

# Complete Genome Sequences of Mycobacteriophages Balomoji, Crumble, Frederick, ShowerHandel, TinyTimmy and Zabiza

Zachery Middleton<sup>1</sup>, Giovanna Alves<sup>1</sup>, Em Bavuso<sup>1</sup>, Phuong-Anh Bui<sup>1</sup>, Kai Jandial<sup>1</sup>, Drew Kirchner<sup>1</sup>, Taylor Kiskamp<sup>1</sup>, Blair Morgan<sup>1</sup>, Kezie Osei<sup>1</sup>, Rohit Paradkar<sup>1</sup>, Sofia Taylor<sup>1</sup>, Allison A. Johnson<sup>2§</sup>

<sup>1</sup>Virginia Commonwealth University, Richmond, Virginia, United States

<sup>2</sup>Center for Biological Data Science, Virginia Commonwealth University, Richmond, Virginia, United States

<sup>§</sup>To whom correspondence should be addressed: aajohnson@vcu.edu

## Abstract

We report the genome sequences of six novel phages infecting *Mycobacterium smegmatis* mc<sup>2</sup>155. Each phage has siphoviral morphology and a double-stranded DNA genome. These genomes represent five different clusters of Mycobacterium phages, including a cluster K1 phage with a genome deletion that impacts the genes involved in lysogeny.



A) Plaque morphology of mycobacteriophage Zabiza shows clear plaques with no halo. B) Genome map region of cluster K1 mycobacteriophages ActinUp and Zabiza. The genome regions encoding the integrase and immunity repressor in ActinUp and the corresponding region in Zabiza are shown, with alignment of the genomes by *cro* and *excise* proteins. Protein coding genes are indicated by colored rectangles. Proteins with predicted function are labeled. The dotted line rectangle represents a large open reading frame in Zabiza, with a purple region that matches the first 97 amino acids of the 126 amino acid repressor protein from ActinUp (ActinUp gp43). The Zabiza ORF lacks a corresponding stop codon and continues until 29523. Purple and pink shading between the genomes indicates homologous regions with high percent identity and low *e*-values in blastn.

## Description

Discovery of mycobacteriophages advances our understanding of phage biology. Mycobacteriophage infect and lyse *Mycobacterium* hosts. With the increasing prominence of phage therapy as treatment for antibiotic-resistant nontuberculous mycobacterium infections (Dedrick et al., 2023) understanding of phage genome diversity can provide invaluable insights. This article describes the discovery and genome characteristics of six *Mycobacterium* phages.

Phage were isolated from soil samples (Table 1) and characterized through established methodologies (Zorawik et al., 2024). Samples were extracted in liquid medium (Middlebrook 7H9 supplemented with 145 mM NaCl, 5% Albumin fraction V, 2% dextrose and 1 mM CaCl<sub>2</sub>), the wash filtered (0.22 μm pore), and the filtrate either directly plated in top agar containing *M. smegmatis* mc<sup>2</sup>155 (Frederick) or first inoculated with *M. smegmatis* and incubated at 37°C with shaking for 24–48 hr to enrich for mycobacteriophages before being re-filtered and plated in top agar. Phage were purified by several rounds of plaque assays. Then, a high titer lysate was prepared by flooding a plate with a high density of plaques with phage buffer (10 mM Tris pH 7.5, 10 mM MgSO<sub>4</sub>, 68 mM NaCl, 1 mM CaCl<sub>2</sub>). Negative staining (1% uranyl acetate) and transmission electron microscopy revealed siphoviral morphologies for all the phages. (Table 1).

DNA was isolated from phage lysates using the Promega Wizard DNA cleanup kit. Genomic DNA for phage Frederick was sequenced by 454 FLX sequencer following library preparation using GS FLX Titanium series reagents whereas genomic DNA for all other phages were sequenced by Illumina MiSeq sequencer (v3 reagents) following library preparation using NEBNext Ultra II FS kit. Reads were assembled into a single contig for each genome using Newbler (v1.1 for phage Frederick, v2.9 for all other phages). Genomes were visually checked for end type and completeness using Consed (v20 for phage Frederick or v29 for all other phages)<sup>3</sup>. Sequencing and genome details are summarized in Table 1.

To annotate the genes in each genome, DNA Master (<http://cobamide2.bio.pitt.edu>) or PECAAN (Reinehart et al., 2016), annotation platforms that integrate data from genomic databases and bioinformatics tools, were used. Open reading frames (ORFs) were predicted by Glimmer v3.02 (Salzberg et al., 1998) and Genemark v2.5 (Besemer & Borodovsky, 2005), and tRNA genes were predicted by Aragorn v1.2.38 (Laslett & Canback, 2004) and tRNAScan v2.0 (Lowe & Chan, 2016). Predicted protein functions were determined using HHPred (PDB\_mmCIF70, Pfam-A, NCBI\_Conserved\_Domains databases (Söding et al., 2005)), Blastp (NCBI non-redundant database (Altschul et al., 1997)), and DeepTMHMM (Hallgren et al., 2022). Default settings were used for all software. Positional and functional annotations for each gene were manually curated (Table 1). Phages were assigned to clusters based on gene content similarity of at least 35% to phages in the Actinobacteriophage database, <https://phagesdb.org/> (Table 1) (Hatfull et al., 2010; Russell & Hatfull, 2017; Gauthier & Hatfull, 2023).

Based on identifiable integrase and immunity repressor gene functions, phages Balamoji, Crumble, and ShowerHandel are predicted to be temperate. Experimental studies will be necessary to demonstrate lysogeny. Genes typically involved in lysogen formation were not identified in Frederick, and TinyTimmy, which are thus unlikely to establish lysogeny. Cluster K1 phage Zabiza genome contains a deletion that impacts integrase and repressor, while retaining *excise* and *cro*. The deletion results in the absence of the tyrosine integrase (Figure 1). The Zabiza genome has also retained a region with coding potential and 100% homology to 78% query coverage of ActinUp repressor protein, but lacks a corresponding stop codon. Wet lab experiments would be needed to confirm whether this repressor is functional. This is an important observation since phages in this predominantly temperate cluster have been shown to infect both fast- and slow-growing mycobacterial hosts (Pope et al., 2011). Examination of 115 cluster K1 mycobacteriophages (Russell and Hatfull, 2017) revealed three other cluster K phages that appear to lack an identifiable integrase, and would be predicted to not form lysogens. Among the 641 total putative ORFs identified across all 6 phage genomes, two are novel genes not found in the Actinobacteriophage database. These novel genes lack predicted function. Finally, Crumble and ShowerHandel encode mycobacteriophage mobile elements (MMPE1 and MMPE2, respectively), which are recalcitrant and found in distinct regions across related phage genomes (Sampson et al., 2009).

### Data availability

GenBank and Sequence Read Archive (SRA) accession numbers are provided in Table 1.

**Table 1. Phage and genome sequence characteristics**

Phage	<b>Balomoji</b>	<b>Crumble</b>	<b>Frederick</b>	<b>ShowerHandel</b>	<b>TinyTimmy</b>	<b>Zabiza</b>
Sample Location (City, State)	Richmond, VA	Richmond, VA	Richmond, VA	Richmond, VA	Richmond, VA	Richmond, VA
Sample location GPS coordinates	37.54572 N, 77.44931 W	37.544917 N, 77.455 W	37.545955 N, 77.453922 W	37.544897 N, 77.454942 W	37.5435 N, 77.4478 W	37.54574 N, 77.45141 W
Predicted lifestyle	Temperate	Temperate	Lytic	Temperate	Lytic	May be lytic
Plaque Morphology	clear with halo	clear	clear	clear	clear	clear with halo
Plaque Size	2.5 ± 0.8 mm (n = 10)	1.7 ± 0.7 mm (n = 10)	1 mm (n=1)	1.2 ± 0.2 mm (n = 10)	0.9 ± 0.2 mm (n = 10)	2.9 ± 0.6 mm (n= 10)
Capsid Size (nm)	61.8 ± 1.2 nm (n = 3)	49.3 ± 2.4 nm (n = 3)	54 nm (n=1)	61.8 ± 1.2 nm (n = 3)	63.2 ± 1.2 nm (n = 3)	52.8 ± 2.4 nm (n = 3)
Tail Length (nm)	254.9 ± 7.9 nm (n = 3)	144.4 ± 2.4 nm (n = 3)	175 nm (n=1)	198.6 ± 3.2 nm (n = 3)	138.9 ± 6.0 nm (n = 3)	179.2 ± 2.1 nm (n = 3)
SRA accession number	SRX25734218	SRX25734219	SRX25734228	SRX25734221	SRX25734223	SRX25734225
# of reads	742,315	291,815	6,185	163,063	3,790,164	2,692,655
Approximate read coverage	1420	717	34	400	15804	11874
Genbank accession number	PP978826	PP978871	PQ244021	PP978878	PQ362662	PQ362688
Cluster	E	F1	B4	F1	A11	K1
Genome length (bp)	76419	53424	58937	57562	51840	56117

# ORFs (# with predicted function)	144 (50)	99 (40)	98 (40)	107 (42)	97 (42)	91 (51)
# Orphans	1	0	0	0	0	1
# tRNAs	2	2	0	0	1	1
GC content (%)	62.8	61.4	61.4	61.4	63.7	66.7
End type	Defined end	Defined end	Circularly permuted	Defined end	Defined end	Defined end
Length of 3'-sticky overhang	9 bases (CGCTTGTC A)	10 bases (CGGACGGCGC)	N/A	10 bases (CGGACGGCGC)	10 bases (CGGTCCGGTTA)	11 bases (CTCGTAGGCAT)

**Acknowledgements:** We thank Daniel Russell and Rebecca Garlena for sequencing and assembling the genomes, and the HHMI SEA-PHAGES program and VCU Life Sciences for support. We thank the SEA PHAGES community for review of this manuscript.

## References

- Altschul S. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-3402. DOI: [25.17.3389](https://doi.org/10.1093/nar/25.17.3389)
- Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research* 33: W451-W454. DOI: [10.1093/nar/gki487](https://doi.org/10.1093/nar/gki487)
- Dedrick RM, Smith BE, Cristinziano M, Freeman KG, Jacobs-Sera D, Belessis Y, et al., Hatfull. 2022. Phage Therapy of *Mycobacterium* Infections: Compassionate Use of Phages in 20 Patients With Drug-Resistant Mycobacterial Disease. *Clinical Infectious Diseases* 76: 103-112. DOI: [10.1093/cid/ciac453](https://doi.org/10.1093/cid/ciac453)
- Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, Krogh A, Winther O. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. : 10.1101/2022.04.08.487609. DOI: [10.1101/2022.04.08.487609](https://doi.org/10.1101/2022.04.08.487609)
- Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, et al., Hendrix. 2010. Comparative Genomic Analysis of 60 Mycobacteriophage Genomes: Genome Clustering, Gene Acquisition, and Gene Size. *Journal of Molecular Biology* 397: 119-143. DOI: [10.1016/j.jmb.2010.01.011](https://doi.org/10.1016/j.jmb.2010.01.011)
- Laslett D. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research* 32: 11-16. DOI: [10.1093/nar/gkh152](https://doi.org/10.1093/nar/gkh152)
- Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research* 44: W54-W57. DOI: [10.1093/nar/gkw413](https://doi.org/10.1093/nar/gkw413)
- Pope WH, Ferreira CM, Jacobs-Sera D, Benjamin RC, Davis AJ, DeJong RJ, et al., Hatfull. 2011. Cluster K Mycobacteriophages: Insights into the Evolutionary Origins of Mycobacteriophage TM4. *PLoS ONE* 6: e26750. DOI: [10.1371/journal.pone.0026750](https://doi.org/10.1371/journal.pone.0026750)
- Reinehart, C.; Gaffney B.; Wood J.D.; Smith J.. (2016). *PECAAN, a Phage Evidence Collection And Annotation Network*. <https://discover.kbrinsgd.org>
- Russell DA, Hatfull GF. 2016. PhagesDB: the actinobacteriophage database. *Bioinformatics* 33: 784-786. DOI: [10.1093/bioinformatics/btw711](https://doi.org/10.1093/bioinformatics/btw711)
- Salzberg SL, Delcher AL, Kasif S, White O. 1998. Microbial gene identification using interpolated Markov models. *Nucleic Acids Research* 26: 544-548. DOI: [10.1093/nar/26.2.544](https://doi.org/10.1093/nar/26.2.544)

6/23/2025 - Open Access

Sampson T, Broussard GW, Marinelli LJ, Jacobs-Sera D, Ray M, Ko CC, et al., Hatfull. 2009. Mycobacteriophages BPs, Angel and Halo: comparative genomics reveals a novel class of ultra-small mobile genetic elements. *Microbiology* 155: 2962-2977. DOI: [10.1099/mic.0.030486-0](https://doi.org/10.1099/mic.0.030486-0)

Soding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* 33: W244-W248. DOI: [10.1093/nar/gki408](https://doi.org/10.1093/nar/gki408)

Zorawik M, Jacobs-Sera D, Freise AC, SEA-PHAGES, Reddi K. 2024. Isolation of Bacteriophages on Actinobacteria Hosts. *Methods in Molecular Biology, Phage Engineering and Analysis* : 273-298. DOI: [10.1007/978-1-0716-3798-2\\_17](https://doi.org/10.1007/978-1-0716-3798-2_17)

**Funding:** No external funding supported this work.

**Author Contributions:** Zachery Middleton: data curation, investigation, writing - original draft. Giovanna Alves: data curation, writing - original draft, investigation. Em Bavuso: data curation, investigation, writing - original draft. Phuong-Anh Bui: data curation, investigation. Kai Jandial: data curation, investigation, writing - original draft. Drew Kirchner: data curation, investigation. Taylor Kiskamp: investigation. Blair Morgan: data curation, investigation, writing - original draft. Kezie Osei: data curation, investigation. Rohit Paradkar: data curation, investigation, writing - original draft. Sofia Taylor: data curation, investigation, writing - original draft. Allison A. Johnson: data curation, project administration, writing - review editing.

**Reviewed By:** Anonymous

**History:** Received December 18, 2024 **Revision Received** June 8, 2025 **Accepted** June 18, 2025 **Published Online** June 23, 2025  
**Indexed** July 7, 2025

**Copyright:** © 2025 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Citation:** Middleton Z, Alves G, Bavuso E, Bui PA, Jandial K, Kirchner D, et al., Johnson AA. 2025. Complete Genome Sequences of Mycobacteriophages Balomaji, Crumble, Frederick, ShowerHandel, TinyTimmy and Zabiza. *microPublication Biology*. [10.17912/micropub.biology.001478](https://doi.org/10.17912/micropub.biology.001478)