

# Genome Sequences of Bacteriophages Natasha and RustyBoy, Isolated on *Arthrobacter globiformis*, and Fazel, Isolated on *Gordonia rubripertincta* in Southwestern New Mexico

Rebecca Merchant<sup>1\*</sup>, Violet Kirk<sup>1\*</sup>, Isabella McClish<sup>1\*</sup>, Meghan Ferrara<sup>1\*</sup>, Luke Garcia<sup>1\*</sup>, Riley Oesch<sup>1\*</sup>, Itxel Barrera-Moncayo<sup>1\*</sup>, Alan Blair<sup>1\*</sup>, Erika Eaton<sup>1\*</sup>, Gwenevere Gatto<sup>1\*</sup>, Laurel Geoffrion<sup>1\*</sup>, Micah Kirkman<sup>1\*</sup>, Abey Kouchich-Martens<sup>1\*</sup>, Lalitha Madduri<sup>1\*</sup>, Sophia Rosa<sup>1\*</sup>, Marisa Valdez<sup>1\*</sup>, Ana Zamora<sup>1\*</sup>, Linda DeVeaux<sup>1\*</sup>, Nathaniel Jobe<sup>1§\*</sup>, Kaarin Goncz<sup>1§\*</sup>

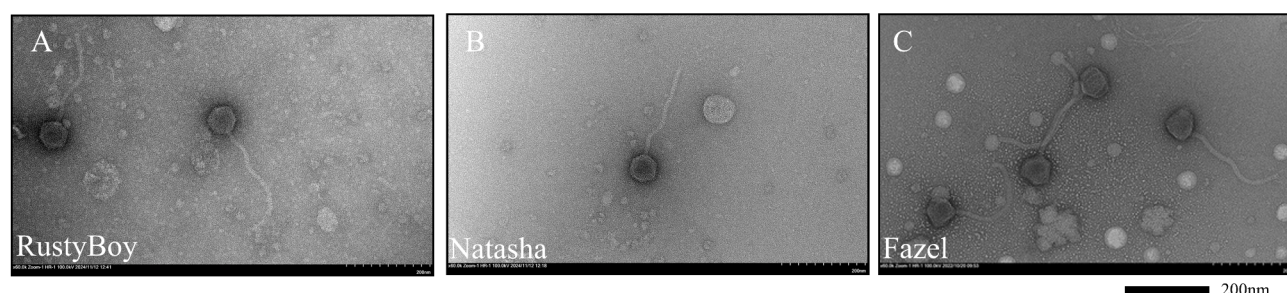
<sup>1</sup>Biology, New Mexico Institute of Mining and Technology, Socorro, NM, US

<sup>§</sup>To whom correspondence should be addressed: nathaniel.jobe@student.nmt.edu; kaarin.goncz@nmt.edu

\*These authors contributed equally.

## Abstract

We report the discovery and genome sequence of three phages with siphovirus morphology isolated from soil collected in Socorro, New Mexico. Phages Natasha and RustyBoy, which were isolated using *Arthrobacter globiformis* B-2979, are both assigned to actinobacteriophage cluster AW based on gene content. Phage Fazel was isolated using *Gordonia rubripertincta* NRRL B-16540 and assigned to cluster DJ.



**Figure 1. Transmission electron micrographs of phage lysate display a siphovirus morphology:**

RustyBoy (A), Natasha (B), and Fazel (C). Lysates were placed on 200-mesh formvar-covered, carbon-coated copper grids (EMS) stained with 2% uranyl acetate and imaged in a Hitachi HT7800 TEM at 100 kV. Scale bar is 200 nm. Images by Tagide deCarvalho.

## Description

Bacteriophages are abundant, present across diverse environments, and are genetically diverse (Clokiet al., 2011). They have multifaceted roles in bacterial ecology and evolution (Chevallereau et al., 2022). Their isolation and genetic characterization continues to advance molecular biology, biotechnology, and medicine (Sharma et al., 2017). Here, we describe the isolation and genomic characteristics of three novel bacteriophages.

All three phages were isolated from soil samples collected in Socorro, New Mexico, (GPS coordinates provided in Table 1) using standard methods (Zorawik et al., 2024). Briefly, soil samples were resuspended in PYCa liquid medium and the suspension subsequently filtered (0.2 micron pore size). Fazel was isolated from a plaque that formed after the filtrate was plated directly in top agar with *Gordonia rubripertincta* NRRL B-16540, following incubation at 30° C for 3 days. RustyBoy and Natasha were isolated from plaques after the filtrate was first inoculated with *Arthrobacter globiformis* B-2979 and allowed to incubate with shaking for 72 hours at 30°C before the resulting culture was filtered and that filtrate plated in top agar with *Arthrobacter globiformis*. All phages were purified through three rounds of additional plating. Plaque characteristics are presented in Table 1. Negative stain transmission electron microscopy revealed all three phages to have a siphovirus morphology (Figure 1).

DNA was isolated from lysates prepared for each phage using the Wizard DNA Cleanup Kit (Promega) and prepared for sequencing using the NEB Ultra II FS library prep kit. DNA for phage Fazel was sequenced on an Illumina MiSeq (v.3 reagents), with raw reads subsequently assembled with Newbler (v2.9), and verified using Consed (v29) (Russell, 2018). DNA for phages RustyBoy and Natasha were sequenced on an Illumina NextSeq1000 (XLEAP-P1 kit), with raw reads first trimmed with Cutadapt 4.7 (using the option: -nextseq-trim 30) and filtered with Skewer 0.2.2 (using the options: -q 20 -Q 30 -n -l 50) prior to assembly. The number of reads, depth of sequencing, and genome characteristics are presented in Table 1.

Table 1. Phage Characteristics and Genome Information.

|                                       | <b>Fazel</b>               | <b>RustyBoy</b>             | <b>Natasha</b>              |
|---------------------------------------|----------------------------|-----------------------------|-----------------------------|
| <b>GPS</b>                            | 34.06532 N<br>106.906919 W | 34.066472 N<br>106.903833 W | 34.066472 N<br>106.903833 W |
| <b>Tail Length, nm</b>                | 242 - 295 (n=4)            | 232 - 293 (n=4)             | 216 - 253 (n=3)             |
| <b>Capsid Diameter, nm</b>            | 60 - 75 (n=4)              | 62 - 70 (n=4)               | 63 - 68 (n=3)               |
| <b>Plaque Size (n=10)</b>             | 1.6 – 2.4 mm               | 0.9 – 1.2 mm                | 0.7 – 1.1 mm                |
| <b>Plaque Morphology</b>              | Clear                      | Cloudy                      | Clear                       |
| <b>Number of Reads</b>                | 110,495 spots              | 2,563,562 spots             | 2,654,698 spots             |
| <b>Sequencing Coverage</b>            | 150-base single-end reads  | 100-base single-end reads   | 100-base single-end reads   |
| <b>Genome Length (bp)</b>             | 60,088                     | 54,635                      | 54,436                      |
| <b>GC Content</b>                     | 52.0%                      | 51.4%                       | 51.4%                       |
| <b>Cluster Assignment</b>             | DJ                         | AW                          | AW                          |
| <b>Number of Genes</b>                | 92                         | 89                          | 88                          |
| <b>Genes with predicted Functions</b> | 29                         | 19                          | 21                          |

Genomes were then annotated using PECAAN v20241104 (Rinehart et al.), which utilized Glimmer v3.02 (Delcher et al., 2007) and GeneMark (v2.5 for Fazel and v2.5p for RustyBoy and Natasha) (Besemer and Borodovsky, 2005) to automatically identify putative genes. Gene start sites were then manually refined using Starterator (<http://phages.wustl.edu/starterator/>), while putative functions were assigned using HHPred (Zimmerman et al., 2018) searches against the PDB\_mmCIF70 (Berman et al., 2000), SCOPe70 (Fox et al., 2014), Pfam-A v37 (Finn et al., 2014), and NCBI\_Conserved\_Domains databases (Marchler-Bauer et al., 2017), BLASTp searches (Madden, 2013) against the Actinobacteriophage and NCBI non-redundant protein databases, and Phamerator (Cresawn et al., 2011) with the Actino\_draft database v578. Default settings were used for all software.

Phages RustyBoy and Natasha are genetically similar, sharing 98 % nucleotide identity. Both phages encode 89/88 predicted genes, 86 of which are conserved between the two phages. Genes 54 and 61 in both phages are sorted into different gene families or “phams” based on limited protein sequence similarity (Cresawn et al., 2011). Indeed, gene 61 from Natasha is an “orpham”, for which no homologues exist within PhagesDB. Putative functions were identified for a subset of genes, including multiple HNH endonucleases, major capsid and protease fusion protein, and Cas4 exonuclease. Based on gene content similarity of at least 35 % to phages in the actinobacteriophage database, PhagesDB, both phages were assigned to cluster AW. As with other AW phages, all genes in RustyBoy and Natasha are encoded unidirectionally, with no identifiable immunity repressor or integrase functions, suggesting these phages are unable to establish lysogeny. Another AW phage, Sporto, that was isolated in Pittsburgh shared 92 % nucleotide identity to RustyBoy. Through genome alignment and phylogeny based on neighbour-joining trees using the MAFFT webserver (Katoh et al., 2019; Kuraku et al., 2013), Natasha, RustyBoy, and Sporto cluster more closely to one another within the AW cluster, warranting consideration for being grouped as an AW subcluster.

Fazel is predicted to encode 92 genes, 6 of which are orphans and 29 of which are assigned a putative function, including deoxycytidylate deaminase, PnuC-like nicotinamide riboside transporter, Thyx-like thymidylate synthase, VRR-NuC domain protein and Cas4 exonuclease. Fazel is assigned to cluster DJ. Consistent with a majority of DJ phages, lysin A and holin functions are encoded next to each other, with lysin B encoded further downstream. All genes are transcribed unidirectionally, which is true for all members of this cluster apart from phage Schwartz33, which contains one predicted

gene of unknown function that is transcribed in the opposite direction. Approximately 8 kbp within the central region of the genome contain multiple repeat motifs seen in other DJ cluster phages. Within this region, Fazel contains one predicted orpham, gene 46 (Henson et al., 2022). Based on the absence of predicted immunity repressor or integrase functions, Fazel and other cluster DJ phages are predicted to be lytic phages.

### Nucleotide sequence accession numbers

Fazel is available at GenBank with Accession No. PP978862 and Sequence Read Archive (SRA) No. SRX24123904. Natasha is available at GenBank with Accession No. PV876962 and Sequence Read Archive (SRA) No. SRX28943173. RustyBoy is available at GenBank with Accession No. PV876932 and Sequence Read Archive (SRA) No. SRX28943179.

**Acknowledgements:** We thank Kaarin Goncz and Nathaniel Jobe for their contribution to the 2025 Phage research program. Thank you to Dr. Goncz for leading us through the wet lab portion of the course and helping us with isolations and plaque assays. Thank you to Nathaniel for his patience on the genome annotation portion and for reviewing genes to ensure accurate results. Lastly, thank you to the SEA-PHAGES program for their resources, helping us in publishing our paper and establishing our phages in the worldwide database.

### References

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28, 235–242.
- Besemer, J., & Borodovsky, M. (2005). GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research*, 33, W451–W454
- Chevallereau A, Pons BtJ, van Houte S, Westra ER. 2021. Interactions between bacterial and phage communities in natural environments. *Nature Reviews Microbiology* 20: 49-62. DOI: [10.1038/s41579-021-00602-y](https://doi.org/10.1038/s41579-021-00602-y)
- Clokier, M. R., Millard, A. D., Letarov, A. V., & Heaphy, S. (2011). Phages in nature. *Bacteriophage*, 1, 31–45
- Cresawn, S.G, Bogel, M., Day, N., Jacobs-Sera, D., Hendrix, R. W., & Hatfull, G. F. (2011). Phamerator: a bioinformatics tool for comparative bacteriophage genomics. *BMC Bioinformatics*. 12, 395. <https://github.com/scresawn/phamerator>
- Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, 23, 673–679
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., & Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Research*. 42, D222–D230
- Fox, N. K., Brenner, S. E., & Chandonia, J.-M. (2014). SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research*. 42, D304–D309
- Henson, M, Fratarcangeli, M., Park, S., Minot, A. & Gainey, M. D. (2022) Genome Characteristics of Vardy, a Cluster DJ Actinobacteriophage Isolated on *Gordonia rupripertincta* in Western North Carolina. *Microbiology Resource Announcements*, 11, e00704-22
- Katoh, K., Rozewicki, J. & Yamada, K. D. (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, *Briefings in Bioinformatics* 20, 1160-1166
- Kuraku, S., Zmasek, C. M., Nishimura, O. & Katoh, K. (2013) aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research*, 41, W22-W28
- Madden, T. (2003). The BLAST Sequence Analysis Tool (National Center for Biotechnology Information (US). <https://www.ncbi.nlm.nih.gov/books/NBK21097/>
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y., & Bryant, S. H. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*. 45, D200–D203
- Rinehart, C. A., Gaffney, B., Smith, J., & Wood, J. D. PECAAN; <https://discover.kbrinsgd.org/evidence/summary>
- Russell DA. 2017. Sequencing, Assembling, and Finishing Complete Bacteriophage Genomes. *Methods in Molecular Biology, Bacteriophages* : 109-125. DOI: [10.1007/978-1-4939-7343-9\\_9](https://doi.org/10.1007/978-1-4939-7343-9_9)
- Sharma S, Chatterjee S, Datta S, Prasad R, Dubey D, Prasad RK, Vairale MG. 2016. Bacteriophages and its applications: an overview. *Folia Microbiologica* 62: 17-55. DOI: [10.1007/s12223-016-0471-x](https://doi.org/10.1007/s12223-016-0471-x)

12/15/2025 - Open Access

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., & Alva, V. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*. 430, 2237–2243

Zorawik, M., Jacobs-Sera, D., Freise, A. C., SEAPHAGES, & Reddi, K. (2024) Isolation of Bacteriophages on Actinobacteria Hosts. *Methods in Molecular Biology*. 2793, 273-298

**Funding:** The authors declare no additional sources of funding

**Author Contributions:** Rebecca Merchant: investigation, writing - original draft. Violet Kirk: investigation, writing - original draft. Isabella McClish: investigation, writing - original draft. Meghan Ferrara: investigation, writing - original draft. Luke Garcia: investigation, writing - original draft. Riley Oesch: investigation. Itxel Barrera-Moncayo: investigation. Alan Blair: investigation. Erika Eaton: investigation. Gwenevere Gatto: investigation. Laurel Geoffrion: investigation. Micah Kirkman: investigation. Abey Kouchich-Martens: investigation. Lalitha Madduri: investigation. Sophia Rosa: investigation. Marisa Valdez: investigation. Ana Zamora: investigation. Linda DeVeaux: supervision, investigation, validation. Nathaniel Jobe: supervision, investigation, validation. Kaarin Goncz: supervision, investigation, validation, writing - review editing.

**Reviewed By:** Anonymous, Jamie Wallen

**History:** Received October 29, 2025 **Revision Received** November 21, 2025 **Accepted** December 12, 2025 **Published Online** December 15, 2025 **Indexed** December 29, 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:** Merchant R, Kirk V, McClish I, Ferrara M, Garcia L, Oesch R, et al., Goncz K. 2025. Genome Sequences of Bacteriophages Natasha and RustyBoy, Isolated on *Arthrobacter globiformis*, and Fazel, Isolated on *Gordonia rubripertincta* in Southwestern New Mexico. microPublication Biology. [10.17912/micropub.biology.001988](https://doi.org/10.17912/micropub.biology.001988)