Genome Sequence of Mycobacterium Phage CrystalP

Christine L. Fleischacker,[a] Miriam Segura-Totten,[b] SEA-PHAGES 2016 Bioinformatics Workshop,[d] Rebecca A. Garlena,[c] Deborah Jacobs-Sera,[c] Welkin H. Pope,[c] Daniel A. Russell,[c] Graham F. Hatfull[c]

Department of Biology, University of Mary, Bismarck, North Dakota, USA[a]; Department of Biology, University of North Georgia, Dahlonega, Georgia, USA[b]; Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA[c]; HHMI Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science Program, Chevy Chase, Maryland, USA[d]

ABSTRACT Mycobacteriophage CrystalP is a newly isolated phage infecting *Mycobacterium smegmatis* strain mc²155. CrystalP has a 76,483-bp genome and is predicted to contain 143 protein-coding and 2 tRNA genes, including repressor and integrase genes consistent with a temperate lifestyle. CrystalP is related to the mycobacteriophages Toto and Kostya and to other Cluster E phages.

The large collection of phages with sequenced genomes that infect the same host strain, *Mycobacterium smegmatis* mc²155, reveals substantial diversity and at least 30 separate lineages represented as clusters and singletons (1, 2). Mycobacteriophage CrystalP was isolated from soil collected from North Huntingdon, PA, using enrichment and *M. smegmatis* mc²155 as the host. CrystalP forms plaques 2 mm in diameter with a clear center and turbid halo, and examination with electron microscopy reveals that it is a member of the *Siphoviridae*, with a 70-nm-diameter capsid and a flexible, noncontractile tail that is 240 nm in length. Double-stranded DNA was isolated and sequenced using an Illumina MiSeq 150-bp single-end run. Trimmed reads were assembled using Newbler, yielding a 76,483-bp contig with 500-fold coverage, and the viral genome was determined to have 9-base 3′ single-stranded extensions (5′-CGCTTGTCA). The G+C content is 63.0%.

The CrystalP genome was annotated using DNA Master (http://cobamide2.bio.pitt.edu/), Glimmer (3), GeneMark (4), Aragorn (5), tRNAscan-SE (6), BLASTP (7), HHPred (8), and Phamerator (9). Annotation identified 145 protein-coding genes, 43 of which were assigned putative functions, and 2 tRNA genes. CrystalP is a member of Cluster E, one of the least diverse mycobacteriophage clusters, and its closest relative is mycobacteriophage Toto (GenBank accession number JN006061), with which it shares 99% nucleotide identity spanning 99% of their genome lengths.

The CrystalP genome is organized with its virion structure and assembly genes in the leftmost 30 kbp of the centrally located lysis and immunity functions and its nonstructural genes in the rightmost 35 kbp of the genome. Most of the genes are transcribed rightward, with the exceptions of a group of 14 small (<500 bp) leftward-transcribed genes centrally located between the lysis and immunity functions and a set of 16 leftward-transcribed genes near the right genome end; the putative immunity repressor (gene *55*) is also transcribed leftward. CrystalP encodes a putative recombination system with an exonuclease (gene *66*) and Erf-like DNA pairing proteins, in addition to a RecA-like protein (gene *116*). It also codes for an RNA ligase (gene *90*) and a polynucleotide kinase (gene *87*) implicated in tRNA repair, along with 2 tRNAs (tRNA^arg and tRNA^gly, genes *107* and *108*, respectively), which collectively may play roles in countering host defenses (10). CrystalP also encodes an adenine-specific DNA methyltransferase (gene *98*), which is present in most but not all Cluster E genomes as well as

in some Cluster F genomes, and which may play a role in defending against host restriction systems.

Approximately 57% of the CrystalP-predicted protein-coding genes are not present outside the group of 84 Cluster E phages; these include many of the putative virion structure and assembly genes (terminase large subunit, portal, protease, capsid, head-to-tail connectors, major tail subunit, tape measure protein, and two minor tail subunit genes). We noted that the putative immunity repressor (gene *55*) is conserved among all 84 of the Cluster E phages and these may form a homoimmune group.

Accession number(s). CrystalP is available at GenBank with accession number KY319168.

ACKNOWLEDGMENT

REFERENCES

1. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs WR, Hendrix RW, Lawrence JG, Hatfull GF; Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science; Phage Hunters Integrating Research and Education; Mycobacterial Genetics Course. 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Elife 4:e06416.
2. Russell DA, Hatfull GF. 2017. PhagesDB: the actinobacteriophage database. Bioinformatics 33:784–786. https://doi.org/10.1093/bioinformatics/btw711.
3. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27:4636–4641. https://doi.org/10.1093/nar/27.23.4636.
4. Borodovsky M, McIninch J. 1993. Recognition of genes in DNA sequence with ambiguities. Biosystems 30:161–171. https://doi.org/10.1016/0303-2647(93)90068-N.
5. 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.
6. Lowe TM, Eddy SR. 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.0955.
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. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248. https://doi.org/10.1093/nar/gki408.
9. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics 12:395. https://doi.org/10.1186/1471-2105-12-395.
10. Zhu H, Yin S, Shuman S. 2004. Characterization of polynucleotide kinase/phosphatase enzymes from mycobacteriophages omega and Cjw1 and vibriophage KVP40. J Biol Chem 279:26358–26369. https://doi.org/10.1074/jbc.M403200200.