



Brief Report

Microbacterium Cluster EA Bacteriophages: Phylogenomic Relationships and Host Range Predictions

Mark Milhaven ¹, Cyril J. Versoza ¹, Aman Garg ², Lindsey Cai ², Sanjana Cherian ², Kamalei Johnson ², Kevin Salas Perez ², Madison Blanco ², Jackelyn Lobatos ², Corinne Mitra ², Maria Strasser ² and Susanne P. Pfeifer ^{1,*} 

¹ Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85281, USA

² School of Life Sciences, Arizona State University, Tempe, AZ 85281, USA

* Correspondence: susanne.pfeifer@asu.edu

Abstract: Bacteriophages are being widely harnessed as an alternative to antibiotics due to the global emergence of drug-resistant pathogens. To guide the usage of these bactericidal agents, characterization of their host specificity is vital—however, host range information remains limited for many bacteriophages. This is particularly the case for bacteriophages infecting the *Microbacterium* genus, despite their importance in agriculture, biomedicine, and biotechnology. Here, we elucidate the phylogenomic relationships between 125 *Microbacterium* cluster EA bacteriophages—including members from 11 sub-clusters (EA1 to EA11)—and infer their putative host ranges using insights from codon usage bias patterns as well as predictions from both exploratory and confirmatory computational methods. Our computational analyses suggest that cluster EA bacteriophages have a shared infection history across the *Microbacterium* clade. Interestingly, bacteriophages of all sub-clusters exhibit codon usage preference patterns that resemble those of bacterial strains different from ones used for isolation, suggesting that they might be able to infect additional hosts. Furthermore, host range predictions indicate that certain sub-clusters may be better suited in prospective biotechnological and medical applications such as phage therapy.

Keywords: bacteriophage; cluster EA; comparative genomics; host range



Citation: Milhaven, M.; Versoza, C.J.; Garg, A.; Cai, L.; Cherian, S.; Johnson, K.; Salas Perez, K.; Blanco, M.; Lobatos, J.; Mitra, C.; et al.

Microbacterium Cluster EA

Bacteriophages: Phylogenomic

Relationships and Host Range

Predictions. *Microorganisms* **2023**, *11*,

170. <https://doi.org/10.3390/microorganisms11010170>

Academic Editors: Igor V. Babkin and Grzegorz Węgrzyn

Received: 29 November 2022

Revised: 4 January 2023

Accepted: 5 January 2023

Published: 10 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

First discovered in the early 1900s, bacteriophages (i.e., viruses that infect, and ultimately kill, bacteria) are the most abundant biological entities on earth, with an estimated 10^{31} viral particles [1]. Yet, despite their abundance across environmental systems and their important impact on the evolution and community dynamics of the bacterial biosphere, much bacteriophage diversity remains uncharacterized to date.

Due to their host specificity and bactericidal nature, lytic bacteriophages are promising agents in many applications ranging from agriculture (e.g., to treat crops infected with pathogenic bacteria [2]) to biomedicine (e.g., to develop therapies for patients infected with multi-drug resistant *Microbacterium* strains [3]) to food safety (e.g., to prevent zoonotic pathogens in poultry, pork, beef, and fish [4]) and to wastewater treatment (e.g., to prevent sludge foaming [5]). Hence, gaining a better understanding of the genomic diversity of bacteriophages and the bacterial strains that they are able to infect is an important direction of current scientific research.

To aid in this endeavor, undergraduate researchers of the Science Education Alliance—Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) project have isolated, sequenced, and genomically characterized nearly 4000 bacteriophages over the last decade [6]. Of these, 514 bacteriophages infect Gram-positive, rod-shaped aerobic *Microbacterium* hosts that have been isolated from a variety of sources, including humans, where they can cause opportunistic infections in immuno-compromised individuals ([7]

and references therein), as well as food [8], soil [9], and plants [10,11] (for additional details, see the Actinobacteriophage Database: <https://www.phagesdb.org> (accessed on 29 November 2022) [12]).

Based on their nucleotide similarity, bacteriophages infecting microbacterial hosts can be classified into 15 clusters (clusters EA–EM and GA–GF), many of which are further separated into sub-clusters whose members share common genomic characteristics [13]. Among these 15 clusters, the cluster with the largest number of members, cluster EA (148 members), consists almost exclusively (~96%) of obligatory lytic bacteriophages isolated from *M. foliorum* NRRL B-24224, with the exception of bacteriophages Ixel, Leafus, Mercedes, and Nebulous isolated from *M. liquefaciens* LMG 16120 [14], Theresita isolated from *M. natoriense* ATCC BAA-1032 [13], and WilliamStrong isolated from *M. paraoxydans* NRRL B-14843. Consequently, the complete host range of bacteriophages from this representative cluster remains relatively poorly understood.

Bacteriophage isolation and cultivation remains the gold standard for characterizing bacteriophage–host interactions. At the “School of Life Sciences”; same time, experimental approaches are cost-, labor-, and time-intensive, thus limiting the number of studies that can feasibly be performed in many laboratories, especially as part of course-based undergraduate research programs with often limited budgets. However, recent advances in high-throughput sequencing technologies as well as associated bioinformatic methods have provided an alternative means to computationally predict bacteriophage–host interactions based on genomic features shared between bacteriophages and their hosts due to their co-evolution (see review of [15]). For example, as obligate parasites [16], bacteriophages require the bacterial host machinery to synthesize proteins—a strategy that is generally most efficient when the codon usage patterns of the bacteriophage closely match those of its host [17]. Consequently, patterns of codon usage bias (i.e., the preferential usage of synonymous codons) can provide important insights into the evolutionary relationships between bacteriophages and their hosts as well as regions of horizontal gene transfer (see review of [18]). To obtain a more complete picture, this information can then be combined with genome-wide levels of virus–host similarity, for example, based on the frequencies of different oligonucleotides in the virus and host genomes [19].

As part of a course-based undergraduate research experience at Arizona State University, we here characterize the phylogenomic relationships and computationally infer the putative host ranges of all cluster EA bacteriophages known to date. These novel insights will aid the future design of experimental assays to investigate bacteriophage–host relationships and may elucidate potential applications of cluster EA bacteriophages.

2. Materials and Methods

Publicly available whole-genome sequence data for 125 cluster EA bacteriophages were downloaded from NCBI GenBank (for accession numbers, see Supplementary Table S1) to characterize their phylogenomic relationships. In brief, a whole-genome multiple-sequence alignment (MSA) was generated between the bacteriophage genomes using the fast Fourier transform (FFT-NS-2) in MAFFT v.7.402 [20], pairwise average nucleotide identities (ANIs) calculated using the genome comparison tool in DNA Master v.5.23.6 and plotted in R v.4.0.2 [21] using the ggplot2 package [22]. In addition, a neighbor-joining tree was built from the MSA in MEGA v.11 [23] using a phylogeny test with 10,000 bootstrap replicates and subsequently visualized using FigTree v.1.43 (<http://tree.bio.ed.ac.uk/software/figtree/> (accessed on 29 November 2022)). For each group in the neighbor-joining tree, nucleotide sequence relatedness was analyzed using dot plots from Gepard v.2.1.0 [24].

To obtain insights into which *Microbacterium* species each bacteriophage might be able to infect, COUSIN v.0.4 [25] was used to determine the COdon Usage Similarity INDEX (COUSIN₅₉) for each of the 125 cluster EA bacteriophages (Supplementary Table S1) across 14 putative microbacterial host species (Supplementary Table S2), including *M. foliorum* B-24224 (i.e., the experimentally validated host for the majority of the isolated cluster

EA bacteriophages). Following Howell, Versoza et al. [26], host ranges were predicted using both exploratory and confirmatory methods—PHERI v.0.2 [27] and WIsH v.1.1 [19], respectively. PHERI is a machine-learning-based tool that capitalizes on protein sequence similarity, while WIsH utilizes the oligonucleotide frequency profiles of bacteriophages to predict prospective hosts. Thereby, likelihoods are calculated under trained Markov models, each corresponding to a potential host genome, and the model that produces the highest likelihood is determined to be the likely host. All software was executed using default settings.

3. Results

Comparative genomic analyses on 125 cluster EA bacteriophages—including 88 members of sub-cluster EA1, 7 of sub-cluster EA2, 3 of sub-cluster EA3, 8 of sub-cluster EA4, 7 of sub-cluster EA5, 4 of sub-cluster EA6, 1 of sub-cluster EA7, 1 of sub-cluster EA8, 2 of sub-cluster EA9, 3 of sub-cluster EA10, and 1 of sub-cluster EA11 (Supplementary Table S1)—demonstrated high levels of sequence relatedness in both the dot plot analysis (Supplementary Figure S1 and Supplementary Table S3) and the pairwise ANIs calculated between the cluster EA bacteriophages (Supplementary Figure S2), with groupings in the neighbor-joining tree (Figure 1) in agreement with previous cluster assignments [13]. Interestingly, however, sub-cluster EA1 appears further sub-divided into seven distinct groups as well as five singletons (Baines, Calix, Gelo, Nattles, and StingRay).

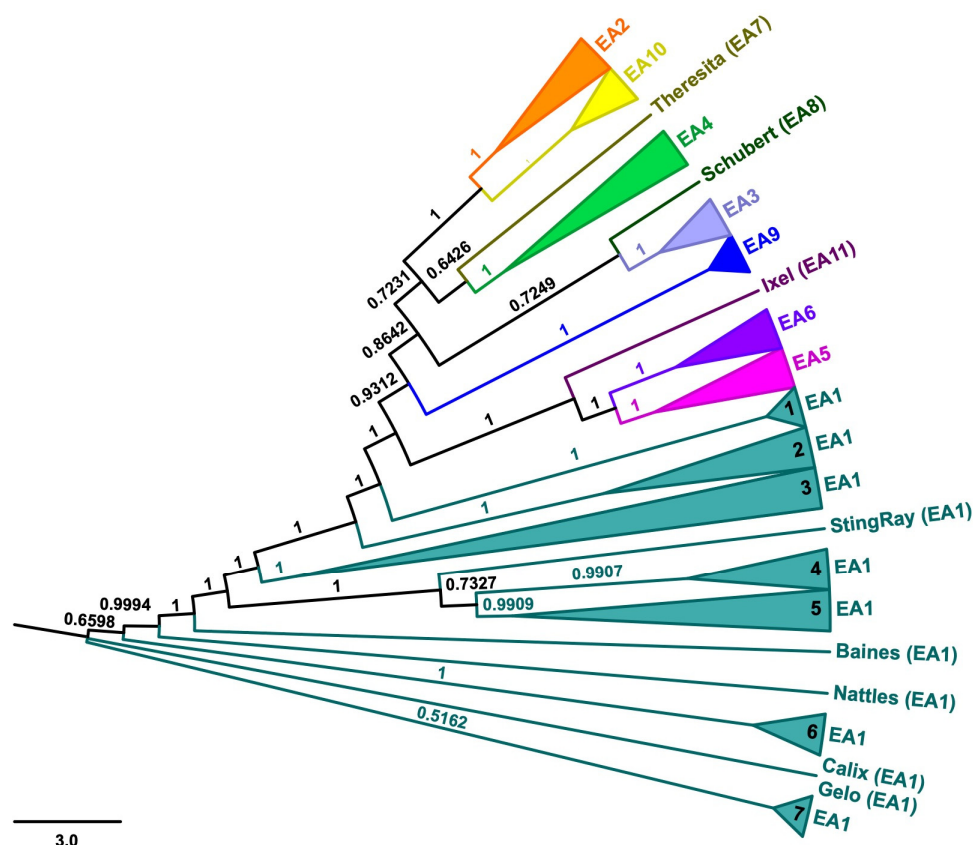


Figure 1. Neighbor-joining tree. Neighbor-joining tree generated using a multiple-sequence alignment of 125 cluster EA bacteriophage genomes (Supplementary Table S1) with 10,000 bootstrap replicates. Colors highlight bacteriophage membership in sub-clusters EA1–EA11. Numbers on the branches indicate bootstrap values.

Patterns of codon usage bias among the 125 bacteriophages show a strong agreement in codon usage preferences (CUPrefs) across the 14 potential *Microbacterium* hosts (Figure 2). Of particular note, bacteriophages of all sub-clusters more closely resemble CUPrefs of

bacterial strains different from those used for isolation, suggesting that they might be able to infect additional hosts. Specifically, bacteriophages from sub-clusters EA1, EA4–5, and EA8–10 exhibit CUPrefs that are most similar to *M. protaetiae*, EA2, EA6, and EA11 to *M. amylolyticum*, EA3 to *M. fandaimingii*, and EA7 to *M. liquefaciens*. Despite this, *M. foliorum* (the host used for isolation for the majority of the cluster EA bacteriophages) and *M. liquefaciens* (used to isolate Ixel, Leafus, Mercedes, and Nebulous) are consistently predicted as the most likely bacterial hosts across the cluster EA bacteriophages based on their similarity in oligonucleotide frequency profiles (Figure 3). Notably, members of the EA2, EA3, and EA8 (Schubert) sub-clusters have relatively low log likelihoods compared to the remainder of the sub-clusters, suggesting that more suitable hosts might exist outside of the 14 strains tested. This observation is further supported by the results of the exploratory host prediction analysis which highlighted that the members of these sub-clusters are also likely to infect hosts of the *Burkholderia*, *Mycolicibacterium*, and *Rhizobium* genera (Supplementary Table S4). In addition, these analyses suggest that Theresita—a sub-cluster EA7 singleton—likely exhibits a much broader host range across *Microbacterium* species.

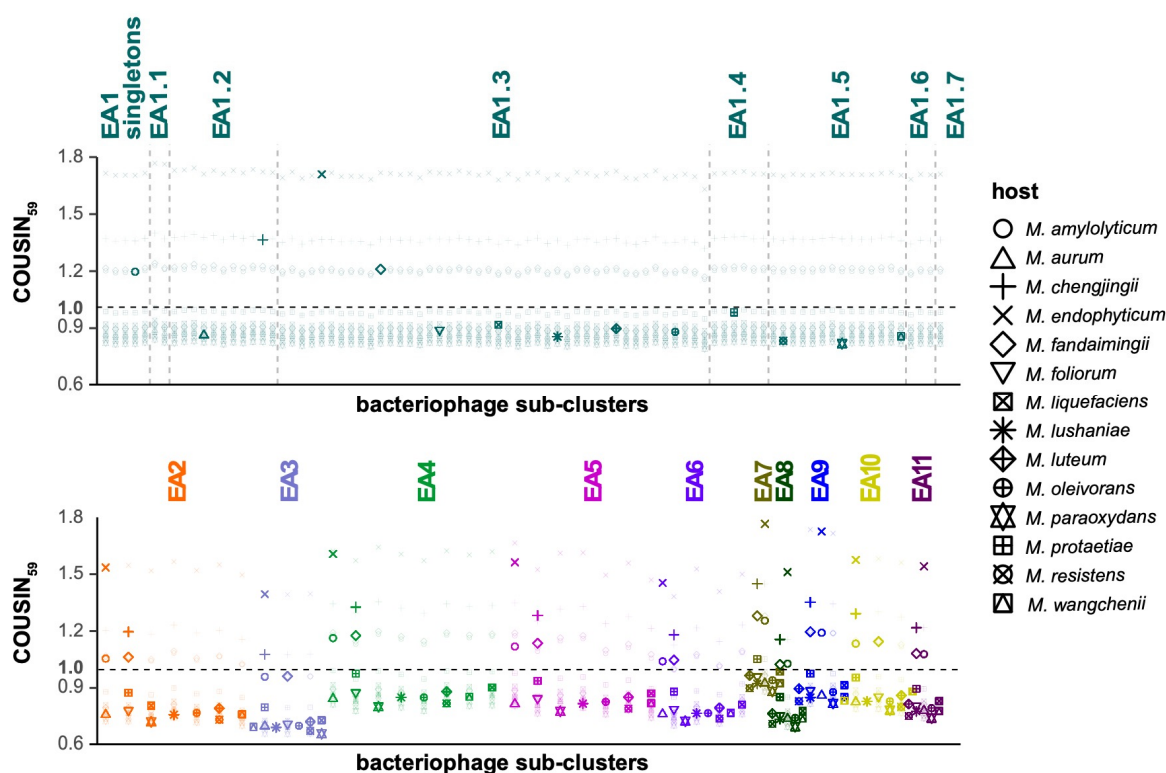


Figure 2. Codon usage bias. COdon Usage Similarity Index (COUSIN₅₉) of 125 cluster EA bacteriophages (Supplementary Table S1) across 14 potential *Microbacterium* host species (Supplementary Table S2), ordered by sub-cluster assignment of the bacteriophages. A COUSIN₅₉ score of 1 (horizontal black dashed line) indicates that a bacteriophage–host pair exhibits similar codon usage preferences. Colors highlight bacteriophage membership in sub-clusters EA1–EA11; vertical gray dashed lines indicate sub-cluster EA1 groupings in the neighbor-joining tree (Figure 1); shapes highlight *Microbacterium* species; individual bacteriophage–host COUSIN₅₉ scores are shown as transparent symbols and mean COUSIN₅₉ scores for each potential host in a sub-cluster are shown as non-transparent symbols.

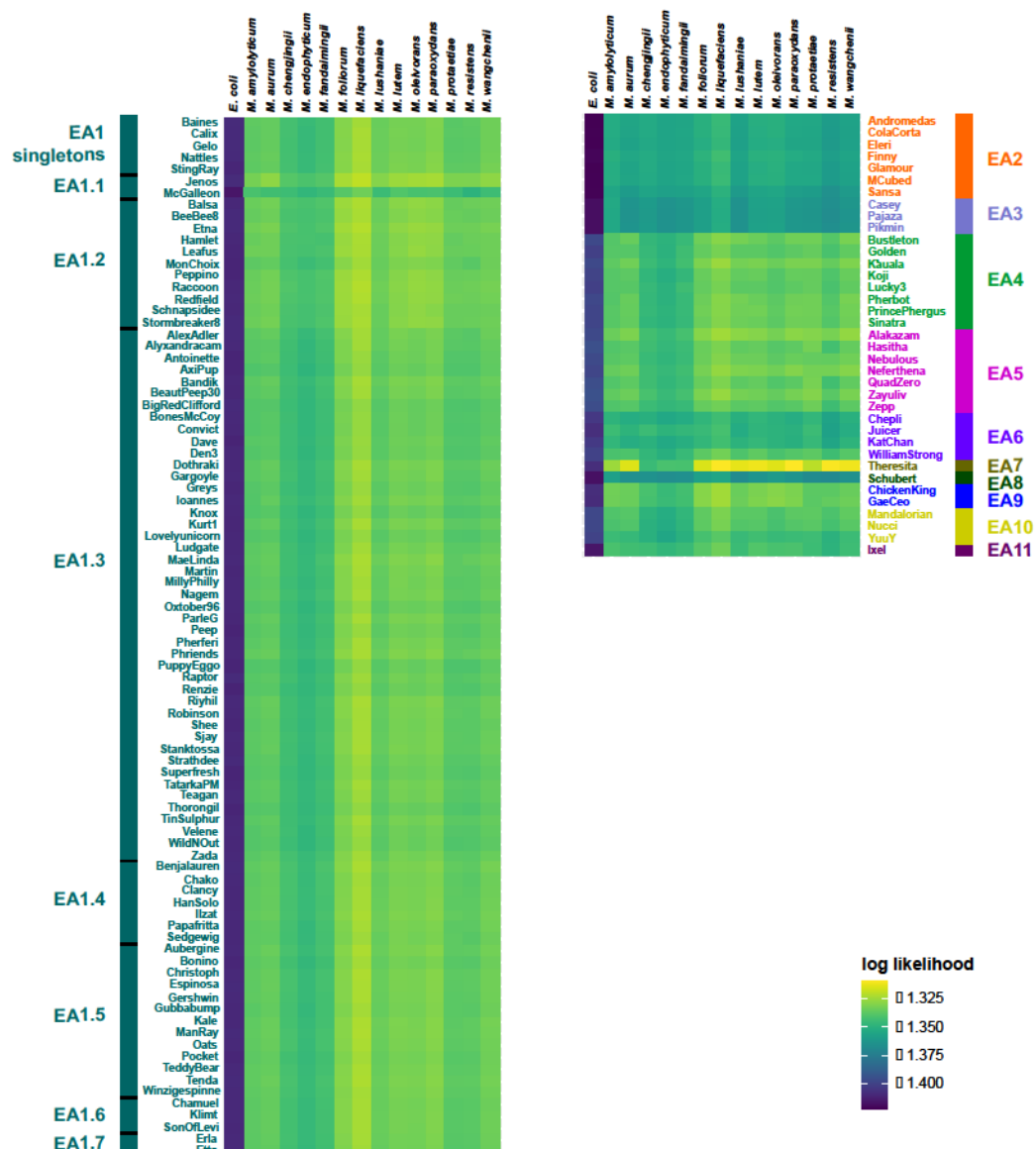


Figure 3. Predicted host ranges. Heatmap of log likelihoods of bacteriophage–host pairs, including 125 cluster EA bacteriophages (Supplementary Table S1), 14 potential *Microbacterium* host species (Supplementary Table S2), and *Escherichia coli* as a negative control as predicted by WISH. Colored names and bars highlight bacteriophage membership in sub-cluster EA1–EA11 groupings in the neighbor-joining tree (Figure 1); higher values in the heatmap correspond to more likely bacteriophage–host interactions.

4. Discussion

Bacteriophages are utilized for a variety of applications in agriculture, biomedicine, biotechnology, and diagnostics [28]. To effectively implement these bactericidal agents, it is crucial to understand phylogenomic relationships between bacteriophages and to ascertain their host ranges. Comparative genomic analyses confirmed previously established cluster memberships of the analyzed bacteriophages, yet the additional groupings of cluster EA1 bacteriophages in the neighbor-joining tree (Figure 1) suggest that relationships cannot be fully resolved at the sub-cluster level. With regards to host specificity, the reported codon usage bias (indicated by the COUSIN₅₉ index; Figure 2) demonstrates a shared pattern across bacteriophages, with clustering of CUPrefs scores in accordance with sub-cluster assignment. Strikingly, cluster EA bacteriophages harboring a median GC content of 63.4% (Supplementary Table S1) exhibit codon usage preferences that are similar to those

found in bacterial strains with GC contents ranging from 63.3% (*M. fandaimeingii*) to 68.3% (*M. liquefaciens*) (Supplementary Table S2). In contrast, patterns of codon usage bias across bacteriophages are distinctly different from the microbial strains with both the lowest and highest GC content—*M. chengjingii* (61.8%) and *M. endophyticum* (61.9%) as well as *M. wangchenii* (70.6%), *M. lushaniae* (70.7%), and *M. resistens* (71.3%)—suggesting that cluster EA bacteriophages are more likely to have shared an evolutionary infection history with the remainder of the host species. Consistent with the patterns of codon usage preference, and based on the oligonucleotide frequency similarity between bacteriophage and host genomes, the confirmatory tool WISH predicted host species with low and high GC content to be less likely hosts of cluster EA bacteriophages. At the same time, both confirmatory (WISH) and exploratory (PHERI) analyses predicted that several additional members of the *Microbacterium* genera might be potential hosts for cluster EA bacteriophages. These host range predictions are in agreement with experimentally validated results of successful *M. foliorum*, *M. liquefaciens*, and *M. paraoxydans* infections (Supplementary Table S1). While the majority of bacteriophages exhibit similar likelihood estimates across bacterial hosts, the observation of elevated log likelihood values suggests that certain bacteriophages should be favored in antimicrobial strategies. For example, a phage cocktail comprised of a combination of sub-cluster EA1 and EA7 members would allow for a broadening of host range when dealing with different *Microbacterium* species and strains. As such, bacteriophages of these sub-clusters will be important candidates for follow-up experimental validations of host specificity to aid future applications using bacteriophages of cluster EA as bactericidal agents.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms11010170/s1>, Figure S1: Dot plots; Figure S2: Average nucleotide identities; Table S1: Cluster EA bacteriophages included in the comparative analyses; Table S2: Host bacteria included in the comparative analyses; Table S3: Average nucleotide identities of representative cluster EA bacteriophages used in the dot plot analysis; Table S4: Exploratory host range prediction.

Author Contributions: Conceptualization and funding acquisition, S.P.P.; data curation, M.M.; formal analysis and visualization, M.M., C.J.V., A.G., L.C., S.C., K.J., K.S.P., M.B., J.L., C.M. and M.S.; M.M., C.J.V. and S.P.P. wrote this manuscript with input from all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a National Science Foundation CAREER grant to SPP, grant number DEB-2045343.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Genomic data for cluster EA bacteriophages and putative bacterial host species can be downloaded from NCBI GenBank using the accession numbers provided in Supplementary Table S1 and Supplementary Table S2, respectively.

Acknowledgments: We are grateful to Suhail Ghafoor for IT support.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Whitman, W.B.; Coleman, D.C.; Wiebe, W.J. Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6578–6583. [CrossRef] [PubMed]
- Svircev, A.; Roach, D.; Castle, A. Framing the future with bacteriophages in agriculture. *Viruses* **2018**, *10*, 218. [CrossRef] [PubMed]
- Kakasis, A.; Panitsa, G. Bacteriophage therapy as an alternative treatment for human infections. A comprehensive review. *Int. J. Antimicrob. Agents* **2019**, *53*, 16–21. [CrossRef] [PubMed]
- Monk, A.B.; Rees, C.D.; Barrow, P.; Hagens, S.; Harper, D.R. Bacteriophage applications: Where are we now? *Lett. Appl. Microbiol.* **2010**, *51*, 363–369. [CrossRef] [PubMed]
- Dyson, Z.A.; Tucci, J.; Seviour, R.J.; Petrovski, S. Lysis to Kill: Evaluation of the lytic abilities, and genomics of nine bacteriophages infective for *Gordonia* spp. and their potential use in activated sludge foam biocontrol. *PLoS ONE* **2015**, *10*, e0134512. [CrossRef]

6. Heller, D.; Sivanathan, V. Publishing student-led discoveries in genetics. *G3 Genes Genomes Genet.* **2022**, *12*, jkac141. [\[CrossRef\]](#)
7. Gneiding, K.; Frodl, R.; Funke, G. Identities of *Microbacterium* spp. encountered in human clinical specimens. *J. Clin. Microbiol.* **2008**, *46*, 3646–3652. [\[CrossRef\]](#)
8. McLean, R.A.; Sulzbacher, W.L. *Microbacterium thermosphactum*, spec. nov.; a nonheat resistant bacterium from fresh pork sausage. *J. Bacteriol.* **1953**, *65*, 428–433. [\[CrossRef\]](#)
9. Liu, J.; Nakayama, T.; Hemmi, H.; Asano, Y.; Tsuruoka, N.; Shimomura, K.; Nishijima, M.; Nishino, T. *Microbacterium natoriense* sp. nov., a novel d-aminoacylase-producing bacterium isolated from soil in natori, Japan. *Int. J. Syst. Evol. Microbiol.* **2005**, *55*, 661–665. [\[CrossRef\]](#)
10. Behrendt, U.; Ulrich, A.; Schumann, P. Description of *Microbacterium foliorum* sp. nov. and *Microbacterium phyllosphaerae* sp. nov., isolated from the phyllosphere of grasses and the surface litter after mulching the sward, and reclassification of *Aureobacterium resistens* (Funke et al. 1998) as *Microbacterium resistens* comb. nov. *Int. J. Syst. Evol. Microbiol.* **2001**, *51*, 1267–1276. [\[CrossRef\]](#)
11. Park, M.-J.; Kim, M.K.; Kim, H.-B.; Im, W.-T.; Yi, T.-H.; Kim, S.-Y.; Soung, N.-K.; Yang, D.-C. *Microbacterium ginsengisoli* sp. nov., a glucosidase-producing bacterium isolated from soil of a ginseng field. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 429–433. [\[CrossRef\]](#)
12. Russell, D.A.; Hatfull, G.F. PhagesDB: The actinobacteriophage database. *Bioinformatics* **2017**, *33*, 784–786. [\[CrossRef\]](#)
13. Jacobs-Sera, D.; Abad, L.A.; Alvey, R.M.; Anders, K.R.; Aull, H.G.; Bhalla, S.S.; Blumer, L.S.; Bollivar, D.W.; Bonilla, J.A.; Butela, K.A.; et al. Genomic diversity of bacteriophages infecting *Microbacterium* spp. *PLoS ONE* **2020**, *15*, e0234636. [\[CrossRef\]](#)
14. Frost, V.J.; Westover, K.M. Complete genome sequences of *Microbacterium liquefaciens* phages Mercedes, Leafus, Nebulous, and Ixel. *Microbiol. Resour. Announc.* **2021**, *10*, e00068-21. [\[CrossRef\]](#)
15. Versoza, C.J.; Pfeifer, S.P. Computational prediction of bacteriophage host ranges. *Microorganisms* **2022**, *10*, 149. [\[CrossRef\]](#)
16. Węgrzyn, G. Should bacteriophages be classified as parasites or predators? *Pol. J. Microbiol.* **2022**, *71*, 3–9. [\[CrossRef\]](#)
17. Esposito, L.A.; Gupta, S.; Streiter, F.; Prasad, A.; Dennehy, J.J. Evolutionary interpretations of mycobacteriophage biodiversity and host-range through the analysis of codon usage bias. *Microb. Genom.* **2016**, *2*, e000079. [\[CrossRef\]](#)
18. Thankeswaran Parvathy, S.; Udayasuriyan, V.; Bhadana, V. Codon usage bias. *Mol. Biol. Rep.* **2022**, *49*, 539–565. [\[CrossRef\]](#)
19. Galiez, C.; Siebert, M.; Enault, F.; Vincent, J.; Söding, J. WISH: Who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* **2017**, *33*, 3113–3114. [\[CrossRef\]](#)
20. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#)
21. Dalgaard, P.; R Development Core Team. R: A Language and Environment for Statistical Computing. 2010. Available online: <https://www.r-project.org/> (accessed on 29 November 2022).
22. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*, 2nd ed.; Springer International Publishing: Cham, Switzerland, 2016; ISBN 9783319242774.
23. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Krumsiek, J.; Arnold, R.; Rattei, T. Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **2007**, *23*, 1026–1028. [\[CrossRef\]](#)
25. Bourret, J.; Alizon, S.; Bravo, I.G. COUSIN (COdon Usage Similarity INdex): A normalized measure of codon usage preferences. *Genome Biol. Evol.* **2019**, *11*, 3523–3528. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Howell, A.A.; Versoza, C.J.; Cerna, G.; Johnston, T.; Kakde, S.; Karuku, K.; Kowal, M.; Monahan, J.; Murray, J.; Nguyen, T.; et al. Phylogenomic analyses and host range prediction of cluster P mycobacteriophages. *G3 Genes Genomes Genet.* **2022**, *12*, jkac244. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Baláz, A.; Kajsík, M.; Budiš, J.; Szemeš, T.; Turňa, J. PHERI—Phage Host Exploration Pipeline. *bioRxiv* **2020**. Available online: <https://www.biorxiv.org/content/10.1101/2020.05.13.093773v3> (accessed on 20 December 2022).
28. Gibb, B.; Hyman, P.; Schneider, C. The many applications of engineered bacteriophages—An overview. *Pharmaceuticals* **2021**, *14*, 634. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.