Complete genome sequences of 38 Gordonia sp. bacteriophages

Welkin H. Pope  
*University of Pittsburgh*

Matthew T. Montgomery  
*University of Pittsburgh*

Randall J. DeJong  
*Calvin University*

Follow this and additional works at: [https://digitalcommons.calvin.edu/calvin_facultypubs](https://digitalcommons.calvin.edu/calvin_facultypubs)

Part of the Genomics Commons

**Recommended Citation**

Pope, Welkin H.; Montgomery, Matthew T.; and DeJong, Randall J., "Complete genome sequences of 38 Gordonia sp. bacteriophages" (2017). *University Faculty Publications*. 250.  
[https://digitalcommons.calvin.edu/calvin_facultypubs/250](https://digitalcommons.calvin.edu/calvin_facultypubs/250)

This Article is brought to you for free and open access by the University Faculty Scholarship at Calvin Digital Commons. It has been accepted for inclusion in University Faculty Publications by an authorized administrator of Calvin Digital Commons. For more information, please contact dbm9@calvin.edu.
Complete Genome Sequences of 38 *Gordonia* sp. Bacteriophages

Welkin H. Pope, Matthew T. Montgomery, J. Alfred Bonilla, Randall Dejong, Rebecca A. Garlena, Carlos Guerrero Bustamante, Karen K. Klyczek, Daniel A. Russell, John T. Wertz, Deborah Jacobs-Sera, Graham F. Hatfull

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; Department of Biology, University of Wisconsin – River Falls, River Falls, Wisconsin, USA; Department of Biology, Calvin College, Grand Rapids, Michigan, USA

**ABSTRACT** We report here the genome sequences of 38 newly isolated bacteriophages using *Gordonia terrae* 3612 (ATCC 25594) and *Gordonia neofelifaecis* NRRL 59395 as bacterial hosts. All of the phages are double-stranded DNA (dsDNA) tail phages with siphoviral morphologies, with genome sizes ranging from 17,118 bp to 93,843 bp and spanning considerable nucleotide sequence diversity.

The bacteriophage population is vast, dynamic, and old, with an estimated population of $10^{31}$ virions and $10^{23}$ productive infections/s on a global scale (1). The genomic diversity of the population is poorly understood, with fewer than 3,000 complete genome sequences in GenBank. In general, phages isolated on phylogenetically unrelated hosts share little or no sequence similarity, but considerable insights can be gleaned by comparative genomics of phages isolated on a common host, as illustrated for enterobacteriophages and mycobacteriophages (2, 3). The Howard Hughes Medical Institute (HHMI) Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program provides an undergraduate course-based research experience that contributes to our understanding of phage diversity and evolution through bacteriophage discovery and genomics, using *Actinobacteria*, including mycobacteria and *Gordonia* sp. strains, as isolation hosts.

*Gordonia* phages were isolated by enrichment or direct plating of filtered soil samples using *Gordonia terrae* 3612 or *Gordonia neofelifaecis* NRRL 59395 as a host (Table 1). Thirty-eight individual phages were isolated, and electron microscopy shows that all have siphoviral morphotypes. Plaque-purified phages were amplified, and their double-stranded DNA (dsDNA) was extracted and sequenced using an Illumina MiSeq, as described previously (4). The 140-base reads were assembled using Newbler and Consed, with average coverages between 447- and 3,241-fold. Sequence ambiguities and genome termini were resolved either by sequencing directly from genomic templates or from PCR products. Genomes were annotated using DNA Master (http://cobamide2.bio.pitt.edu), coding sequences were predicted using GeneMark (5) and Glimmer (6), and tRNAs were predicted using Aragorn (7) and tRNAscan-SE (8). Functional assignments were made using BLASTP (9) and HHpred (10, 11) against the publically available databases GenBank, the Protein Data Bank, and Pfam.

The 38 newly isolated *Gordonia* phages exhibit considerable diversity (Table 1). The smallest genomes, Jeanie and McGonagall, at ~17,000 bp, have the highest G+C content (68%) and are each predicted to contain only 25 genes, including those encoding structural proteins, integrase and immunity repressor, endolysin, and a DnaQ-like subunit of DNA polymerase III. Three phages (PatrickStar, Kampe, and Orchid) have G+C contents (47%) that are strikingly lower than that of their host (67.77%), and lower than the G+C% of any mycobacteriophage; these phages may be relatively...
recent arrivals to the *Gordonia* neighborhood (12) (Table 1). These phages, together with Kvothe, Jumbo, and Demothenes, have genomes with direct terminal repeats, a feature not observed in any mycobacteriophages. Many of the *Gordonia* phage genomes have defined ends with 3′ single-stranded extensions (Table 1), and only three (Terapin, Twister6, and Wizard) are circularly permuted.

Most of the *Gordonia* phages form turbid plaques, and 27 of the 38 encode either tyrosine or serine integrases; another six phages encode putative ParAB partitioning systems. Temperate lifestyles thus appear to be common for these phages. Some of the phages have all or part of a second integrase gene, and although these are mostly predicted to be nonfunctional, they perhaps reflect relatively recent genomic rearrangements. Finally, we note that six phages, KatherineG, Rosalind, Strosahl, Remus, Soups, and JSwag, are sufficiently similar to some mycobacteriophages to warrant grouping within Cluster A (13).

**Accession number(s).** Nucleotide sequence accession numbers are shown in Table 1.

### Table 1 *Gordonia* phage genomics

| Phage name  | GenBank accession no. | Genome size (bp) | G+C content (%) | No. of tRNAs | No. of CDSs | End typea | Host strain |
|-------------|-----------------------|------------------|-----------------|-------------|-------------|-----------|-------------|
| Bachita     | KU998247              | 93,843           | 61.9            | 8           | 182         | CGCGACGCTC | G. terrae 3612 |
| Bantam      | KX557272              | 92,580           | 64.7            | 2           | 168         | CGCAGACCTC | G. terrae 3612 |
| BatStar     | KX557273              | 53,432           | 66.6            | 0           | 83          | CGGCTGGGGA | G. terrae 3612 |
| Blueberry   | KU998236              | 54,990           | 67              | 0           | 86          | TGGCCGTTGA | G. terrae 3612 |
| BritBrat    | KU998233              | 55,524           | 65              | 0           | 98          | CGTATGGCAT | G. terrae 3612 |
| CaptainKirk2| KX557274              | 47,898           | 67.4            | 0           | 79          | TCGCGCCGTA | G. terrae 3612 |
| CarolAnn    | KX557275              | 54,167           | 66.9            | 0           | 80          | TGGCCGTTGA | G. terrae 3612 |
| ClubL       | KU998246              | 92,618           | 61.9            | 9           | 179         | CGCGACGCTC | G. terrae 3612 |
| Cozz        | KU998239              | 46,600           | 60              | 0           | 68          | CGTATGGCTT | G. terrae 3612 |
| Cucurbita   | KX557276              | 93,686           | 62              | 9           | 178         | CGCGACGCTC | G. terrae 3612 |
| Demothenes  | KU998242              | 74,073           | 59.3            | 0           | 95          | Dir. Term. Repeat | G. terrae 3612 |
| Eyre        | KX557277              | 44,929           | 67.5            | 0           | 74          | CCCCCTGCGTG | G. terrae 3612 |
| Ghobes      | KX557278              | 45,285           | 65.2            | 0           | 59          | TGGCCGAGGT | G. terrae 3612 |
| Hedwig      | KX557279              | 44,536           | 67.2            | 0           | 70          | TCCGGCGGTA | G. terrae 3612 |
| Howe        | KU252585              | 53,182           | 65              | 0           | 79          | TGGCCGTTGA | G. terrae 3612 |
| JSwag       | KX557280              | 52,726           | 61.9            | 3           | 101         | CGGCGGTGA | G. terrae 3612 |
| Jumbo       | KX557281              | 78,302           | 54.5            | 0           | 102         | Dir. Term. Repeat | G. terrae 3612 |
| Kampe       | KU998254              | 80,649           | 47              | 2           | 115         | Dir. Term. Repeat | G. terrae 3612 |
| KatherineG  | KU998251              | 52,689           | 61.9            | 3           | 99          | CGGCGGTGA | G. terrae 3612 |
| Kvothe      | KU998243              | 75,462           | 59.5            | 0           | 99          | Dir. Term. Repeat | G. terrae 3612 |
| Nyceirae    | KX557282              | 41,857           | 67.5            | 0           | 61          | CGGCGGGGA | G. terrae 3612 |
| OneUp       | KU998245              | 93,577           | 61.5            | 9           | 163         | CGCGACGCTC | G. terrae 3612 |
| Orchid      | KU998253              | 80,650           | 47              | 2           | 114         | Dir. Term. Repeat | G. terrae 3612 |
| PatrickStar | KU998252              | 80,729           | 47              | 2           | 115         | Dir. Term. Repeat | G. terrae 3612 |
| Remus       | KX557283              | 52,738           | 62              | 3           | 98          | CGGCTGGTTA | G. terrae 3612 |
| Rosalind    | KU998250              | 52,684           | 61.9            | 3           | 99          | CGGCTGGTTA | G. terrae 3612 |
| Smoothie    | KU998244              | 93,139           | 61.9            | 8           | 179         | CGGCGGCTC | G. terrae 3612 |
| Soups       | KU998249              | 52,924           | 61.9            | 3           | 98          | CGGCTGGTTA | G. terrae 3612 |
| Splinter    | KU998238              | 45,858           | 66.1            | 0           | 80          | TCGGCGGCGGTA | G. terrae 3612 |
| Strosahl    | KX557284              | 52,738           | 62              | 3           | 98          | CGGCTGGTTA | G. terrae 3612 |
| Terrapin    | KX557285              | 66,611           | 59.6            | 0           | 97          | Circ. Permuted | G. terrae 3612 |
| Twister6    | KX557286              | 57,804           | 67.7            | 0           | 93          | Circ. Permuted | G. terrae 3612 |
| Utz         | KU998248              | 49,768           | 67.7            | 0           | 71          | TCGCCGTTGA | G. terrae 3612 |
| Vendetta    | KU998237              | 45,858           | 66.1            | 0           | 81          | TCGGCCGCGTA | G. terrae 3612 |
| Wizard      | KU998234              | 58,308           | 67.9            | 0           | 89          | Circ. Permuted | G. terrae 3612 |
| Zirinka     | KX557278              | 52,077           | 66.7            | 0           | 79          | CGGCTGGGGA | G. terrae 3612 |
| Jeanie      | KU998256              | 17,118           | 68.6            | 0           | 25          | AGCCGCCGCTG | G. neofelisaecis |
| McGonagall  | KU998255              | 17,119           | 68.6            | 0           | 25          | AGCCGCCGCTG | G. neofelisaecis |

**a**CDSs, coding sequences.

**b**End types are 3′-single-stranded overhangs, unless otherwise noted as Dir. Term. Repeat (direct terminal repeat) or Circ. Permuted (circularly permuted).

**c**Phage Hunters Integrating Research and Education (PHIRE) program, University of Pittsburgh.

**d**Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES), University of Wisconsin-River Falls.

**e**SEA-PHAGES, University of Pittsburgh.

**f**SEA-PHAGES, Calvin College.

**g**This total includes one transfer-messenger RNA (tmRNA).
ACKNOWLEDGMENTS

We thank Marcie Warner, Becky Bortz, Sarah Grubb, Emily Furbee, and the students of the SEA-PHAGES programs at the University of Pittsburgh, Calvin College, and the University of Wisconsin–River Falls for their invaluable contributions in phage discovery and phage genomics.

REFERENCES

1. Hendrix RW. 2002. Bacteriophages: evolution of the majority. Theor Popul Biol 61:471–480. https://doi.org/10.1006/tpbi.2002.1590.

2. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cesaawrn 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.

3. Grose JH, Casjens SR. 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. Virology 468–470:421–443. https://doi.org/10.1016/j.virol.2014.08.024.

4. 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. 2016. Complete genome sequences of 61 mycobacteriophages. Genome Announc 4:e00389-16. https://doi.org/10.1128/genomeA.00389-16.

5. Besemer J, Borodovsky M. 2005. GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451–W454. https://doi.org/10.1093/nar/gki408.

6. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinformatics 23:673–679. https://doi.org/10.1093/bioinformatics/btm009.

7. 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.

8. Lowe TM, Eddy SR. 1997. TRINscan-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/25.5.955.

9. 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.

10. Remmert M, Biegert A, Hauser A, Söding J. 2011. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods 9:173–175. https://doi.org/10.1038/nmeth.1818.

11. 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/gkj408.

12. Pope WH, Jacobs-Sera D, Russell DA, Rubin DH, Kajee A, Msibi ZN, Larsen MH, Jacobs WR, Jr, Lawrence JG, Hendrix RW, Hatfull GF. 2014. Genomics and proteomics of mycobacteriophage patience, an accidental tourist in the mycobacterium neighborhood. mBio 5:e02145. https://doi.org/10.1128/mBio.02145-14.

13. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germaine KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O’Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cesaawrn SG, Hendrix RW. 2010. Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. J Mol Biol 397:119–143. https://doi.org/10.1016/j.jmb.2010.01.011.