

Challenges of isolating bacteriophage mRNA for chemical analysis

We struggled to isolate enough phage mRNA for HPLC as we searched for new nucleoside chemistries. Ribodepleting the abundant extracted rRNA introduced contaminating DNA, and we were still left with more bacterial mRNA than phage transcripts. We suggest an alternative approach.

Version 2, published Aug 1, 2023. Originally published Mar 13, 2023.

 Arcadia Science

DOI: 10.57844/arcadia-j03a-mz33

Purpose

As part of our effort to identify new phage nucleic acid modifications, we sought to isolate bacteriophage mRNA for chemical analysis by high-performance liquid chromatography (HPLC) and liquid-chromatography mass spectrometry (LC-MS).

While we are not continuing this project, we thought it would be helpful to share roadblocks we encountered in isolating sufficient quantities of high-quality, intact phage mRNA for chemical analysis. We describe our experience and suggest ideas for troubleshooting that others might try if pursuing similar goals.

- This pub is part of the **project**, "[Exploring bacteriophage nucleic acid chemistries](#)." Visit the project narrative for more background and context.
- **Data** from this pub, including raw reads, is accessible in the [European Nucleotide Archive \(ENA\)](#).
- You can find step-by-step **protocols** for [calculating MOI](#) and performing [timed phage infection and cell harvest](#) on protocols.io.
- All associated **code** is available in [this GitHub repository](#).

Project background and goals

As part of the [bacteriophage nucleic acid modifications project](#), we wished to set up a new workflow to identify phage mRNA modifications. Our aim was to isolate intact bacteriophage mRNA and carry out chemical analysis to identify base modifications. We hit a snag in isolating sufficient amounts of bacteriophage mRNA from infected cells and started to think through workarounds before deciding to discontinue the project altogether.

The approach

We have to extract bacteriophage mRNA from host bacteria since host cell machinery is used for phage mRNA synthesis. Within bacterial cells, the total cellular RNA consists of ~95–97% ribosomal RNA (rRNA) and tRNA [1], so ribosomal RNA (rRNA) depletion is necessary to enrich mRNA for downstream analysis. For these reasons, isolating sufficient quantities of phage mRNA for chemical analysis is extra challenging.

Our goal was to develop a protocol that would maximize the yield of phage mRNA transcripts. This is likely essential for detection of RNA base modifications, particularly if they occur at low frequencies within phage transcripts. To this end, we set up a pilot study where we carried out timed infections of *Escherichia coli* strain B and *Bacillus subtilis* strain 168 with T4 and SPO1 phages, respectively. We then extracted cellular RNA, and depleted host ribosomal RNA in an effort to enrich phage mRNA.

In our pilot study, we infected *E. coli* with phage T4 at a multiplicity of infection (MOI) of 100, and infected *B. subtilis* with SPO1 phage at an MOI of 2.5. While a higher MOI would have been preferred for the SPO1 infection, this was the highest MOI infection we could achieve with our phage stock. We harvested cells 15 minutes post-infection. Based on previous literature, we predicted that this time point would correspond to middle–late infection in the life cycle of T4, and early–middle infection for SPO1 [2][3]. We subsequently extracted cellular RNA using the Monarch Total RNA Miniprep Kit, as per the manufacturer's instructions, and then depleted rRNA using the NEBNext rRNA Depletion Kit (Bacteria). We nuclease-

digested the RNA we obtained, and chemically analyzed by HPLC, comparing samples with and without rRNA depletion. Concurrently, we determined the proportion of phage mRNA relative to the bacterial host using RNA sequencing.

Detailed methods

Phage infection and cell harvest

We're sharing a protocol for calculating MOI and another that explains how to carry out timed infections of *E. coli* and *B. subtilis* with phages T4 and SPO1, respectively. This can be adapted for any phage-host pair.

TRY IT: Protocols for "[Calculating multiplicity of infection \(MOI\)](#)" and "[Phage infection and timed harvest of *E. coli* and *B. subtilis* cells](#)" are available on protocols.io.

RNA extraction and rRNA depletion

We used the Monarch Total RNA Miniprep Kit (NEB, #T2010S) for RNA extractions and added 10 mg/ml lysozyme (ThermoFisher Scientific, #90082) to lyse cells. We carried out ribosomal RNA depletion using the NEBNext rRNA Depletion Kit (Bacteria) (NEB, #E7850S).

HPLC analysis

We performed HPLC analysis of nucleic acids as described in [this protocol](#).

RNA sequencing and analysis

Novogene carried out RNA QC, rRNA depletion, library prep, and sequencing (Illumina NovaSeq 6000 platform). We pooled both the *E. coli* and *B. subtilis* samples into one library prep to cut down on cost, and used read-mapping to disambiguate the samples.

We then determined the fraction of the RNA-seq sample that mapped to *E. coli*, *B. subtilis*, phage T4, and phage SPO1. We first downloaded the reference genome for each organism (Table 1) and then combined the genomes into a single reference FASTA file. We used `bwa mem` (version 0.7.17) [4] to map the RNA-seq reads back to

the single reference file, and used `samtools idxstats` and `samtools depth` (version 1.16.1) [5] to determine the number of reads that mapped to each reference and at each position in each reference. We also used the `featureCounts()` function in the Rsubread package (version 2.8.1) [6] to count the number of reads that mapped to each gene in each reference. Last, we used `tidyverse` (version 1.3.2) [7] to visualize these analyses.

Genome	GenBank accession
<i>E. coli</i>	GCF_001559635.1
<i>B. subtilis</i>	GCF_000009045.1
Phage T4	GCF_000836945.1
Phage SPO1	GCF_000881675.1

Table 1. GenBank Genome accessions used in this pub.

The **code** we used to analyze RNA-seq data is available at [this GitHub repository](#) (DOI: [10.5281/zenodo.7719755](https://doi.org/10.5281/zenodo.7719755)).

Data deposition

We deposited reads (FASTQ file) in the [ENA](#) (project PRJEB60535).

The results

Upon isolating RNA from phage-infected *E. coli* and *B. subtilis* cells harvested at 15 minutes post-infection, we successfully depleted rRNA with the NEBNext rRNA Depletion Kit ([Figure 1](#)). We used the maximum input, one microgram of RNA, and recovered a low amount of mRNA, consistent with the expected proportion of mRNA to rRNA within the cells (Table 2). This was only sufficient for one round of HPLC analysis, near the limit of detection of our instrument.

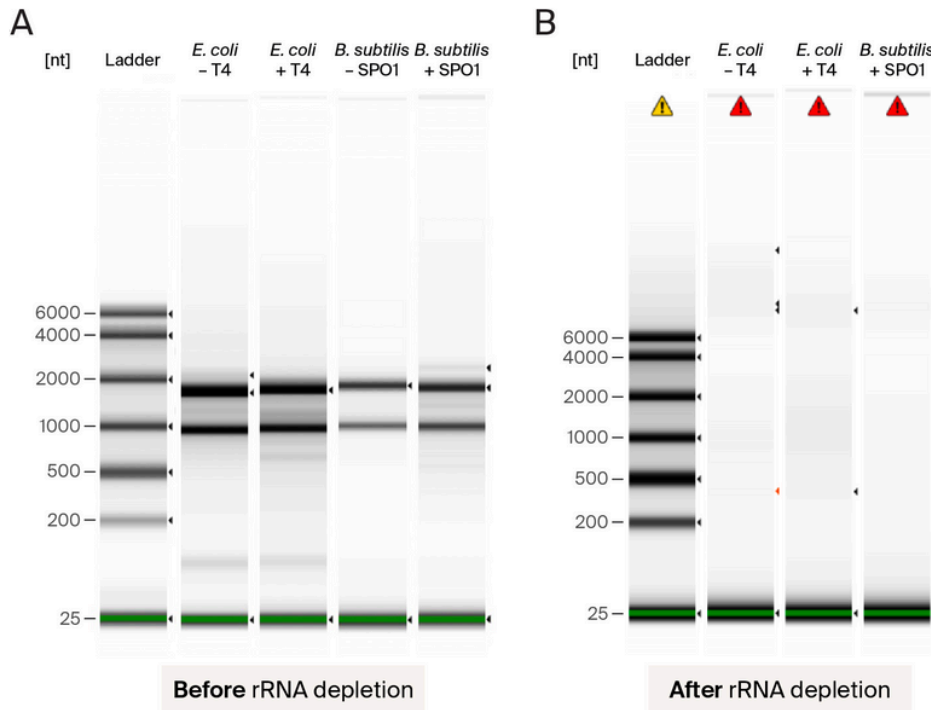


Figure 1. **TapeStation electrophoresis gel showing successful rRNA depletion.**
 (A) Total cellular RNA samples analyzed by TapeStation before rRNA depletion.
 (B) Samples analyzed after rRNA depletion using the NEBNext rRNA Depletion kit.

Sample	<i>Before rRNA depletion</i>		<i>After rRNA depletion</i>	
	[RNA] (ng/ μL)	Total RNA (ng)	[RNA] (ng/ μL)	Total RNA (ng)
<i>E. coli</i> (control)	111	1110	5.64	56.4
<i>E. coli</i> + phage T4 (15 min post-infection)	95.7	957	6.96	69.6
<i>B. subtilis</i>	30.9	309	-	-
<i>B. subtilis</i> + phage SPO1 (15 min post-infection)	50	500	4.66	46.6

Table 2. RNA yield before and after ribosomal RNA depletion. Measured on TapeStation (Agilent).

By analyzing HPLC runs of RNA from *E. coli* and *E. coli* infected with phage T4, we further discovered that rRNA depletion introduced DNA, which co-eluted with the purified mRNA (Figure 2). The rRNA depletion kit uses DNA oligonucleotide probes for hybridization with rRNA, facilitating subsequent degradation of the hybrid by RNaseH [8]. DNase I is then added to degrade these DNA probes, and the user isolates the mRNA using a bead-based cleanup. The contaminating DNA in the

HPLC analysis originates either from undigested oligonucleotide probes that co-purified with the mRNA during the bead cleanup or digested DNA nucleotides that were carried over at a low level. These contaminating DNA nucleosides are present in extremely low amounts, and are only an issue because our mRNA yield is very low. These contaminating DNA peaks present a problem because they could easily obscure signal from the modified nucleosides we were hoping to find.

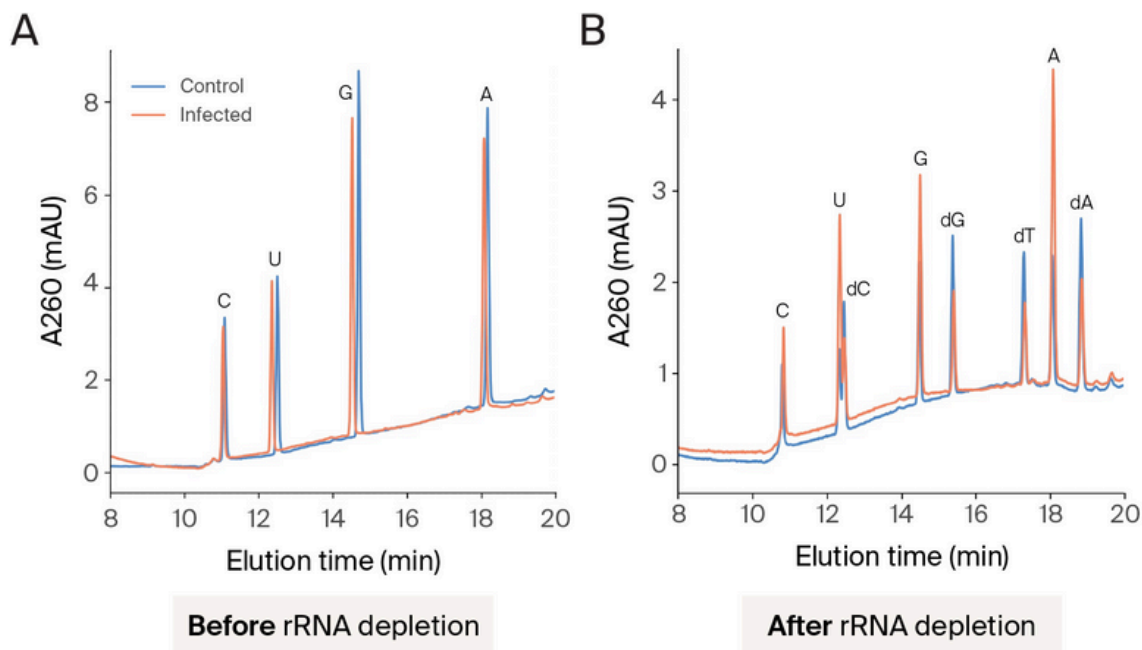


Figure 2. **Depleting ribosomal RNA in phage RNA extracts introduces DNA nucleosides.**

HPLC chromatograms of RNA obtained from *E. coli* (blue, control) and *E. coli* infected with phage T4 (orange) before (A) and after (B) rRNA depletion. We've labeled peaks with the corresponding nucleosides (d: deoxy) based on retention times of nucleoside standards. We diluted the RNA samples from before rRNA depletion to match the concentration of RNA obtained after rRNA depletion, which we injected undiluted into the HPLC column.

To measure the ratio of phage RNA to bacterial RNA in our samples, we performed RNA sequencing on T4 and SPO1 infection that underwent rRNA depletion. This revealed that 39% of the reads from the T4 infection samples mapped to T4, and 17.9% of the reads from the SPO1 infection mapped to SPO1 (Figure 3). This means that the majority of the RNA we analyzed with HPLC was bacterial in origin. This further limits our ability to detect RNA modifications that are specific to phage mRNA.

SHOW ME THE DATA: Access our [RNA sequencing data](#).

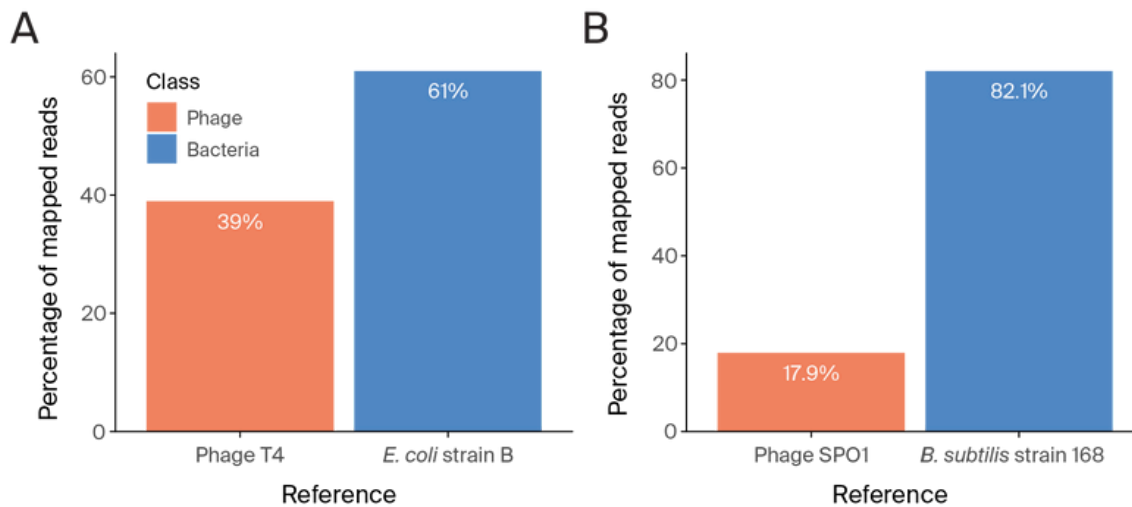


Figure 3. **The majority of reads from sequencing ribodepleted samples map to bacterial hosts.**

(A) Percentage of reads that map to the phage T4 vs. *E. coli* genomes in RNA we obtained from *E. coli* cells that we harvested 15 minutes post-infection with phage T4 at an MOI of 100.

(B) Percentage of reads that map to the SPO1 and *B. subtilis* genomes in RNA we obtained from *B. subtilis* cells that we harvested 15 minutes post-infection with phage SPO1 at an MOI of 2.5.

Sequencing coverage depths across the T4 and SPO1 genomes revealed transcriptional activity across the T4 genome while the SPO1 genome only had transcriptional activity at the start of its genome (Figure 4). This region of the SPO1 genome encodes the early-expressed “host shut-off” genes [9]. It’s likely that the higher level of T4 transcripts relative to SPO1 transcripts is due to T4 being farther into its life cycle than SPO1. If we were to repeat this experiment, we would extract SPO1 mRNA at a later time point.

The **code** we used to analyze RNA-seq data is available at [this GitHub repository](#) (DOI: [10.5281/zenodo.7719755](https://doi.org/10.5281/zenodo.7719755)).

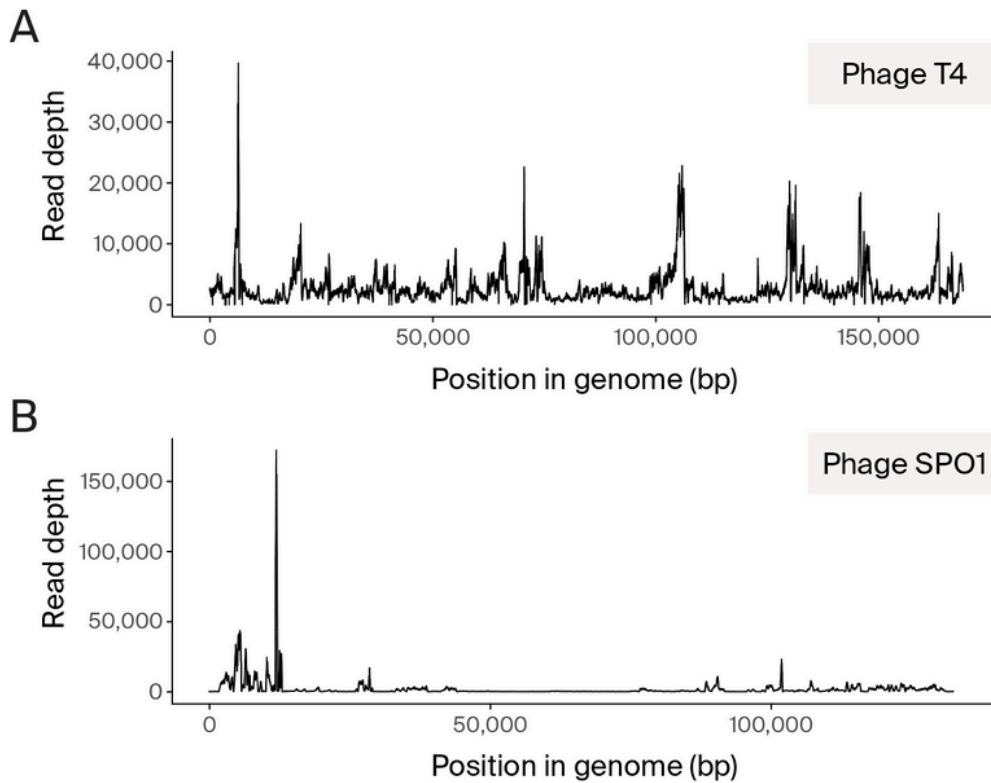


Figure 4. **Phage T4 transcripts that we detected by RNA-seq map come from throughout the genome, while phage SPO1 transcripts predominantly map to the start of its genome.**

RNA sequencing read depth across (A) phage T4 and (B) phage SPO1 genomes.

Overall, these observations highlight that when harvesting phage mRNA, it is important to consider MOI input to ensure that all cells are productively infected, and to harvest phage late in the infection life cycle to obtain maximum phage mRNA yield. Also, an ideal method would more effectively deplete bacterial mRNAs to facilitate targeted analysis of the phage mRNA fraction.

Key takeaways

Analyzing phage mRNA chemistry is much more challenging than analyzing phage DNA chemistry. To study phage DNA chemistry, we can purify phage particles away from their bacterial hosts before extracting DNA. This allows us to obtain high-purity phage DNA. However, phage mRNA is only produced within the context of an infected cell. Thus, obtaining significant amounts of high-purity phage mRNA becomes a challenging numbers game.

When we harvest total RNA from infected cells, the vast majority is bacterial rRNA. After removing the rRNA with ribodepletion, the remaining mRNA will be a mix of phage and bacterial transcripts. We find that bacterial transcripts still comprise the majority of our RNA harvest with the infection parameters we used. We also find that the mRNA yields from standard ribodepletion approaches are quite low, meaning that background levels of contaminating DNA nucleosides comprise a significant fraction of the total nucleoside content.

We anticipate that by scaling up the RNA depletion reaction, we could increase the mRNA signal to a level significantly above the DNA background. We also anticipate that we could increase the sensitivity of our assay by using LC-MS/MS instead of HPLC to detect modified RNA nucleosides.

An idea for scaling up depletion of bacterial transcripts

Based on the cost and low yield of using an rRNA depletion kit, as well as the need to deplete bacterial mRNA, we brainstormed methods for an in-house bacterial RNA depletion protocol. While we don't intend to continue this project, we're sharing these ideas in case others working in this area would like to pursue them.

We first considered designing our own DNA probes against the entire host transcriptome, which would allow us to deplete all bacterial RNAs and also scale up the depletion reaction to capture higher amounts of phage mRNA. However, we wanted to analyze mRNA chemistry from a range of phages with different hosts, and generating multiple sets of custom probes would be prohibitively expensive. This would also require us to have sequenced the genome or the transcriptome of the bacterial host ahead of time.

Instead, we think it could be possible to generate host-specific depletion probes in-house, without sequencing the genome. Our idea is to harvest total RNA from an uninfected host strain, use reverse transcriptase to generate cDNA complementary to the RNA, and then use RNaseH to digest away all RNA. One could then use the resulting cDNAs as a host-specific DNA probe set to facilitate

RNaseH digestion of bacterial transcripts. While this approach would not remove host transcripts that are uniquely present during phage infection, we anticipate that it could substantially cut down on the bacterial signal. Also, by generating probe sets in-house, one could scale up rRNA depletion reactions to obtain higher yields of phage mRNAs for analysis.

Acknowledgements

Thank you to Novogene Corporation Inc. for RNA QC, library prep, and sequencing.

Contributors (A-Z)

- **Januka Athukoralage:** Formal Analysis, Investigation, Methodology, Visualization, Writing
- **Adair L. Borges:** Conceptualization, Editing, Formal Analysis, Methodology, Supervision, Visualization
- **Megan L. Hochstrasser:** Editing, Visualization
- **Taylor Reiter:** Formal Analysis, Visualization

References

1. Rosenow C. (2001). Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches. <https://doi.org/10.1093/nar/29.22.e112>
2. Wolfram-Schauerte M, Pozhydaieva N, Viering M, Glatter T, Höfer K. (2022). Integrated Omics Reveal Time-Resolved Insights into T4 Phage Infection of E. coli on Proteome and Transcriptome Levels. <https://doi.org/10.3390/v14112502>
3. Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. <https://doi.org/10.48550/arxiv.1303.3997>
4. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1GPDP. (2009). The Sequence Alignment/Map format and SAMtools. <https://doi.org/10.1093/bioinformatics/btp352>
5. Liao Y, Smyth GK, Shi W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. <https://doi.org/10.1093/bioinformatics/btt656>

6. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemond G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. (2019). Welcome to the Tidyverse. <https://doi.org/10.21105/joss.01686>
7. <https://www.neb.com/products/e7850-nebnext-rrna-depletion-kit-bacteria>
8. Stewart CR, Gaslightwala I, Hinata K, Krolikowski KA, Needleman DS, Peng AS-Y, Peterman MA, Tobias A, Wei P. (1998). Genes and Regulatory Sites of the “Host-Takeover Module” in the Terminal Redundancy of *Bacillus subtilis* Bacteriophage SPO1. <https://doi.org/10.1006/viro.1998.9197>