Genome Sequence of *Bacillus* Phage Saddex

Elizabeth Greguske  
*Saint Anselm College*

Alexis Nadeau  
*Saint Anselm College*

Emily Fitzmeyer  
*Saint Anselm College*

Karolina Fučíková  
*Assumption College, k.fucikova@assumption.edu*

Follow this and additional works at: https://digitalcommons.assumption.edu/sciences-faculty

Part of the Life Sciences Commons

Recommended Citation


This Article is brought to you for free and open access by the Physical and Biological Sciences Department at Digital Commons @ Assumption College. It has been accepted for inclusion in Physical and Biological Sciences Department Faculty Works by an authorized administrator of Digital Commons @ Assumption College. For more information, please contact digitalcommons@assumption.edu.
Genome Sequence of Bacillus Phage Saddex

Elizabeth Greguske,a Alexis Nadeau,a Emily Fitzmeyer,a* Karolina Fucikovab

aDepartment of Biology, Saint Anselm College, Manchester, New Hampshire, USA
bDepartment of Natural Science, Assumption College, Worcester, Massachusetts, USA

ABSTRACT The complete genome of Bacillus phage Saddex was determined and annotated in this study. Saddex has distinct sections with similarities to other Bacillus phages, such as Kida, even though these phages were isolated more than 800 km apart by separate laboratories.

Saddex is a novel Bacillus bacteriophage isolated, characterized, and annotated by students in the Howard Hughes Medical Institute (HHMI) Phage Hunters program. Saddex is able to lyse multiple Bacillus host species, including Bacillus cereus, which is of particular interest, as this bacterium lives in the gut of poultry as well as in soil and causes an estimated 2% of all cases of food poisoning (1).

Saddex was isolated from lawn soil sampled (at coordinates N42.791838, W71.069913) in Haverhill, Massachusetts. In brief, log phase Bacillus thuringiensis subspecies kurstaki cells were mixed with lawn soil and allowed to grow overnight while being shaken at 37°C. Phage were isolated via centrifugation for 10 min at 3,000 rpm, followed by filtering with a 0.22-μm sterile syringe filter (2). DNA was isolated from purified phage with the Qiagen viral DSP spin kit version 1 and sequenced via the HiSeq 2500 Illumina platform at the Hubbard Center for Genome Studies (University of New Hampshire, Durham, NH), resulting in 779,058 paired-end reads and an average 250-bp read length. Reads were trimmed with Trimmomatic (3), assembled into contigs with QUAST (4), and then refined with Geneious version 10.2 (5) reference assemblies with custom low sensitivity (allowing only 2% mismatches for precise mapping). The average depth of coverage was 539.4× with no areas of poor coverage noted. Saddex was autoannotated in Geneious version 10.2 with default settings, with a known Bacillus phage genome, Evoli (GenBank accession number KJ489398), for comparison. The genome was then visually cross-checked against 11 other complete Bacillus phage genomes available in GenBank.

The complete genome of Saddex is 142,353 bp of linear, double-stranded DNA with a G+C content of 39.0%. All genes in Saddex were found to have at least one homolog in other published Bacillus phages (2, 6–10) with a BLAST nucleotide (BLASTn) analysis (11), indicating that no novel genes were identified. However, the nucleotide sequence identity similarity of these genes ranged from as low as 63% to as high as 99%, with an average 81.9% similarity; this shows that Saddex could have unique polymorphisms. Of the 208 predicted proteins, 54 were assigned a function, typically for tail and capsid structure, nuclease activity, and lytic activity. Three tRNAs were identified, all three of which were found to have 100% similarity to at least one other Bacillus phage with a BLASTn analysis (11).

The proposed Bacillus phage cluster guidelines (12), in which pairwise average nucleotide identity across the genome is used to group similar phage, place Saddex in the C1 cluster of Bacillus phages.

Data availability. The complete genome sequence of the Bacillus phage Saddex is available in GenBank under accession number MH538193. Raw reads are available in the SRA under accession number SRP158918.
ACKNOWLEDGMENT

This research was supported by New Hampshire-INBRE through an Institutional Development Award (IDeA), number P20GM103506, from the National Institute of General Medical Sciences of the National Institutes of Health (NIH).

REFERENCES