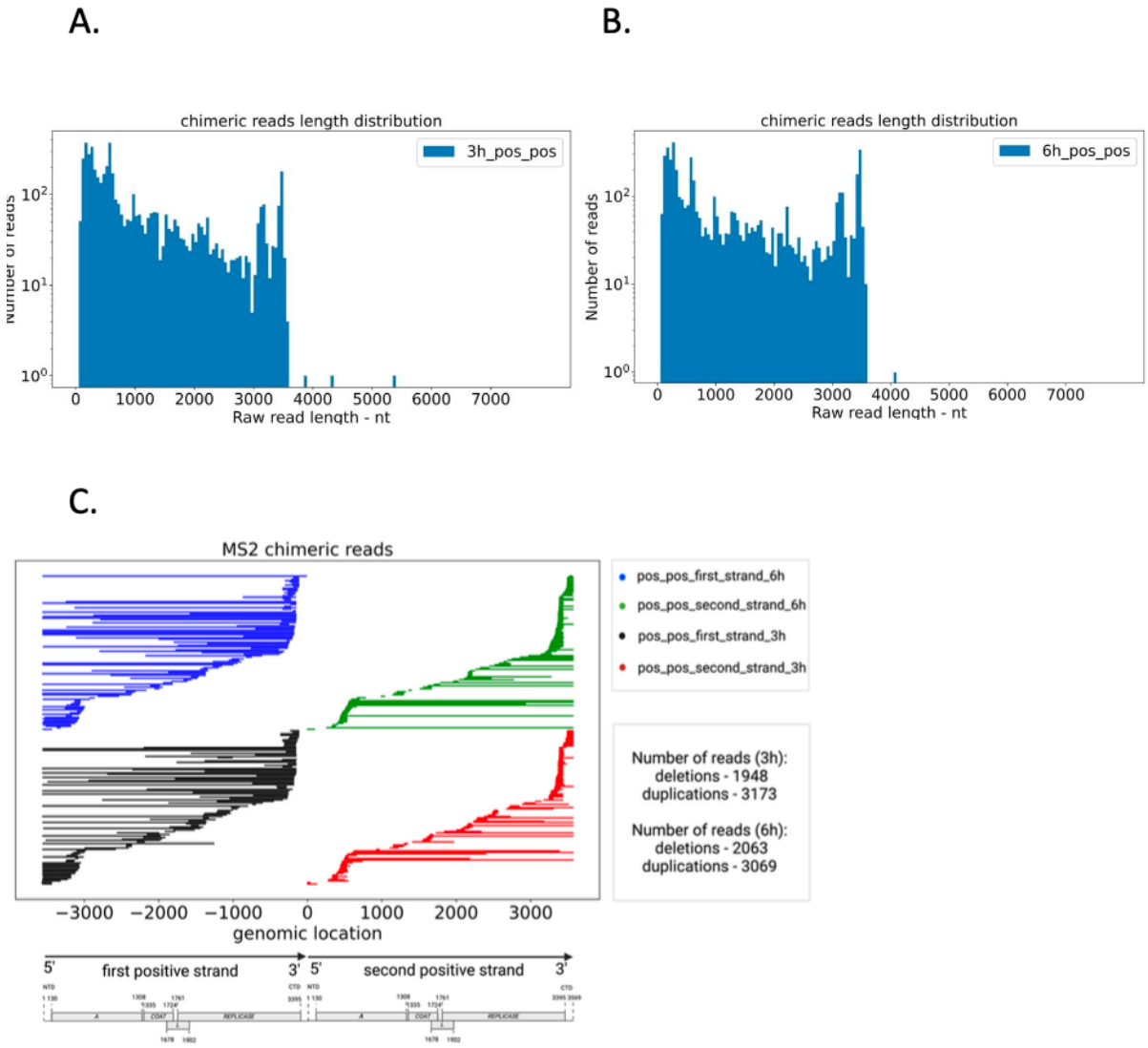

Figure 4. Systematically summarized the characteristics of MS2 chimeric reads (chimeras). A-B. Read length of chimeric reads (pos_pos) histogram at 3h and 6h p.i. **C.** Features of chimeric reads of positive-strand and positive-strand (pos_pos) of MS2. Ligation events happen at the 3' end of the first strand and the 5' of the second strand. 2% of total pos_pos chimeric reads were visualized. (right) pos_pos reads summary at 3h and 6h p.i.

Static and dynamic RNA modification on MS2 phage during replication

Previous works have reported the complex interplay between viral replication and RNA modification (Wang et al., 2023). MS2 genome doesn't encode any modification enzymes and must hijack the bacterial RNA modification enzymes to methylate its own transcript. Previous works have identified the presence of ψ modification on MS2 phages using either mass spectrometry and/or DRS using nanopore sequencing technique (Jones, 2023). We wondered whether new modifications occurred during replication and persisted in progenies. Since RNA modification is thought to be co-transcriptional, we postulated that the combination of the RNA secondary structure and modification affected the MS2 replication and translation efficiency. We predicted and mapped m5C and ψ modifications throughout the MS2 life cycle on the RNA positive and negative strands (**Figure 5 A-D**). We found m^5C, and ψ modifications were widely distributed in both positive- and negative strands (**Figure 5 A-D**). On full-length transcripts (> 3,000 nt), we found no evidence of newly acquired modifications on the positive and negative strands for m5C and ψ. We identified instead commonly modified static sites (present at 0h, 3h and 6h p.i.) with high stoichiometry throughout the transcripts (**Figure 5 A-D**) (marked in red). We found m5C variation in the replicase with and at the 5' end of negative-strand transcripts with a decrease in stoichiometry across time (**Figure 5 A-B**) We also identified unchanged ψ modification across the life cycle at 924 nt mark on MS2 RNA with a high difference score compared with IVT sample (**Figure 5C**), consistent with the previous study validating our approach (Jones, 2023). We observed no modification pattern for m6A on MS2 RNA (**Suppl. Figure S7 A-F**). Overall, it suggests that RNA modifications on MS2 full transcripts are not *de novo* acquired during replication but vary in stoichiometry during the life cycle. We next assessed the RNA modification landscape on shorter transcripts < 800 nt (**Suppl. Figure S8**) to assess the likelihood of acquiring new modifications throughout the life cycle. We found mostly the static modification sites for m5C and ψ were similar to the full-length transcripts.

Interestingly, we found shorter transcripts decrease in m5C stoichiometry across time (**Suppl. Figure 8 A-B**). We also found additional differential methylation sites absent in the full transcript in position 82, 1069, and 1628 nt in the *mat* and *coat* ORFs on the negative strand (**Suppl. Figure S7 B**), and position 674 nt on the positive strand in the *mat* ORF for ψ modification (**Suppl. Figure S8 C**). We found additional methylated sites in positions 2382 nt and 2413 nt in the *rep* ORF for ψ modification on the negative strand (**Suppl. Figure S7 D**). Again, we observed no modification pattern for m6A on MS2 RNA (**Suppl. Figure S7 A-D**). We assessed the secondary structures at these differentially methylated sites and the modified motifs. We found a specific pattern for ψ modification corresponding to a canonical site modification, whereas we found no specific modification sites or secondary structures for m5C modification. The site of ψ installation on the MS2 RNA was conserved; bacterial ψ synthase recognizes the secondary structure and installs ψ either within the loops or at the stem-loop junctions (**Suppl Figure S9 A-J**). MEME motif pattern showed the specificity of ψ synthase to recognize a conserved 5'- ψU-3' motif (**Suppl Figure S9 K**). No specific patterns (secondary structures and motifs) were observed for m5C modification. We assessed whether these newly formed methylated sites and differentially methylated sites coincided with the region of transcriptional activity during replication. We found no evidence that these methylated sites were important for translation. We noticed, however, that these regions were more amenable to RNA modification, such as U-rich hairpins for ψ modification. 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.

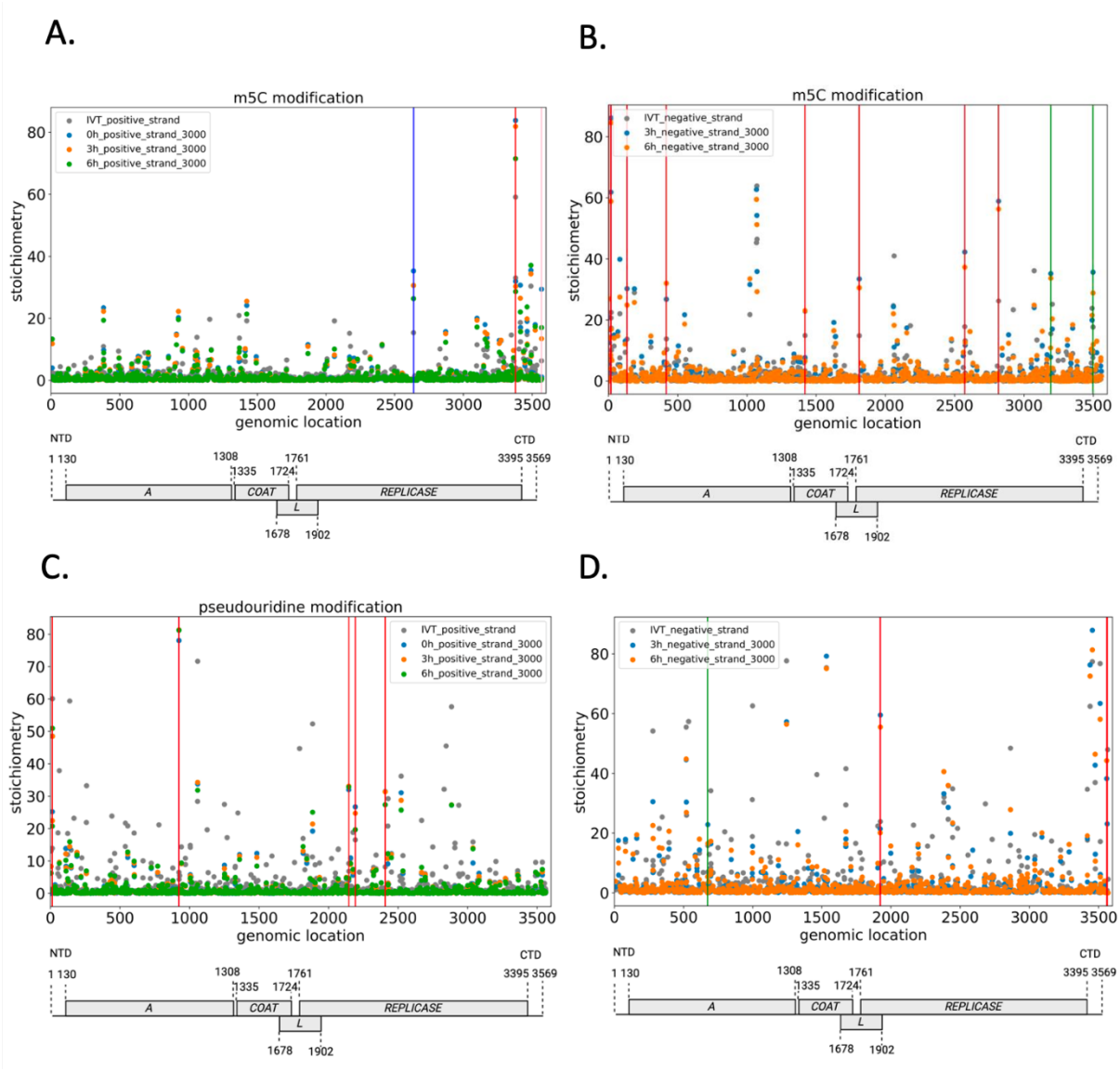

Figure 5. Frequent potential sites of ψ and m^{5}C RNA modification dynamics on MS2 phage positive and negative RNA (>3000 nt length) via Dorado and Modkit software packages. Synthetic modification-free RNA as the negative control. A. m5C modification (positive strand) landscapes according to the counts of modified and unmodified based generated by Modkit. **B.** m5C modification (negative strand) landscapes **C. ψ** modification (positive strand) landscapes **D.** pseudouridine modification (negative strand) landscapes. With selection criteria based on single-base analysis (those that meet all the requirements are highlighted in vertical lines): Presence of PTMs at all 0h, 3h, 6h (vertical red line), only 0h vs. 6h decrease (vertical blue line), only 3h vs. 6h decrease (vertical green line), only 3h vs. 6h increase (vertical black line), 0h vs. 3h decrease then 3h vs. 6h increase (vertical pink line). Stoichiometry (test) > 20%, 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 (FPR) (Benjamini & Hochberg, 1995). 5-methylcytidine, m^{5}C.

Discussion:

RNA bacteriophages such as Qβ and MS2 have been extensively studied for their ability to rapidly replicate and assemble into a mature virus (Chang et al., 2022; Vasilyev et al., 2013; Wagner, Weise, and Mutschler, 2022). However, the dynamics of replication during its life cycle remain poorly studied and relatively unknown. This study aimed to detail in real-time the dynamics of MS2 transcriptional activity using the Nanopore DRS approach. We found that MS2 phage maintained a high transcriptional activity, especially expressing high amounts of small transcripts containing the coat protein-encoding region, hybrid RNA and byproducts from the replicase. The replication initiation of the positive-strand small transcripts is mediated by cryptic ribosomal binding sites on the negative strand, recognized by the replicase complex. By investigating the RNA modification landscapes, we conserved site (5'-ΨU-3') and secondary structure (stem-loops) amenable to Ψ modification installation on MS2 phage RNA. Together, direct RNA Nanopore sequencing has allowed the dissection of the detailed dynamics of MS2 transcriptional activity during its life cycle. This has led to our main finding that despite carrying a minimal genome, MS2 could considerably enhanced its transcriptional activity and harness its error-prone replicase to rapidly attach, replicate, and lyse the bacterial cell.

High transcriptional/replication level of MS2 within bacterial cells

Our first finding indicated a high level of transcriptional activity of MS2 within the bacterial cell, evolving across the life cycle leading to replication and lysis. Previous works investigated in depth the replication pattern (including host factors) of RNA phages (Vasilyev et al., 2013) and RNA viruses (Hillen et al., 2020) to generate full-length infective mRNA (or genomic RNA). MS2/Qβ replicase complex has been extensively investigated since the years sixties, including the host factors responsible for forming the complex (Wagner et al., 2022) and their underlying roles (Vasilyev et al., 2013). However, little is known about the dynamics of the

transcriptional activity during the MS2 phage life cycle. A key reason is most previous studies were performed in an *in-vitro* or a transcription-translation system for MS2 (Wagner et al., 2022) and Qβ (Vasilyev et al., 2013; Tomita, Ichihashi, and Yomo, 2015). Using Nanopore RNA DRS, we showed and dissected for the first time the high level of transcriptional activity of MS2 phage. We reported significant changes in MS2 transcripts level across its life cycle by increased expression of *mat* and *cap* and downregulation of *lys* and *rep,* indicating the balance up- to down-regulation of the phage ORFs during the life cycle, which is potentially tightly regulated at each stage to either, absorb, grow, replicate, assemble and lyse. Strikingly, we noticed an over 500-fold increase in full-length MS2 transcripts in only three hours post-inoculation, indicative of high activity of the phage 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 (Viehweger et al., 2019) 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) (Kim et al., 2020). As such, our observation is in line with previous reports on other viruses suggesting the ability of a virus to rapidly transcribe within hours of inoculation. Together, our study reports that RNA phages such as MS2 require significant upregulating and fine-tuning of the expression of their ORFs to rapidly replicate, assemble, and lyse. An important question remains whether MS2 error-prone replicase is more prone to rapidity over precision.

Formation of cryptic small ORF transcripts during MS2 life cycle.

Previous works have reported that MS2 is transcribed as a polycistronic single full-length RNA and utilizes its negative strand for replication via the replicase complex (Dykeman, 2023). Importantly, the coat protein plays a crucial role in phage encapsidation and translational

repression of the replicase by binding to specific stem loops on MS2 RNA (Rolfsson et al., 2016).

As we identified numerous small transcripts mapped throughout the MS2 genome, we wondered whether some are required for replication/encapsidation and fine-tuning. Our second important finding is we identified hotspots of mapped transcripts corresponding to the full-length coat protein ORF and the, for the first time, presence of cryptic ribosomal binding sites (RBS) on the negative strand of the MS2 genome during replication. Previous works have reported the position of a now well-described and heavily used in biotechnology applications RBS upstream of the coat protein. and suggested the likely presence of additional RBS across the genome (Kastelein et al., 1983). Using the RNA DRS technique and mapping the starts of the transcripts, we have confirmed this hypothesis and found additional RBS totalling five across the genome. Additionally, we reported that this transcription is regulated by the occupancy of the coat protein on the positive strand on specific RNA binding sites, aborting the transcription/translation from the replicase complex and confirming previous findings (Rolfsson et al., 2016). Together, it suggests that MS2 phage produces, in addition to the cistrons, a high level of coat protein transcripts to repress the replicase complex. There are large amounts of subgenomic transcripts containing ORFs of structural proteins, such as nucleocapsid protein (N) and membrane protein (M), being detected in the SARS-CoV-2 transcriptome (Kim et al., 2020). It indicated that transcripts encoding structural proteins should remain at a high level for the translation of enough structural protein for assembly, both important for prokaryotic viruses and eukaryotic viruses. Although the MS2 phage and eukaryotic viruses both utilize negative strands to synthesize their positive strand during transcription, the mechanisms are different. Eukaryotic viruses utilize discontinuous transcription for highly effective transcription (Grelewska-Nowotko, Elhag, and Turowski,

2025). Together, our data and previous work suggest independent high-level production of coat transcripts during the MS2 life-cycle to assemble, repress, and control the transcription/translation required for virus replication, relaying on the coat protein replicase activities.

Error-prone replicase complex generates large amounts of hybrid transcripts

The replicase plays a critical role in phage replication. An important requirement for some viruses to replicate and evolve is to harbour either proofreading (Robson et al., 2020) or non-proofreading replicase (Whitfield et al., 2020). Previous works have reported that MS2 replicase generates large amounts of phage transcripts rapidly (Wagner, Weise, and Mutschler, 2022). An unknown was the trade-off between efficiency and precision in transcription. Using direct RNA sequencing, our third major finding is a large amount of short sub-genomic RNA transcripts are produced from MS2 during its life cycle. We identified the nature of these transcripts by mapping these transcripts to the MS2 genome. We found that RNA species were either "parasitic RNA" or gRNA containing truncated fragments, which were considered defective. Previous studies showed that the replicase complex was able to produce short-amplifying RNAs as by-products during MS2 genome replication (Wagner, Weise, and Mutschler, 2022). Previous work reported that Qβ replicase generates huge amounts of small RNA, but template-independent (random RNAs) unmapped to the Qβ genome (Chetverin, Chetverina, and Munishkin, 1991). It suggests that MS2 replicase produces a large number of by-product transcripts while replicating its own genome. We hypothesized these by-products were full cistron transcripts, template switch RNA, or hybrid RNA products of the error-prone replicase.

Another mechanism for viruses to acquire novel traits and evolve is to generate mutations and structural variants or form hybrids with other viruses or within the same virus (Simon-Loriere and Holmes, 2011). We wondered whether the error-prone replicase would favour the generation of *de-novo* mutations and such hybrids RNA to form new viral combinations. Using direct RNA sequencing, we reported for the first time long intraspecies hybrids transcripts (>3569 nt), including fold-back RNAs. We found positive-to-positive strand, negative-to-negative strand, and positive-to-negative strand hybrids in addition to small transcripts spanning the entire genome. We also reported the emergence of *de-novo* mutations across the MS2 life cycle. The presence of hybrid RNA or recombination events has been reported for other eukaryotic viruses such as SARS-CoV-2 (Garushyants, Rogozin, and Koonin, 2021) or plant viruses (Johnson et al., 2024). Hybrid RNAs also have been reported for MS2 in plasmid systems (van Meerten et al., 2002). Confirming previous works, our analysis reports the formation of hybrids RNA and *de novo* mutation during transcription. Together, they indicate the ability of MS2 to rapidly evolve, recombine, and acquire new mutations from its error-prone replicase.

Conserved sequence (5'-ΨU-3') and secondary structure (stem-loops) of Ψ installation sites on native MS2 RNA

Previous works have evidenced the crucial role of RNA modification in viral replication (Bhattacharya et al., 2022). To achieve high translation efficiency and evade host immunity, some viruses encode their own methyltransferase or other modifying enzymes (Li and Rana, 2022). However, the MS2 genome doesn't harbor any modifying enzymes and requires hijacking the host modification machinery to modify its RNA. Using direct RNA sequencing, we aimed to assess whether new modifications were acquired and removed during the phage cycle to modulate MS2 transcriptional activity. Our fourth major finding evidenced the

presence of Ψ installation site (5'-ΨU-3') and secondary structure (stem-loop) amenable to Ψ modification. Until now, there are no ψ erasers identified (Eyler et al., 2019), indicating conserved modification patterns in transcripts >3000 nt across the MS2 life cycle. Using bacterial cell lysates, mRNAs with ψ modifications were not translated (Karikó et al., 2008), which is consistent with another study that showed ψ-installation impedes translation elongation (Eyler et al., 2019). We observed the presence of ψ on the positive strand of MS2 transcripts (>3000 nt) within the *mat*-encoding region and *rep*-encoding region. Interestingly, we found a dynamic of loss and gain in methylation of the small transcripts during the MS2 life cycle. It suggests that, unlike the full transcripts, small transcripts are more amenable to RNA modification and were more prevalent on the negative strand. However, we couldn't establish clear functional evidence of their role in MS2 cistrons translation. RNA modification might play a role in helping viruses escape host defense, enhancing viral protein translation, and keeping viral mRNA stability (Lu et al., 2020).

Conclusion

Our investigation capturing the dynamics of MS2 transcriptional activity across its life cycle using the Nanopore direct RNA sequencing approach has revealed a complex interplay between the ORFs, 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 likely required for replication. We finally reported the presence or rapidly emerging de novo mutations in the genome and the dynamics of acquisition and removal of modified RNA in small transcripts. Together, the direct RNA sequencing approach provides significant insights into the transcriptional dynamics that have led to the confirmation of previous hypotheses formulated 30 years ago and has revealed new insights into MS2 biology.

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 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:

Adachi, H., Hengesbach, M., Yu, Y.T. and Morais, P., 2021. From antisense RNA to RNA modification: therapeutic potential of RNA-based technologies. Biomedicines, 9(5), p.550.

Anderson, B.R., Muramatsu, H., Nallagatla, S.R., Bevilacqua, P.C., Sansing, L.H., Weissman, D. and Karikó, K., 2010. Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. Nucleic acids research, 38(17), pp.5884-5892.

Anderson, B.R., Muramatsu, H., Jha, B.K., Silverman, R.H., Weissman, D. and Karikó, K., 2011. Nucleoside modifications in RNA limit activation of 2′-5′-oligoadenylate synthetase and increase resistance to cleavage by RNase L. Nucleic acids research, 39(21), pp.9329-9338.

Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994.

Bayoumi, M. and Munir, M., 2021. Evolutionary conservation of the DRACH signatures of potential N6-methyladenosine (m6A) sites among influenza A viruses. *Scientific Reports*, *11*(1), p.4548.

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.

Betancourt, A.J., 2009. Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. *Genetics*, *181*(4), pp.1535-1544.

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.

Boni, I.V., lsaeva, D.M., Musychenko, M.L. and Tzareva, N.V., 1991. Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic acids research*, *19*(1), pp.155-162.

Callanan, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R.P. and Hill, C., 2018. RNA phage biology in a metagenomic era. Viruses, 10(7), p.386.

Charette, M. and Gray, M.W., 2000. Pseudouridine in RNA: what, where, how, and why. IUBMB life, 49(5), pp.341-352.

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.

Chrisman, B.S., Paskov, K., Stockham, N., Tabatabaei, K., Jung, J.Y., Washington, P., Varma, M., Sun, M.W., Maleki, S. and Wall, D.P., 2021. Indels in SARS-CoV-2 occur at template-switching hotspots. *BioData mining*, *14*, pp.1-16.

Comeau, A.M., Hatfull, G.F., Krisch, H.M., Lindell, D., Mann, N.H. and Prangishvili, D., 2008. Exploring the prokaryotic virosphere. *Research in microbiology*, *159*(5), pp.306-313.

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.

Davis, D.R., 1995. Stabilization of RNA stacking by pseudouridine. Nucleic acids research, 23(24), pp.5020-5026.

De Coster, W. and Rademakers, R., 2023. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics*, *39*(5), p.btad311.

De Smit, M.H. and Van Duin, J., 1993. Translational initiation at the coat-protein gene of phage MS2: native upstream RNA relieves inhibition by local secondary structure. *Molecular Microbiology*, *9*(5), pp.1079-1088.

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.

Dykeman, E.C., Grayson, N.E., Toropova, K., Ranson, N.A., Stockley, P.G. and Twarock, R., 2011. Simple rules for efficient assembly predict the layout of a packaged viral RNA. *Journal of molecular biology*, *408*(3), pp.399-407.

Eyler, D.E., Franco, M.K., Batool, Z., Wu, M.Z., Dubuke, M.L., Dobosz-Bartoszek, M., Jones, J.D., Polikanov, Y.S., Roy, B. and Koutmou, K.S., 2019. Pseudouridinylation of mRNA coding sequences alters translation. Proceedings of the National Academy of Sciences, 116(46), pp.23068-23074.

Fedewa, G., Radoshitzky, S.R., Chī, X., Dǒng, L., Zeng, X., Spear, M., Strauli, N., Ng, M., Chandran, K., Stenglein, M.D. and Hernandez, R.D., 2018. Ebola virus, but not the Marburg virus, replicates efficiently and without required adaptation in snake cells. *Virus Evolution*, *4*(2), p.vey034.

Fernández, I.S., Ng, C.L., Kelley, A.C., Wu, G., Yu, Y.T. and Ramakrishnan, V., 2013. Unusual base pairing during the decoding of a stop codon by the ribosome. Nature, 500(7460), pp.107-110.

Gabay, O., Shoshan, Y., Kopel, E., Ben-Zvi, U., Mann, T.D., Bressler, N., Cohen-Fultheim, R., Schaffer, A.A., Roth, S.H., Tzur, Z. and Levanon, E.Y., 2022. Landscape of adenosine-to-inosine RNA recoding across human tissues. Nature Communications, 13(1), p.1184.

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.

Golovina, A.Y., Sergiev, P.V., Golovin, A.V., Serebryakova, M.V., Demina, I., Govorun, V.M. and Dontsova, O.A., 2009. The yfiC gene of E. coli encodes an adenine-N6 methyltransferase that specifically modifies A37 of tRNA1Val (cmo5UAC). Rna, 15(6), pp.1134-1141.

Grelewska-Nowotko, K., Elhag, A.E. and Turowski, T.W., 2025. Transcription Kinetics in the Coronavirus Life Cycle. *Wiley Interdisciplinary Reviews: RNA*, *16*(1), p.e70000.

Groeneveld, H.E.R.M.A.N., Thimon, K.A.F.I.I.N.E. and Van Duin, J., 1995. Translational control of maturation-protein synthesis in phage MS2: a role for the kinetics of RNA folding?. Rna, 1(1), p.79.

Gu, R., Zhang, Z., DeCerbo, J.N. and Carmichael, G.G., 2009. Gene regulation by sense–antisense overlap of polyadenylation signals. *Rna*, *15*(6), pp.1154-1163.

Harb, L., Chamakura, K., Khara, P., Christie, P.J., Young, R. and Zeng, L., 2020. ssRNA phage penetration triggers detachment of the F-pilus. Proceedings of the National Academy of Sciences, 117(41), pp.25751-25758.

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.

Johnson, B. and Szekely, M., 1977. Specific binding site of E. coli initiation factor 3 (IF3) at a 3′-terminal region of MS2 RNA. *Nature*, *267*(5611), pp.550-552.

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

Karikó, K., Buckstein, M., Ni, H. and Weissman, D., 2005. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. Immunity, 23(2), pp.165-175.

Karikó, K., Muramatsu, H., Welsh, F.A., Ludwig, J., Kato, H., Akira, S. and Weissman, D., 2008. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. Molecular therapy, 16(11), pp.1833-1840.

Keller, M.W., Rambo-Martin, B.L., Wilson, M.M., Ridenour, C.A., Shepard, S.S., Stark, T.J., Neuhaus, E.B., Dugan, V.G., Wentworth, D.E. and Barnes, J.R., 2018. Direct RNA sequencing of the coding complete influenza A virus genome. Scientific reports, 8(1), p.14408.

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.

Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G. and Smith, D.B., 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic acids research*, *46*(D1), pp.D708-D717.

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.

Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E. and Jaffrey, S.R., 2015. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nature methods, 12(8), pp.767-772.

Liu, J., Chen, L., Guo, X., Zhao, B. and Jiang, J., 2024. Emerging role of N6-methyladenosine RNA modification in regulation of SARS-CoV-2 infection and virus-host interactions. *Biomedicine & Pharmacotherapy*, *173*, p.116231.

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.

Lu, M., Zhang, Z., Xue, M., Zhao, B.S., Harder, O., Li, A., Liang, X., Gao, T.Z., Xu, Y., Zhou, J. and Feng, Z., 2020. N 6-methyladenosine modification enables viral RNA to escape recognition by RNA sensor RIG-I. *Nature microbiology*, *5*(4), pp.584-598.

Mawassi, M., Satyanarayana, T., Albiach-Martí, M.R., Gowda, S., Ayllón, M.A., Robertson, C. and Dawson, W.O., 2000. The fitness of citrus tristeza virus defective RNAs is affected by the lengths of their 5′-and 3′-termini and by the coding capacity. *Virology*, *275*(1), pp.42-56.

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.

Morais, P., Adachi, H. and Yu, Y.T., 2021. The critical contribution of pseudouridine to mRNA COVID-19 vaccines. Frontiers in cell and developmental biology, 9, p.789427.

Naylor, R., Ho, N.W. and Gilham, P.T., 1965. Selective chemical modifications of uridine and pseudouridine in polynucleotides and their effect on the specificities of ribonuclease and phosphodiesterases. Journal of the American Chemical Society, 87(18), pp.4209-4210.

Newby, M.I. and Greenbaum, N.L., 2001. A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. Rna, 7(6), pp.833-845.

Newby, M.I. and Greenbaum, N.L., 2002. Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. Proceedings of the National Academy of Sciences, 99(20), pp.12697-12702.

Newby, M.I. and Greenbaum, N.L., 2002. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. Nature structural biology, 9(12), pp.958-965.

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

Parr, C.J., Wada, S., Kotake, K., Kameda, S., Matsuura, S., Sakashita, S., Park, S., Sugiyama, H., Kuang, Y. and Saito, H., 2020. N 1-Methylpseudouridine substitution enhances the performance of synthetic mRNA switches in cells. Nucleic Acids Research, 48(6), pp.e35-e35.

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.

Patterson, J.B., Cornu, T.I., Redwine, J., Dales, S., Lewicki, H., Holz, A., Thomas, D., Billeter, M.A. and Oldstone, M.B., 2001. Evidence that the hypermutated M protein of a subacute sclerosing panencephalitis measles virus actively contributes to the chronic progressive CNS disease. *Virology*, *291*(2), pp.215-225.

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.

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.

Rūmnieks, J., Liekniņa, I., Kalniņš, G., Šišovs, M., Akopjana, I., Bogans, J. and Tārs, K., 2020. Three-dimensional structure of 22 uncultured ssRNA bacteriophages: Flexibility of the coat protein fold and variations in particle shapes. *Science advances*, *6*(36), p.eabc0023.

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

Simon, A.E. and Gehrke, L., 2009. RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. *Biochimica Et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1789*(9-10), pp.571-583.

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

Slotkin, W. and Nishikura, K., 2013. Adenosine-to-inosine RNA editing and human disease. *Genome medicine*, *5*, pp.1-13.

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.

Spitale, R.C. and Incarnato, D., 2023. Probing the dynamic RNA structurome and its functions. Nature Reviews Genetics, 24(3), pp.178-196.

Teng, H., Stoiber, M., Bar-Joseph, Z. and Kingsford, C., 2024. Detecting m6A RNA modification from nanopore sequencing using a semi-supervised learning framework. *bioRxiv*.

Thongchol, J., Lill, Z., Hoover, Z. and Zhang, J., 2023. Recent advances in structural studies of single-stranded RNA bacteriophages. Viruses, 15(10), p.1985.

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.

Wang, H., Feng, J., Fu, Z., Xu, T., Liu, J., Yang, S., Li, Y., Deng, J., Zhang, Y., Guo, M. and Wang, X., 2024. Epitranscriptomic m5C methylation of SARS-CoV-2 RNA regulates viral replication and the virulence of progeny viruses in the new infection. *Science Advances*, *10*(32), p.eadn9519.

Wang, X., Chang, Z., Zhao, T., Zhong, W., Shi, J., Wang, G. and Xu, X., 2023. The role of post-transcriptional regulation in SARS-CoV-2 infection and pathogenicity. *Frontiers in Immunology*, *14*, p.1256574.

Wells, H.L., Bonavita, C.M., Navarrete-Macias, I., Vilchez, B., Rasmussen, A.L. and Anthony, S.J., 2023. The coronavirus recombination pathway. *Cell host & microbe*, *31*(6), pp.874-889.

Westhof, E., 2019. Pseudouridines or how to draw on weak energy differences. Biochemical and biophysical research communications, 520(4), pp.702-704.

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).

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.