Complete Genome Sequences of *Vibrio cholerae*-Specific Bacteriophages 24 and X29

Sudhakar G. Bhandare,a Andrew Warry,a,b Richard D. Emes,a,b
Steven P. T. Hooton,c Paul A. Barrow,a Robert J. Atterburya

University of Nottingham, School of Veterinary Medicine and Science, Sutton Bonington, United Kingdoma; University of Nottingham, Advanced Data Analysis Centre, Sutton Bonington, United Kingdomb; University of Nottingham, School of Biosciences, Division of Food Science, Sutton Bonington, United Kingdomc

ABSTRACT The complete genomes of two *Vibrio cholerae* bacteriophages of potential interest for cholera bacteriophage (phage) therapy were sequenced and annotated. The genome size of phage 24 is 44,395 bp encoding 71 putative proteins, and that of phage X29 is 41,569 bp encoding 68 putative proteins.

Genome sequencing of bacteriophage therapy candidates is required to determine whether they carry genes related to virulence, antimicrobial resistance, and/or lysogeny. Two phages which infect *Vibrio cholerae* (24 and X29) were obtained for sequencing from the Felix d'Herelle Reference Centre for Bacterial Viruses, Canada, prior to initiating phage therapy studies. Phage 24 was isolated from sewage in Bangladesh (1), while phage X29 was isolated from cholera stools in India (2). Both of these phages belong to the *Myoviridae* family. The measurements of head diameter and tail length for these phages are 64 nm and 79 nm (phage 24) and 64 nm and 142 nm (phage X29), respectively (3).

The genomic DNA of these phages was extracted using the Promega DNA extraction kit (A7280 Wizard DNA cleanup system) (4). The DNA samples were prepared according to the Nextera DNA sample preparation guide and sequenced on an Illumina MiSeq platform (2 × 250-bp paired-end reads with 120× coverage). Complete genome sequences were assembled using SPAdes version 3.1.0 (5). Genome annotation was performed using the Rapid Annotations using Subsystems Technology (RAST) server (6) with some additional manual curation. The translated sequence of each predicted gene was compared to known proteins using BLASTp analysis (7), and conserved protein motifs were identified using the Pfam database (8). The genomes were scanned for tRNAs using tRNAscan-SE (9) and ARAGORN (10). The phage 24 genome was found to be circularly permuted with terminal redundancy, in contrast to phage X29, which has defined ends indicated by steep drop-offs in read coverage at the contig ends.

The genome size of phage 24 is 44,395 bp with a GC content of 45.4%. It contains 71 predicted coding DNA sequences (CDSs), of which 19 could be assigned putative functions. Phage 24 has 99% identity with *Vibrio* phage CP-T1 (11), the major difference being an in-frame 75-bp deletion in the pentapeptide repeat-containing protein annotated in phage 24. The conserved proteins identified can be grouped as structural head proteins, packaging proteins, phage tail proteins, and DNA replication proteins. The head proteins were a putative head protein, a major capsid protein, and a phage portal protein. Packaging proteins were small- and large-terminase subunits. The phage tail proteins were a tail fiber/lysozyme protein and a baseplate protein. The DNA replication proteins were a helicase, a primase, and a DNA polymerase.

Address correspondence to Robert J. Atterbury, robert.atterbury@nottingham.ac.uk.

S.G.B. and A.W. contributed equally to this work.

Phage X29 has a genome size of 41,569 bp with a GC content of 46% containing 68 CDSs; predicted functions could be assigned to 36 of them. Phage X29 shares 99% identity with *Vibrio* phage Phi 2 (KJ545483). It contains a conserved translational frameshift in the coding sequences located upstream of its tape measure protein that are analogous to the lambda proteins gpG and gpGT (12). The genome codes for phage structural proteins and several DNA-binding proteins with helix-turn-helix domains. The tail fiber and shaft proteins were a tail length tape measure protein, a putative phage tail protein, a tail fiber assembly-like protein, and a tail fiber protein. The coat proteins were a major capsid protein, a head maturation protease, and a portal protein. Lysogeny-related proteins were phage integrases and a Cro protein homolog. Analysis of both phage genomes revealed an absence of tRNAs.

Accession number(s). This whole-genome shotgun project has been deposited at GenBank under the accession no. KJ572844 (phage 24) and KJ572845 (phage X29). The versions described in this paper are the second versions, KJ572844.2 and KJ572845.2, respectively.

REFERENCES

1. Lee JV, Furniss AL (ed). 1981. The phage typing of *Vibrio cholerae* serovar, p 119–122. *In* Holme T, Holmgren J, Merson MH, Mollby R (ed), Acute enteric infection in children. Elsevier North-Holland, Biomedical Press, Amsterdam.
2. Nicolle P, Gallut J, Ducrest P, Quiniou J. 1962. Recherches en vue d'établir une lysotypie pour les vibrions cholériques et les vibrions d'El-Tor. Rev Hyg Med Soc 10:91–126.
3. Ackermann H-W, Furniss AL, Kasatiya SS, Lee JV, Mbiguino A, Newman FS, Takeya K, Vieu J-F. 1983. Morphology of *Vibrio cholerae* typing phages. Ann Virol 134:387–404. https://doi.org/10.1016/S0769-2617(83)80011-2.
4. Summer EJ. 2009. Preparation of a phage DNA fragment library for whole genome shotgun sequencing. Methods Mol Biol 502:27–46. https://doi.org/10.1007/978-1-60327-565-1_4.
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
6. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
8. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. Nucleic Acids Res 40:D290–D301. https://doi.org/10.1093/nar/gkr1065.
9. Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33:W686–W689. https://doi.org/10.1093/nar/gki366.
10. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.
11. Comeau AM, Tremblay D, Moineau S, Rattei T, Kushkina AI, Tovkach FI, Krisch HM, Ackermann HW. 2012. Phage morphology recapitulates phylogeny: the comparative genomics of a new group of myoviruses. PLoS One 7:e40102. https://doi.org/10.1371/journal.pone.0040102.
12. Xu J, Hendrix RW, Duda RL. 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol Cell 16:11–21. https://doi.org/10.1016/j.molcel.2004.09.006.