Complete Genome Sequence of *Pseudomonas aeruginosa* Phage UF_RH6, Isolated from Human Lung

Abdolrazagh Hashemi Shahraki,ᵃ Majid Vahed,ᵃ Shahla Masood,ᵇ Mehdi Mirsaeidiᵃ

ᵃDivision of Pulmonary, Critical Care, and Sleep Medicine, College of Medicine-Jacksonville, University of Florida, Jacksonville, Florida, USA
ᵇDepartment of Pathology and Laboratory Medicine, College of Medicine-Jacksonville, University of Florida, Jacksonville, Florida, USA

ABSTRACT We report the genome sequence of a lytic phage named UF_RH6, which infects *Pseudomonas aeruginosa*. This phage was isolated from a respiratory secretion sample from a patient with pulmonary *P. aeruginosa*. UF_RH6 belongs to the family *Caudoviricetes* and the genus *Samunavirus*. Its genome is 94,715 bp in length and encodes 130 proteins.

Bacteriophages have been identified in environments that are conducive to bacterial survival, such as soil, water, and wastewater, as well as animal and human tissues (1, 2). We obtained respiratory secretions from a 65-year-old male patient who was diagnosed with *Pseudomonas aeruginosa* pneumonia and was admitted to the University of Florida hospital (University of Florida Health, Jacksonville, FL, USA). The collection of human leftover samples for phage isolation was granted ethical approval by the University of Florida institutional review boards (approval number IRB202102636). A respiratory secretion sample (1 mL) was mixed with SM buffer (9 mL; Thermo Fisher Scientific, USA) and filtered through a 0.2-μm syringe filter. Then, 10 μL of filtered sample was added to 400 μL of *P. aeruginosa* (strain DJ06) cultured in brain heart infusion broth and was incubated at 37°C for 20 min. The phage was isolated using double-layer agar by incubating the plates at 37°C for 24 h (3). The phage was purified via single plaque isolation. DNA was extracted from phage lysate (5 mL) using the QIAamp MinElute Virus kit (Qiagen, USA). The Illumina Nextera XT library preparation kit was used for DNA library preparation, and sequencing was carried out using an Illumina NovaSeq 6000 system (paired-end 150-cycle mode). Bcl2fastq v2.20 (Illumina) was utilized to demultiplex reads, and Cutadapt v2.8 was used to remove sequencing adaptors and low-quality bases (4). Using the read mapper of the STAR package, *P. aeruginosa* DNA was removed from the data (5). The unmapped paired-end reads were then assembled using MetaWRAP v1.2.00 (6), and the resulting consensus sequences with lengths of >5,000 bp were evaluated by QUAST v5.0.2 (7). Centrifuge v1.04b was utilized to analyze the assembled consensus sequences (8). CheckV v1.01 was applied to evaluate the viral genome completeness and to identify closed genomes (9). The taxonomic identity of the virus was characterized by NCBI BLASTn (10). PhageTerm was used to determine the phage termini (11). VICTOR was used for phylogenetic analysis (12). GeneMarkS was used to identify open reading frames (ORFs) (13). The genome was annotated based on PHASTER (14) and BLASTp (10) results, and tRNA sequences were determined by tRNAscan-SE (15). ResFinder v4.0 (16) was used to detect virulence factors, and the Antibiotic Resistance Genes Database (ARDB) (17) was used to detect antibiotic resistance factors. We used default parameters for all software.

We obtained 24,108,049 raw reads (150-bp read length) for the sequenced sample. After removal of host DNA, only one contig with a length of >5,000 bp (94,715 bp) was assembled from the remaining reads (15.8% [3,809,071 reads]), with coverage of 10,768×. The genomic structure of *Pseudomonas* phage UF_RH6 (GenBank accession

TABLE 1 Genome sequence coverage and nucleotide identity of UF_RH5 with respect to its closest relatives

| Phage[a]	Sequence coverage (%)/identity (%) with respect to phage:					
	UF_RH6 (GenBank accession no. OQ383211.1)	SM1 (GenBank accession no. NC_041877.1)	vB_PaeS_FBPa45 (GenBank accession no. ON857939.1)	Pa BHU-15 (GenBank accession no. OL473897.1)	Pa BHU-17 (GenBank accession no. OL622073.1)	BHU-1 (GenBank accession no. OL343755.1)
UF_RH6 (GenBank accession no. OQ383211.1)	100/100	97/99.08	97/98.8	88/96.8	87/96.6	87/97.1
SM1 (GenBank accession no. NC_041877.1)	97/99.08	100/100	99/98.58	90/98.5	89/98.39	90/98.46
vB_PaeS_FBPa45 (GenBank accession no. ON857939.1)	97/98.8	99/98.58	100/100	90/98.1	89/98.39	90/98.99
Pa BHU-15 (GenBank accession no. OL473897.1)	88/96.8	90/98.5	90/98.1	100/100	93/100	93/100
Pa BHU-17 (GenBank accession no. OL622073.1)	87/96.6	89/98.39	89/98.39	93/100	100/100	93/100
BHU-1 (GenBank accession no. OL343755.1)	87/97.1	90/98.46	90/98.99	93/100	93/100	100/100

[a] All phages are classified as *Pseudomonas* phages.

number OQ383211.1) is composed of linear double-stranded DNA, spanning a length of 94,715 bp and exhibiting a GC content of 55.30%. PhageTerm predicted a circularly permuted genome for UF_RH6. CheckV results showed the completeness of the sequence. The genome comprises 130 ORFs and belongs to the family *Caudoviricetes* and the genus *Samunavirus*, as evidenced by similarities to other members (Table 1 and Fig. 1). UF_RH6 shows the greatest nucleotide identity (99.08%) to *Pseudomonas* phage SM1 (GenBank accession number NC_041877.1). The genome of UF_RH6 contains a tRNA sequence; however, no virulence or antibiotic resistance genes were detected.

Data availability. The complete phage genome sequence was deposited in GenBank under accession number OQ383211.1. The raw data are available in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA941099, SRA accession number SRR23702725, and BioSample accession number SAMN33589860.
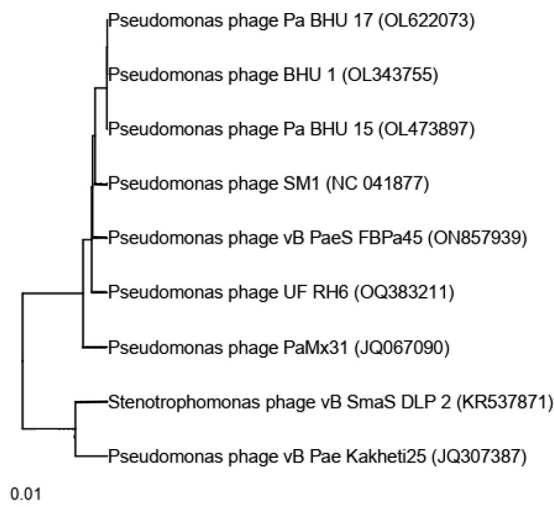
FIG 1 Phylogenetic tree illustrating the taxonomic relationship of UF_RH6 to closely related phages belonging to the genus *Samunavirus* (SM1, vB_PaeS_FBPa45, BHU-1, Pa BHU-15, Pa BHU-17, and PaMx31). Two other phages (*Stenotrophomonas* phage vB_SmaS_DLP_2 and *Pseudomonas* phage vB_Pae_Kakheti25) are *Pseudomonas* phages belonging to the genus *Septimatrevirus*. The tree was generated using the genome BLAST distance phylogeny (GBDP) method. The scale bar indicates the number of substitutions per site.

Interdisciplinary Center for Biotechnology Research (ICBR) Bioinformatics Core Facility, especially Fahong Yu.

REFERENCES

1. Clokie MRJ, Millard AD, Letarov AV, Heaphy S. 2011. Phages in nature. Bacteriophage 1:31–45. https://doi.org/10.4161/bact.1.1.14942.
2. Alexander J, Navas D, Flowers M, Charles A, Carr A. 2020. Isolation of lytic bacteriophages with broad host range activity against *Pseudomonas aeruginosa* strains isolated from respiratory samples from cystic fibrosis patients intended for therapeutic application. Open Forum Infect Dis 7(Suppl 1): S801. https://doi.org/10.1093/ofid/ofaa439.1795.
3. Jiang Y-H, Liu J-Q, Zhao C-Y, Yu S, Sun Y-B, Shi H-Y, Huang H-L. 2020. Isolation and genome sequencing of a novel *Pseudomonas aeruginosa* phage PA-YS35. Curr Microbiol 77:123–128. https://doi.org/10.1007/s00284-019-01792-8.
4. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
5. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635.
6. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP: a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:158. https://doi.org/10.1186/s40168-018-0541-1.
7. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
8. Kim D, Song L, Breitwieser FP, Salzberg SL. 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res 26:1721–1729. https://doi.org/10.1101/gr.210641.116.
9. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol 39:578–585. https://doi.org/10.1038/s41587-020-00774-7.
10. Donkor ES, Dayie NTKD, Adiku TK. 2014. Bioinformatics with basic local alignment search tool (BLAST) and fast alignment (FASTA). J Bioinform Seq Anal 6:1–6. https://doi.org/10.5897/IJBC2013.0086.
11. Garneau JR, Depardieu F, Fortier L-C, Bikard D, Monot M. 2017. Phage-Term: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci Rep 7: 8292. https://doi.org/10.1038/s41598-017-07910-5.
12. Meier-Kolthoff JP, Göker M. 2017. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33:3396–3404. https://doi.org/10.1093/bioinformatics/btx440.
13. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes: implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29: 2607–2618. https://doi.org/10.1093/nar/29.12.2607.
14. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.
15. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413.
16. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother 75:3491–3500. https://doi.org/10.1093/jac/dkaa345.
17. Liu B, Pop M. 2009. ARDB: Antibiotic Resistance Genes Database. Nucleic Acids Res 37(Suppl 1):D443–D447. https://doi.org/10.1093/nar/gkn656.