genomeA
Journals.ASM.org

Complete Genome Sequences of Nine Mycobacteriophages

Jorgelina Judith Franceschelli,[a] Cristian Alejandro Suarez,[a] Lucrecia Terán,[b] Raúl Ricardo Raya,[b] Héctor Ricardo Morbidoni[a]

Laboratorio de Microbiología Molecular, Cátedra de Microbiología, Facultad de Ciencias Médicas, Universidad Nacional de Rosario, Rosario, Argentina[a]; Centro de Referencia para Lactobacilos (CERELA-CONICET), San Miguel de Tucumán, Argentina[b]

Genome analyses of a large number of mycobacteriophages, bacterial viruses that infect members of the genus *Mycobacterium*, yielded novel enzymes and tools for the genetic manipulation of mycobacteria. We report here the complete genome sequences of nine mycobacteriophages, including a new singleton, isolated using *Mycobacterium smegmatis* mc²155 as a host strain.

Address correspondence to Héctor Ricardo Morbidoni, morbiatny@yahoo.com.

Not only have mycobacteriophages been an essential source of materials for the development of genetic tools that allow for mycobacterial genetic manipulation, but also they have proven to be a valuable resource for understanding the molecular evolution of those viruses. Until now, >3,600 mycobacteriophages have been isolated, of which about 600 genomes have been sequenced (see http://www.phagesdb.org), annotated, and characterized. The analysis of those sequences has revealed impressive phage diversity, generating 21 clusters (A·U) of which the phages belonging to the same group have at least 50% nucleotide similarity (1). Eight mycobacteriophages are considered singletons because they did not show enough sequence similarity to be part of the clusters described, suggesting that the genome diversity of this group is not exhausted (2, 3).

We isolated >40 novel mycobacteriophages from soil samples of several geographic locations in Argentina using *Mycobacterium smegmatis* mc²155 as a host; we previously reported the characterization of 18 of them. Several of these phages infected *Mycobacterium tuberculosis*, opening the possibility for their use for diagnostic purposes; at the same time, we found an uncommon mechanism for phage heritance represented by partition genes similar to those of plasmids (4). We report here the genome sequences of nine mycobacteriophages (20ES, CRB1, 40AC, Jolie1, Hosp, 39HC, 40BC, Jolie2, and 32HC), all of them belonging to the *Siphoviridae* class, with genome sizes ranging from 44 kbp to 71 kbp. Mycobacteriophage genome sequencing was performed at the Instituto de Agrobiotecnología de Rosario (IN-DEAR), Argentina, by whole-genome shotgun sequencing using a Life Sciences GS-FLX 454 sequencer. When the genomes had defined ends, these were determined by PCR and Sanger sequencing of the product, as previously reported (4). Phage genome annotation of open reading frames was performed using the DNA Master program, which includes GeneMark version 2.0, Glimmer version 3.02, and Aragorn version 1.1 (available at http://cobamide2.bio.pitt.edu), while probable function assignment was done with HHpred and Pfam (5–8). Three mycobacteriophages, 20ES, CRB1, and 40AC, belong to cluster A according to Hatfull's classification, while the four members of cluster B, Jolie1, Hosp, 39HC, and 40BC, are highly similar to

mycobacteriophage KayaCho (2), except for a few open reading frames (ORFs) not showing any homology to KayaCho. Jolie2 is a member of cluster G, displaying a slightly different G+C content (68% versus 66%), which may reflect the fact that this cluster contains few members, and therefore, subclustering has not yet been proposed. Thus, Jolie2 may represent a different subcluster. As expected, the majority of new isolates belong to the most populous clusters described to date (A and B). Interestingly, mycobacteriophage 32HC is a new singleton, being a temperate phage with an open reading frame (*orf47*) encoding a putative integrase belonging to the tyrosine recombinase family. Bioinformatics analysis suggested a possible tRNA gene (MSMEG_5758) for the integration of this mycobacteriophage into the *M. smegmatis* chromosome. Roughly 50% of the *orf* genes of 32HC displayed homology to *Mycobacterium abscessus* genes encoding hypothetical proteins of unknown function; however, preliminary experiments indicated that 32HC does not infect this mycobacterial species.

Nucleotide sequence accession numbers. The complete genomes of these nine mycobacteriophages have been deposited in GenBank under the following accession no.: KJ028219 (32HC), KJ192196 (40AC), KJ410132 (20ES), KJ410133 (Jolie2), KJ410134 (CRB1), KJ433974 (Hosp), KJ433973 (39HC), KJ433975 (40BC), and KJ433976 (Jolie1).

REFERENCES

1. Hatfull GF. 2012. The secret lives of mycobacteriophages. Adv. Virus Res. 82:179–288. http://dx.doi.org/10.1016/B978-0-12-394621-8.00015-7.
2. Hatfull GF, Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) Program, KwaZulu-Natal Research Institute for Tuberculosis and HIV (K-RITH) Mycobacterial Genetics Course, University of California—Los Angeles Research Immersion Laboratory in Virology, Phage Hunters Integrating Research and Education (PHIRE) Program. 2013. Complete genome sequences of 63 mycobacteriophages. Genome Announc. 1(6):e00847-13. http://dx.doi.org/10.1128/genomeA.00847-13.

3. **Hatfull GF, Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science Program, KwaZulu-Natal Research Institute for Tuberculosis and HIV Mycobacterial Genetics Course Students, Phage Hunters Integrating Research and Education Program.** 2012. Complete genome sequences of 138 mycobacteriophages. J. Virol. **86:**2382–2384. http://dx.doi.org/10.1128/JVI.06870 -11.

4. **Stella EJ, Franceschelli JJ, Tasselli SE, Morbidoni HR.** 2013. Analysis of novel mycobacteriophages indicates the existence of different strategies for phage inheritance in mycobacteria. PLoS One **8:**e56384. http://dx.doi.org/ 10.1371/journal.pone.0056384.

5. **Borodovsky M, Mills R, Besemer J, Lomsadze A.** 2003. Prokaryotic gene prediction using GeneMark and GeneMark.hmm. Curr. Protoc. Bioinformatics Chapter 4:Unit 4.5. http://dx.doi.org/10.1002/0471250953.bi0405s01.

6. **Laslett D, Canback B.** 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. **32:**11–16. http://dx.doi.org/10.1093/nar/gkh152.

7. **Söding J, Biegert A, Lupas AN.** 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. **33:**W244–W2W2. http://dx.doi.org/10.1093/nar/gki162.

8. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M.** 2014. Pfam: the protein families database. Nucleic Acids Res. **42:**D222–D230. http://dx.doi.org/10.1093/nar/gkt1223.