Poster Number: MECH-FRI-550
Date: June 5, 2026

Leveraging Long-Read and Short-Read Next-Generation Sequencing Data to Produce Reference Genomes for a Diverse Collection of Bacteriophages and their Hosts

ATCC®
Credible leads to Incredible®

Noah Wax, MS; Joseph R. Petrone, PhD; Jeanette Rimbey, MS; Nikhita P. Puthuveetil, MS; Emily White, MS; David A. Yarmosh, MS; Amy L. Reese, MS; Corina Tabron, MS; Jade L. Kirkland, MS; Kaitlyn Bentley, MS; James Duncan, MS; Robert Marlow, BS; Stephen King, MS; Scott Nguyen, PhD; Ana Fernandes, BS; John Bagnoli, BS; Briana Benton, BS; Jonathan L. Jacobs, PhD
ATCC, Manassas, VA 20110

Background

Bacteriophages have applications across many scientific disciplines, spanning therapeutics to food sciences, and are notably used for detecting wastewater contamination. *Bacteroides* phages, such as *Bacteroides fragilis* phage B56-3, are commonly used to detect wastewater contamination and, in some cases, can be a more reliable indicator than *E. coli* phages.[1] Public databases can harbor incomplete or incorrect genomic sequences for bacteriophages and their hosts.[2] In 2025, ATCC® launched an initiative to sequence its phage collection, which contains over 340 unique phages spanning 204 host bacteria across 50 genera. ATCC®'s genomic data for both bacteriophages and their hosts are traceable to physical material in the ATCC® repository. Here, we demonstrate how we produced reference genomes for 74 bacteriophages and their hosts, including *Bacteroides fragilis* phage B56-3 (ATCC® 700786-B1™), which currently does not have a full genome deposited in NCBI.

Growth, DNA Extraction, and NGS

Phages and hosts were grown according to the ATCC recommended growth conditions. Phages were extracted using the EZ1&2 Virus Mini Kit v2.0 (QIAGEN). Nucleic acid concentration and purity were determined using the Qubit fluorometer and Nanodrop spectrophotometer. Sequencing was performed on Illumina and, when nucleic acid yield supported long reads, Oxford Nanopore (ONT); low-yield samples were processed with an Illumina-only workflow. Illumina libraries were prepared with the Illumina DNA Prep kit and run on a NextSeq 2000; ONT phage libraries used the Native Barcoding Kit 96 V14 on an R10 flow cell.

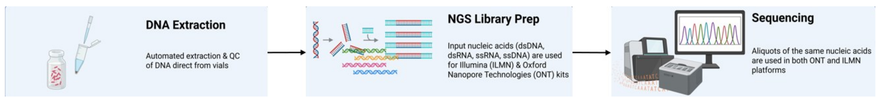

Figure 1: ATCC's standardized process for phage nucleic acid extraction through sequencing. Created with BioRender.com

Bioinformatic Analysis

Once sequenced, if only Illumina short-read data was available, genomes were assembled using SPAdes v4.0.0 with the Illumina short-read-only option. If both Illumina and ONT long-read data were available, genomes were assembled via a hybrid assembly approach utilizing either long-read-first (Autocycler v0.5.2 or Flye v2.9.5-b1801) or a scaffold closing-approach with SPAdes v4.0.0. The final hybrid assembly approach used was dependent on the quality of the final genome. Due to the diversity of phages sequenced they often required multiple assembly attempts to produce a publication quality genome. Together, this integrated sequencing and assembly workflow generated high-quality reference genomes, establishing a reproducible framework for bacteriophage genomics and downstream applications.

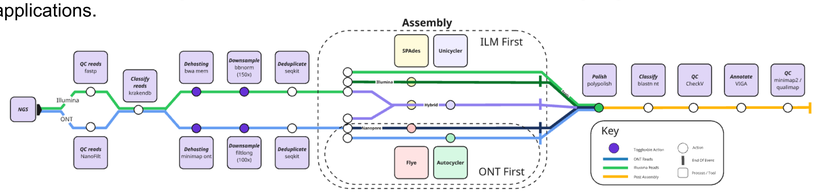

Figure 2: Bioinformatic pipeline for assembly of phage genomes. The green line represents Illumina reads. The light blue line represents ONT reads. Blue circles represent actions that can be toggled on or off. Other colored circles correspond to the color of the bioinformatic tool used (yellow: Spades; light red: Flye; light purple: Unicycler; and light green: Autocycler). The dark green line represents the workflow for an assembly utilizing only Illumina reads, light purple represents the workflow for an Illumina/ONT hybrid assembly, and the dark blue line shows the workflow for a long-read first assembly approach.

Results

Table 2: Novel bacteriophage genomes on the ATCC® Genome Portal.

ATCC® Item Number	Organism Name	Genome Size
700786-B1™	*Bacteroides fragilis* bacteriophage B56-3	47.4 Kb
15261-B3™	*Asticcacaulis excentricus* bacteriophage Ac 24	64.0 Kb

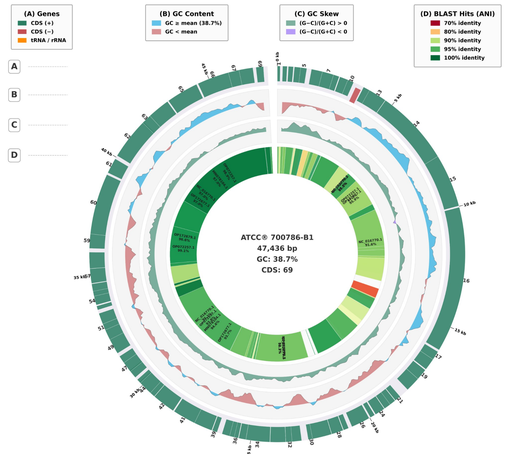

Figure 3. Circos plot for ATCC® 700786-B1™. (A) CDS annotations. (B) GC content. (C) GC skew. (D) ANI to top blast hits.

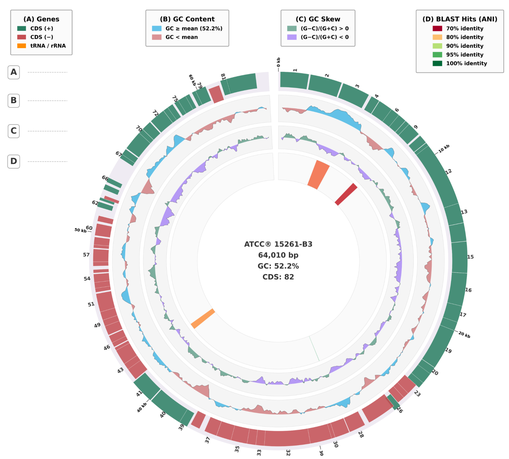

Figure 4. Circos plot for ATCC® 15261-B3™. (A) CDS annotations. (B) GC content. (C) GC skew. (D) ANI to top blast hits.

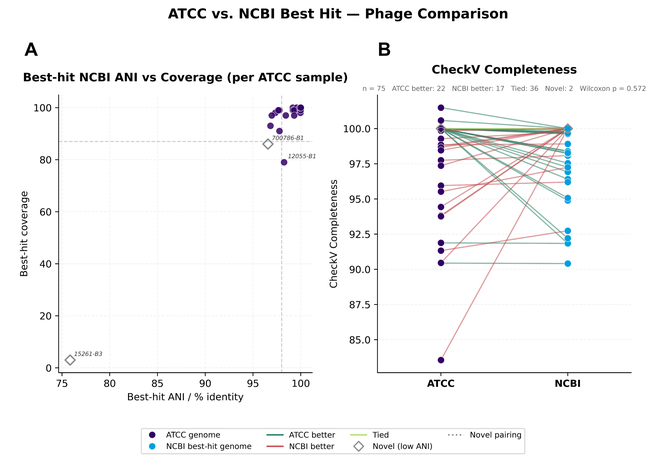

Figure 5: Comparison of best NCBI blast hit coverage (Y-axis) and best NCBI blast hit average nucleotide identity (ANI) (X-axis) of ATCC bacteriophage reference genomes. (A) Comparison of CheckV completeness of ATCC® bacteriophage reference genomes vs best NCBI blast hit. (B) Two genomes; ATCC® 11303-B26™ and ATCC® 19950-B1™ are shown as having a greater than 100% completeness due the presence of multiple contigs.

Conclusions

- Sequencing the ATCC® phage collection has allowed us to produce high-quality, authenticated reference genomes that are not available in NCBI.
- Genomic data for both phages and their bacterial hosts are traceable to physical materials in the ATCC® repository.
- Providing reference genomes for both phages and their hosts could provide novel insights into phage biology and potential predator prey interactions.
- Quarterly microbial genome releases ensure that the data on the AGP continues to expand, and taxonomic classifications/naming conventions are continually updated.

Want to accession a phage with us? Scan here to learn how!

Learn more about the ATCC® Genome Portal

References

1. McLaughlin MR, Rose JB. Application of *Bacteroides fragilis* phage as an alternative indicator of sewage pollution in Tampa Bay, Florida. Estuaries and Coasts 29(2): 246–256, 2006.
2. Yarmosh DA, et al. Comparative Analysis and Data Provenance for 1,113 Bacterial Genome Assemblies. mSphere 7(3): e0007722, 2022.