Complete Genome Sequence of *Streptomyces* Phage Salutena

Jinha Kim, Tyler Higbee, James Clark, Tram Le, Mei Liu, Ben Burrowes

Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, Texas, USA

Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas, USA

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

**ABSTRACT** *Streptomyces* are Gram-negative soil bacteria that can degrade lignin and synthesize antibiotics. Some species cause mycetoma, pneumonitis, and bloodstream infections. Here, we present the genome sequence of the *Streptomyces* sp. strain Mg1 phage Salutena, a siphovirus in the subfamily *Arquatrovirinae*. The genome is 51,993 bp, with 90 predicted protein-coding genes.

*Streptomyces* spp. are Gram-positive, saprotrophic soil bacteria that can degrade lignin and synthesize industrial enzymes and unique antibiotics (1). *Streptomyces* sp. strain Mg1 secretes the antibiotic chalcomycin A, which can degrade *Bacillus subtilis* colonies (2). Furthermore, certain *Streptomyces* spp. can cause mycetoma and, in rare cases, pneumonitis and bloodstream infections (3). Here, we describe a novel phage, Salutena, that infects *Streptomyces* sp. Mg1.

Salutena was isolated from a South Jordan, UT, soil sample taken in August 2019 using the double-overlay agar technique (4). *Streptomyces* sp. Mg1 (provided by Paul Straight, Texas A&M University) was used as the host and grown on nutrient broth or agar at 30°C with 10 mM MgCl₂, 8 mM Ca(NO₃)₂, and 0.5% glucose. Genomic DNA was purified as previously described (5) using a Wizard DNA cleanup kit (Promega). A paired-end sequence library was prepared with 300-bp inserts using the TruSeq Nano kit and was sequenced by an Illumina iSeq 100 instrument. A total of 367,310 reads were visualized (www.bioinformatics.babraham.ac.uk/projects/fastqc), manually trimmed with FastX Toolkit 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/download.html), and assembled by SPAdes v3.5.0 with 86.8-fold contig coverage (6). A paired-end sequence library was prepared with 300-bp inserts using the TruSeq Nano kit and was sequenced by an Illumina iSeq 100 instrument. A total of 367,310 reads were visualized (www.bioinformatics.babraham.ac.uk/projects/fastqc), manually trimmed with FastX Toolkit 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/download.html), and assembled by SPAdes v3.5.0 with 86.8-fold contig coverage (6). Genome closure was confirmed by PCR and Sanger sequencing (forward primer, 5′-GATGTTCGTGGCGTTCAC-3′; reverse primer, 5′-ATCTTGAGCTGGCCGTAC-3′). Initial gene prediction was performed with GLIMMER v3 (7) and MetaGeneAnnotator v1.0 (8). The tRNA and rho-independent terminators were detected with progressiveMauve v2.4 (15). All annotation tools were hosted on the Galaxy platform by the Center for Phage Technology (https://cpt.tamu.edu/galaxy) (16). HHpred v3.2.0 was used for validating functional annotation based on tertiary structure predictions of translated proteins (https://toolkit.tuebingen.mpg.de/tools/hhpred) (17). Default parameters were used for all software unless otherwise specified. After negative staining of the sample with 2% (wt/vol) uranyl acetate, phage morphology was evaluated by transmission electron microscopy (TEM) at the Texas A&M University Microscopy and Imaging Center, and the phage was determined to be a siphovirus (data not shown).

The Salutena genome is 51,993 bp, with a G+C content of 67.50% and a coding density of 92.68%. Based on the annotation results, 90 protein-coding genes were
identified, of which 38 were assigned putative functions. Salutena shared highest nucleotide similarity (71.62%) with Streptomyces platenis MU1A1 phage BartholomewSD (GenBank accession no. MK460245.1). Streptomyces azureus NRRL B-5410 phage Omar (MG593802.1) had the highest number of protein matches (77 proteins). Salutena is therefore a siphovirus of the subfamily Arquatovirinae.

Only 1 trNA and 1 rho-independent terminator were identified. Genes involved in phage morphogenesis, DNA packaging and replication, lysis, and transcription were also identified. A tape measure protein gene was found with tail assembly chaperone genes that are generated with a translational frameshift (18). One lysis protein gene, amidase endolysin, was identified. A GCNS-related N-acetyltransferase gene was identified by protein sequence (BLASTp) and predicted structural (HHpred) homology, indicating a potential regulatory posttranslational modification system through acetylation (19). No introns were identified. None of the assigned gene functions have been verified experimentally.

Data availability. The genome of Salutena has been deposited in GenBank under accession number MT708548.1. The associated BioProject, SRA, and BioSample accession numbers are PRJNA222858, SRR11558352, and SAMN14609629, respectively.

ACKNOWLEDGMENTS
This work was supported by funding from the National Science Foundation (awards EF-0949351 and DBI-1565146). Additional support came from the Center for Phage Technology (CPT).

We thank Paul Straight, Texas A&M University, for providing Streptomyces sp. Mg1.

This announcement was prepared in partial fulfillment of the requirements for BICH464 Phage Genomics, an undergraduate course at Texas A&M University.

REFERENCES