Complete Genome Sequence of *Caulobacter crescentus* Siphophage Sansa

Leonardo Vara, Ashley A. Kane, Jesse L. Cahill, Eric S. Rasche, Gabriel F. Kuty Everett

Center for Phage Technology, Texas A&M University, College Station, Texas, USA

***Caulobacter crescentus* is a Gram-negative dimorphic model organism used to study cell differentiation. Siphophage Sansa is a newly isolated siphophage with an icosahedral capsid that infects *C. crescentus*. Sansa shares no sequence similarity to other phages deposited in GenBank. Here, we describe its genome sequence and general features.**

Address correspondence to Gabriel F. Kuty Everett, cpt@tamu.edu.

Caulobacter crescentus is an aquatic oligotrophic bacterium capable of growing in low-nutrient environments (1). *C. crescentus* forms two types of cells, stalk and swarmer cells, and is a model organism for cell differentiation (2). Bacteriophage Sansa was isolated on *C. crescentus* strain CB15. Bacteriophages have long been a tool used in *C. crescentus* research (3), and Sansa may be useful to further study this model organism.

Bacteriophage Sansa was isolated from a water sample collected in Houston, TX. Phage DNA was sequenced using 454 pyrosequencing at the Emory GRA Genome Center (Emory University, GA). Trimmed FLX Titanium reads were assembled to a single contig of circular assembly at 24.9-fold coverage using the Newbler assembler, version 2.5.3 (454 Life Sciences), with the default settings. The contig was confirmed to be complete by PCR using primers that face the upstream and downstream ends of the contig. Products from the PCR amplification of the junctions of concatemeric molecules were sequenced by Sanger sequencing (Eton Bioscience, San Diego, CA). Genes were predicted using GeneMarkS (4) and corrected using software tools available on the Center for Phage Technology (CPT) Galaxy instance (https://cpt.tamu.edu/galaxy-public/). Morphology was determined using transmission electron microscopy performed at the Texas A&M University Microscopy and Imaging Center.

Sansa is a siphophage with a double-stranded DNA (dsDNA) unit genome of 78,170 bp, a coding density of 95.4%, and a G+C content of 65.8%. A 2,127-bp repeat was determined using the PAUSE method on raw sequencing data (https://cpt.tamu.edu/pause/). The repeat contains 7 coding sequences resulting in a final packaged genome of 80,297 bp with 122 coding sequences. Of the 122 coding sequences, 115 are unique (does not include duplicate coding sequences in the terminal repeat). Thirty-six coding sequences are homologous to proteins of known function, as determined by BLASTp and InterProScan (5, 6). Sansa shows no similarity to other phages in the GenBank database, as determined by BLASTn analysis (7).

Genes encoding proteins involved in DNA replication and recombination, DNA packaging, morphogenesis, and lysis were identified. Genes encoding DNA replication and recombination proteins include those for primase, polymerase III, several DNA-binding proteins, single-stranded DNA (ssDNA)-annealing protein, nucleases, and helicase. The genes encoding the ssDNA-binding protein and the DNA primase share 77% and 76% nucleotide sequence identity with the corresponding regions on the CB15 and K31 genomes, respectively. DNA packaging genes encoding the portal protein, large and small terminase, and a DNA packaging protein were annotated. The DNA packaging protein is in the Terminase_3 superfamily (accession no. cl12054), as determined by CD Search (8). The structural proteins identified include the major capsid, tail, tailspike, and tape measure proteins. Additional structural genes were not identified, likely due to a lack of sequence similarity to other phages. Sansa encodes several methylase proteins, including an adenine-specific DNA methyl-transferase. A lysis cassette was identified that contains a putative holin/antiholin pair, endolysin, and partially embedded inner- and outer-membrane spanins (9, 10).

Nucleotide sequence accession number. The genome sequence of phage Sansa was contributed as accession no. KT001913 to GenBank.

ACKNOWLEDGMENTS

REFERENCES

1. **Curtis PD, Brun YV**. 2010. Getting in the loop: regulation of development in *Caulobacter crescentus*. Microbiol Mol Biol Rev 74:13–41. http://dx.doi.org/10.1128/MMBR.00040-09.
2. **Nierman WC, Feldblyum TV, Laub MT, Paulsen IT, Nelson KE, Eisen JA, Heidelberg JF, Alley MR, Ohta N, Maddock JR, Potocka I, Nelson WC, Newton A, Stephens C, Phadke ND, Ely B, DeBoy RT, Dodson RJ, Durkin AS, Gwinn ML, Haft DH, Kolonay JF, Smit J, Craven MB,**

Khouri H, Shetty J, Berry K, Utterback T, Tran K, Wolf A, Vamathevan J, Ermolaeva M, White O, Salzberg SL, Venter JC, Shapiro L, Fraser CM. 2001. Complete genome sequence of *Caulobacter crescentus*. Proc Natl Acad Sci U S A **98**:4136–4141. http://dx.doi.org/10.1073/pnas.061029298.

3. Bender RA. 1981. Improved generalized transducing bacteriophage for *Caulobacter crescentus*. J Bacteriol **148**:734–735.

4. 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. http://dx.doi.org/10.1093/nar/29.12.2607.

5. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**:421. http://dx.doi.org/10.1186/1471-2105-10-421.

6. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. 2009. InterPro: the integrative protein signature database. Nucleic Acids Res **37**:D211–D215. http://dx.doi.org/10.1093/nar/gkn785.

7. Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol **7**:203–214. http://dx.doi.org/10.1089/10665270050081478.

8. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res **39**:D225–D229. http://dx.doi.org/10.1093/nar/gkq1189.

9. Wang IN, Smith DL, Young R. 2000. Holins: the protein clocks of bacteriophage infections. Annu Rev Microbiol **54**:799–825. http://dx.doi.org/10.1146/annurev.micro.54.1.799.

10. Summer EJ, Berry J, Tran TA, Niu L, Struck DK, Young R. 2007. Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. J Mol Biol **373**:1098–1112. http://dx.doi.org/10.1016/j.jmb.2007.08.045.