

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

Genome Sequences of Mycobacteriophages Kerberos, Pomar16, and StarStuff

Running title: Phages Kerberos, Pomar16, and StarStuff

Deborah Jacobs-Sera¹, Oana Catinas², Mariceli Fernandez-Martinez³, Amelia Garcia⁴, Rebecca A. Garland¹, Carlos A. Guerrero Bustamante¹, Michelle H. Larsen⁵, Rosa H. Medellin⁴, Martin Melendez-Ortiz³, Crystal Y. Melendez-Rivera³, Alondra M. Mercado-Andino³, Abner K. Mercado-Delgado³, Cathia J. Ortiz-Ortiz³, Ana P. Quesada-Gordillo⁴, Jacqueline M. Ramos⁴, Michael Rubin³, Daniel R. Russell¹, Rachna A. Sadana⁴, Sanghamitra Saha⁴, Edwin Vazquez³, David Villarreal⁴, Graham F. Hatfull^{1†}

¹Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA
²Africa Health Research Institute, Durban, South Africa
³Department of Biology, University of Puerto Rico at Cayey, Cayey, PR
⁴Department of Natural Sciences, University of Houston-Downtown, Houston, TX
⁵Department of Medicine, Albert Einstein College of Medicine, New York City, NY

†Corresponding author

52 **Abstract**

53 We describe the genome sequences of three closely-related mycobacteriophages, Kerberos,
54 Pomar16, and StarStuff, isolated at similar times but from geographically distinct regions. All
55 three genomes are similar to other Subcluster A2 phages such as L5 and D29, are temperate,
56 and have siphoviral virion morphologies.

57

58

59 A large collection of sequenced mycobacteriophages – phages that infect mycobacterial hosts –
60 reveals them to be highly genetically diverse and to span a spectrum of genetic diversity (1).
61 They can be grouped into clusters (some of which are divided into subclusters) and singletons
62 according to their overall relatedness (2), and the collection of over 1,300 sequenced phages
63 currently span 26 clusters and six singletons (<http://www.phagesdb.org>). Most of these phages
64 were isolated on a single host strain (*Mycobacterium smegmatis* mc²155), and approximately
65 10% of the phages efficiently infect *Mycobacterium tuberculosis* mc²7000; for some other
66 phages, host range expansion mutants can be isolated that efficiently infect *M. tuberculosis* (3).
67 Those phages that efficiently infect *M. tuberculosis* map within Subclusters A2, A3, and all
68 subclusters within Cluster K (3). Mycobacteriophages have not only provided insights into phage
69 diversity and evolution, but have also been exploited for various tools and applications (4),
70 including the use of D29 in a rapid amplification strategy for tuberculosis diagnosis (5).

71
72 In 2015, phages Kerberos, Pomar16, and StarStuff were isolated from soil samples using *M.*
73 *smegmatis* mc²155 as a host using an enrichment procedure. The samples were collected in
74 geographically distinct regions, Kerberos from Houston, Texas, Pomar16 from Aibonito, Puerto
75 Rico, and StarStuff from Pinetown, South Africa. Following plaque purification and amplification,
76 DNA was isolated and sequenced using Illumina MiSeq 150-bp single-end runs. Trimmed reads
77 were assembled using Newbler, and single contigs were assembled. Genome lengths are
78 52,753 bp, 52,833 bp, and 52,785 and read coverage is 506, 564, and 3,354 for Kerberos,
79 Pomar16 and StarStuff respectively. All three phages have defined sticky ends with 10-base 3'
80 single stranded DNA extensions (5'- CGGTCGGTTA), and all are 63.5% G+C. Electron
81 microscopy shows that all three phages have siphoviral morphologies with icosahedral heads
82 approximately 55 nm diameter, and flexible non-contractile tails approximately 110 nm long.

83

84 All three genomes were annotated using DNA Master (<http://cobamide2.bio.pitt.edu/>), Glimmer
85 (6), GeneMark (7), Aragorn (8), tRNAscan-SE (9), BLASTP (10), HHPred (11) and Phamerator
86 (12). BlastN comparisons show that the three genomes are very closely related to each other
87 and greater than 98% nucleotide identity across their entire genome spans. Each genome
88 contains 93 protein-coding genes and five tRNA genes. Their overall genome architectures are
89 similar to other subcluster A2 phages including L5 and D29 (13, 14), with rightwards-transcribed
90 virion structure and assembly genes in the left arms, and the leftwards-transcribed non-
91 structural genes in the right arms. All encode a putative repressor protein with similarity to the
92 L5 repressor (78% amino acid identity) and the integration systems are closely related to that of
93 D29, with closely-related integrases (99% amino acid identity) and *attP* sites (98% nucleotide
94 identity) and predicted to use the same *attB* site for integration.

95

96 The most closely-related genome to Kerberos, Pomar16, and StarStuff is phage D29, which
97 was previously shown to contain a 3.6 kbp deletion when aligned to phage L5 (14). All three
98 genomes are thus likely to be very close relatives of the putative temperate parent of D29.

99

100 **Nucleotide sequence accession numbers**

101

102 Pomar16, Kerberos, and StarStuff are available at GenBank with Accession Numbers
103 KX574455, KX758538, and KX897981, respectively.

104

105 **Acknowledgements**

106

107 This research was supported by grants from the National Institutes of Health (GM116884) and
108 the Howard Hughes Medical Institute (54308198) to G.F.H, and from the University of Puerto
109 Rico at Cayey to M.R.R.

110 **References**

- 111 1. **Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs**
112 **WR, Hendrix RW, Lawrence JG, Hatfull GF, Science Education Alliance Phage**
113 **Hunters Advancing G, Evolutionary S, Phage Hunters Integrating R, Education,**
114 **Mycobacterial Genetics C.** 2015. Whole genome comparison of a large collection of
115 mycobacteriophages reveals a continuum of phage genetic diversity. *Elife* **4**:e06416.
- 116 2. **Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ,**
117 **Patel MC, Germane KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC,**
118 **Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian**
119 **AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW.**
120 2010. Comparative Genomic Analysis of 60 Mycobacteriophage Genomes: Genome
121 Clustering, Gene Acquisition, and Gene Size. *J Mol Biol* **397**:119-143.
- 122 3. **Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C,**
123 **Boyle MM, Petrova ZO, Dedrick RM, Pope WH, Science Education Alliance Phage**
124 **Hunters Advancing G, Evolutionary Science Sea-Phages P, Modlin RL, Hendrix RW,**
125 **Hatfull GF.** 2012. On the nature of mycobacteriophage diversity and host preference.
126 *Virology* **434**:187-201.
- 127 4. **Hatfull GF.** 2014. Mycobacteriophages: windows into tuberculosis. *PLoS Pathog*
128 **10**:e1003953.
- 129 5. **Wilson SM, al-Suwaidi Z, McNERNEY R, Porter J, Drobniewski F.** 1997. Evaluation of a
130 new rapid bacteriophage-based method for the drug susceptibility testing of *Mycobacterium*
131 *tuberculosis*. *Nat Med* **3**:465-468.
- 132 6. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. Improved microbial gene
133 identification with GLIMMER. *Nucleic Acids Res* **27**:4636-4641.
- 134 7. **Borodovsky M, McIninch J.** 1993. Recognition of genes in DNA sequence with
135 ambiguities. *Biosystems* **30**:161-171.

- 136 8. **Laslett D, Canback B.** 2004. ARAGORN, a program to detect tRNA genes and tmRNA
137 genes in nucleotide sequences. *Nucleic Acids Res* **32**:11-16.
- 138 9. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer
139 RNA genes in genomic sequence. *Nucleic Acids Res* **25**:955-964.
- 140 10. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment
141 search tool. *J Mol Biol* **215**:403-410.
- 142 11. **Soding J, Biegert A, Lupas AN.** 2005. The HHpred interactive server for protein homology
143 detection and structure prediction. *Nucleic Acids Res* **33**:W244-248.
- 144 12. **Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF.** 2011.
145 Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC*
146 *Bioinformatics* **12**:395.
- 147 13. **Hatfull GF, Sarkis GJ.** 1993. DNA sequence, structure and gene expression of
148 mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol Microbiol* **7**:395-
149 405.
- 150 14. **Ford ME, Sarkis GJ, Belanger AE, Hendrix RW, Hatfull GF.** 1998. Genome structure of
151 mycobacteriophage D29: implications for phage evolution. *J Mol Biol* **279**:143-164.
152