Complete Genome Sequence of the Cluster B4 Mycobacteriophage Lolalove, Isolated in Charleston, South Carolina

Christine A. Byrum,a Hannah Marie Rozier,a Toni E. Allison,a Emilia Ballou,a Lauren Bergen,a Reilley A. Chamness,a Madison E. Davis,a Mouna S. DiBenedetto,a Nathaniel C. Elston,a Lyric A. Graham,a Keiana L. Haigh,a Tessa M. Jansen,a Gabrielle S. Kostur,a Nicholas A. Larson,a Fiona L. Lewis,a Carlo Negroni,a Isabella V. Rupert,a Isabel S. Wood,a Anastasia M. Zimmerman,a Veronique A. Delesalleb

aDepartment of Biology, College of Charleston, Charleston, South Carolina, USA
bDepartment of Biology, Gettysburg College, Gettysburg, Pennsylvania, USA

ABSTRACT

Lolalove, a B4 subcluster soil bacteriophage of *Mycobacterium smegmatis*, was isolated in Charleston, SC. It possesses a 71,111-bp linear double-stranded DNA (dsDNA) genome with 99 protein-coding genes and a GC content of 68.9%. Genome BLASTn alignments indicate high sequence identity to the related B4 subcluster *M. smegmatis* phages BrownCNA, Mithril, and Hangman.

The mycobacteriophage Lolalove was discovered in damp soil at a Shell gas station in Charleston, SC (32.787021N, 79.934143W), as part of the Howard Hughes Medical Institute (HHMI) Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) (1) effort to compare genomes of novel actinobacteriophages. This virus infects *Mycobacterium smegmatis* mc2155 and was isolated using enrichment (37°C, 48 h) followed by 3 purification and amplification cycles in 7H9 top agar (see SEA-PHAGES Phage Discovery Guide for details) (2). Transmission electron microscopy revealed that this virus has *Siphoviridae* morphology with a mean capsid diameter of 76.83 nm, tail length of 329.34 nm, and tail diameter of 12.8 nm (Fig. 1).

DNA was extracted using the Promega Wizard DNA cleanup system and a sequencing library prepared with an NEBNext Ultra II library prep kit. Pittsburgh Bacteriophage Institute sequenced the genome using Illumina MiSeq technology (v3 reagents) (3) and 609,904 single-end reads (150 bp) were collected. During assembly, ~20% of reads were from a second B4 subcluster virus. The two genomes were resolved using AceUtil (3) based on differences in contig coverage. The Lolalove reads were assembled *de novo* into a single contig using Newbler v2.9 (4) and verified using Consed v29.0 (3, 5). The Lolalove genome is 71,111 bp with 1,193× coverage. The GC content is 68.9%, and the termini are circularly permuted as the contig ends overlap (3). Base 1 in the genome was selected relative to the terminase sequence (identified using DNA Master) (6) as described by Russell (3).

After sequencing, the genome was annotated using PECAAN (7), and data were subsequently transferred to DNA Master v5.22.23 (https://phagesdb.org/DNAMaster). To identify genome features and assign putative gene functions, programs used included GLIMMER v3.0 (8), GeneMark v2.5 (9), Starterator v1.1 (https://seaphages.org/media/docs/Starterator_Guide_2016.pdf), BLASTp v2.9 (10), HHpred (11), NCBI Conserved Domain Database (12), TMHMM2 (https://www.cbs.dtu.dk/services/TMHMM), TOPCONS v2 (13), Phamerator (14), ARAGORN v1.2.38 (15), and tRNAscan-SE v3.0 (16). Default settings were used in all programs.

Lolalove is a cluster B/subcluster B4 bacteriophage whose genome contains 99 putative protein-coding genes with 31 assigned functions but no tRNAs or transfer-messenger

Citation


Editor Kenneth M. Stedman, Portland State University
Copyright © 2021 Byrum et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Christine A. Byrum, byrumc@cofc.edu.
Received 11 May 2021
Accepted 3 June 2021
Published 1 July 2021
RNAs (tmRNAs) (viruses are in the same cluster if they share >50% nucleotide sequence similarity) (17). Predicted genes for structural assembly occur on the forward strand of the left arm, and as in other B4 subcluster members, no frameshift is detected in the tail chaperone protein (gp23). Putative genes on the reverse strand include gp6, gp7, gp19, gp40 to gp43, gp46 to gp60, and gp75 to gp99.

The genome similarity of Lolalove to related viruses was evaluated using tools online at https://phagesdb.org to measure whole-genome BLASTn alignment (10) and genome content similarity scores (GCSs) (18) (Table 1). Based on percent identity, Lolalove shows the highest nucleotide similarity to the B4 subcluster genomes of BrownCNA, Mithril, Hangman, Waleliano, Zemanar, and Fortunato. These mycobacteriophages were predominately isolated from sites in the southeastern United States.

Data availability. The Lolalove virus is available at the Pittsburgh Bacteriophage Institute in freezer box 93/grid E3. The genome sequence and raw reads appear in DDBJ/ENA/GenBank under accession number MT818419.1 and SRA under accession number SRX9117730.

### TABLE 1 Comparison of the Lolalove genome to related B4 subcluster genomes

<table>
<thead>
<tr>
<th>Phage name</th>
<th>GenBank accession no.</th>
<th>Genome size (bp)</th>
<th>GC content (%)</th>
<th>Fold coverage (×)</th>
<th>Location found</th>
<th>No. of CDSs</th>
<th>% identity</th>
<th>% query coverage</th>
<th>GCS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lolalove</td>
<td>MT818419.1</td>
<td>71,111</td>
<td>68.9</td>
<td>1,193</td>
<td>Charleston, SC</td>
<td>99</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>BrownCNA</td>
<td>KT270441</td>
<td>71,214</td>
<td>68.9</td>
<td>1,433</td>
<td>Winder, GA</td>
<td>97</td>
<td>98.46</td>
<td>98</td>
<td>95.3</td>
</tr>
<tr>
<td>Mithril</td>
<td>MN369759</td>
<td>70,937</td>
<td>69.0</td>
<td>211</td>
<td>Louisville, KY</td>
<td>95</td>
<td>98.41</td>
<td>97</td>
<td>94.2</td>
</tr>
<tr>
<td>Hangman</td>
<td>MH513970</td>
<td>71,376</td>
<td>68.9</td>
<td>2,011</td>
<td>Conway, SC</td>
<td>97</td>
<td>97.87</td>
<td>97</td>
<td>97.4</td>
</tr>
<tr>
<td>Waleliano</td>
<td>MK524486</td>
<td>70,963</td>
<td>68.9</td>
<td>408</td>
<td>Catonville, MD</td>
<td>96</td>
<td>97.83</td>
<td>97</td>
<td>96.9</td>
</tr>
<tr>
<td>Zemanar</td>
<td>JF704104</td>
<td>71,092</td>
<td>68.9</td>
<td>NA</td>
<td>Fredericksburg, VA</td>
<td>95</td>
<td>97.75</td>
<td>97</td>
<td>94.2</td>
</tr>
<tr>
<td>Fortunato</td>
<td>KX589269</td>
<td>70,679</td>
<td>69.0</td>
<td>196</td>
<td>Arkadelphia, AR</td>
<td>94</td>
<td>96.89</td>
<td>96</td>
<td>90.5</td>
</tr>
</tbody>
</table>

*aCDSs, coding DNA sequences.

*bNA, not applicable.
ACKNOWLEDGMENTS

This study was generously supported by the HHMI SEA-PHAGES program as well as the Department of Biology at the College of Charleston.

Special thanks go to Nancy Smythe of the Medical University of South Carolina for her assistance with transmission electron microscopy and to College of Charleston students Stephanie Jackson and Cody Johnson for assistance in isolating Loolalove. We also thank Graham F. Hatfull, Deborah Jacobs-Sera, Daniel A. Russell, Welkin H. Pope, and Rebecca A. Garlena at the University of Pittsburgh for sequencing, quality control, and assembly of the genome.

REFERENCES


