

Exploring Genomic Conservation in Actinobacteriophages with Small Genomes

Vincent Y. Tse¹, Haoyuan Liu¹, Andrew Kapinos¹, Canela Torres¹, Breanna Camille S. Cayabyab¹, Sarah N. Fett¹, Lucy G. Nakashima¹, Mujtahid Rahman¹, Aida S. Vargas¹, Krisanavane Reddi¹, Jordan Moberg Parker^{1,2}, Amanda C. Freise^{1*}

¹*Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles (UCLA), Los Angeles, CA, USA*

²*Current address: Department of Biomedical Science, Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, USA*

Keywords: bacteriophage, *Arthrobacter*, *Microbacterium*, *Gordonia*, *Streptomyces*, *Rhodococcus*, virology, microbiology, comparative genomics, genome size, SEA-PHAGES

*Corresponding Author:
Amanda C. Freise, afreise@ucla.edu

Abstract

Actinobacteriophages of a wide range of genome sizes continue to be isolated and characterized, but only a handful of these have atypically small genomes, defined in this work as genome sizes under 20,000 bp. These “small phages” are relatively rare and have received minimal study thus far. Of the actinobacteriophages published in PhagesDB.org, small phages have been isolated on *Arthrobacter*, *Gordonia*, *Rhodococcus*, and *Microbacterium* hosts. A previous study by Pope *et al.* showed that *Gordonia* small phages have similar gene products and amino acid sequences. Here, we set out to further examine relationships between small *Gordonia* phages as well as small phages that infect other hosts. Of the 3222 sequenced phages listed on PhagesDB, we identified 109 distinctly small phages with genome sizes under 20,000 bp. The majority of the small phages were isolated on *Arthrobacter* or *Microbacterium* hosts. Using comparative genomics, we searched for patterns of similarity among 34 cluster-representative small phages. Dot plot comparisons showed that there was more amino acid conservation than nucleotide identity amongst small phages. Gene content similarity (GCS) analysis revealed that the temperate *Gordonia* phages in Cluster CW share significant GCS values (over 35%) with the lytic *Arthrobacter* phages in Cluster AN, suggesting that some small phages have a considerable degree of genomic similarity with each other. SplitsTree analyses of shared phams (genes with substantial amino acid identity) supported the complexity of clustering criteria in small phages, given shuffling of genes across phages of different clusters and close relationships despite varied cluster membership. We observed this continuum of phage diversity through *Rhodococcus* phage RRH1’s closer similarity to phages in *Gordonia* subcluster

Genomic conservation in small phages

CW1 than CW1 is to *Gordonia* subcluster CW3. Finally, we were able to confirm the presence of conserved phams across not only small *Gordonia* phages but also within small phages from different clusters and hosts. Studying these genomic trends hidden in small phages allows us to better understand and appreciate the overall diversity of phages.

Introduction

Bacteriophages have been described as among the planet's most influential and are the most abundant biological entities on Earth, numbering up to an estimated 10^{31} individual phage particles and playing key roles in the environment [1,2]. The enormous diversity of phages has led to observations of phages that infect a wide variety of hosts, exhibit an array of life cycles, and display an assortment of genomes [3]. Over the years, phages of many different varieties continue to be isolated and added to PhagesDB, a database for actinobacteriophage research [4]. At the time of our study, we observed a sizable gap in the distribution of genome sizes on PhagesDB, wherein actinobacteriophages with genome sizes nearing 20,000 base pairs were separated from the rest of the phage database. As such, we ultimately designated these phages under 20,000 base pairs as atypically small phages. We have identified 109 small actinobacteriophages and characterized 34 of those phages as representative small phages. Few phages of this size have been analyzed, but due to their simplicity, understanding the overarching structure of these genomes can provide a solid foundation for contextualizing future phage research [5].

The goal of this study was to characterize these phages with smaller genomes and shed a clear light on patterns of genomic conservation that have previously been obscure. Bacteriophage genomes have often been shown to display a mosaic nature [6] and a continuum of diversity [7]; as atypically small phages have a very compact and simple genomic structure, we wanted to learn whether small phages share some of these very few genes with each other. A recent study showed that atypically small *Gordonia* phages have similar gene products and amino acid sequences [8] but whether

Genomic conservation in small phages

these *Gordonia* small phages display similarity with other small phages from different hosts and clusters was not clear. Another subsequent study likewise suggested that very small *Microbacterium* phages share similar genome architecture with small *Arthrobacter*, *Gordonia*, and *Rhodococcus* phages [9]. Further analysis of these phages, particularly of newly isolated small phages, can help reinforce genomic patterns that have already been observed and even identify new trends that may have been difficult to see.

In this study, we have identified and characterized atypically small actinobacteriophages. Through dot plot analysis, gene content similarity comparisons, and the construction of a Splitstree network phylogeny, we examined relationships between these small phages and were able to support the findings presented in the aforementioned studies. We additionally observed the conservation of several phams among otherwise diverse groups of small phages, such as genomic similarity between *Arthrobacter* phages, *Gordonia* phages, and RRH1, a *Rhodococcus* phage that is currently the smallest phage in the Actinobacteriophage database [5]. Overall, this indicates that genome mosaicism is prevalent even in small phages with few genes and supports the presence of a continuum of phage diversity.

Methods

Phage preparation, DNA extraction, and genome sequencing and annotation

Various small genome phages were isolated from soil samples and phage DNA was both purified and amplified according to the procedures described by SEA-PHAGES [10,11]. The phage genomes were then sequenced using Illumina-MiSeq and assembled as previously done [12] and auto-annotated using DNA Master [13], Glimmer [14], and GeneMark [15]. Various programs were used to manually verify the accuracy of the auto-annotation, such as Starterator [16], PhagesDB and NCBI BLASTp [4,17], NCBI Conserved Domain Database [18,19], and HHpred [20].

Comparative genomic analyses

A list of all 3222 phages and their genomic data as of November 2020 was obtained from PhagesDB (<https://phagesdb.org/data/>). The phages were plotted by increasing genome size. In order to determine the genome size cutoff for “small” phages, we identified a large gap in genome size from 19,679 bp to 28,876 bp and thus set 19,679 bp as the cutoff for “small” phages. Small phages were subsequently sorted based on their isolation type, host, morphotype, life cycle, GC content, and cluster. Representative small phages (Table 1) were selected such that as much diversity as possible was maintained from the aforementioned categories.

Nucleotide and amino acid FASTA files from each representative small phage were extracted from NCBI and imported into Gepard 1.40 [21] to produce dot plots. Furthermore, gene content similarity (GCS) values of the 34 representative phages were computed using the PhagesDB Explore Gene Content tool [4]. The GCS values of all 34 phages were entered into GraphPad Prism9 to generate a heatmap. Pham data

Genomic conservation in small phages

for all representative small phages was extracted from the Actino_Draft database (version 382) and imported into SplitsTree 4.16.2 [22]. Using default parameters, a network phylogeny was created for these representative small phages.

To expedite the process of genome structure comparison, 11 of the 34 representative phages were chosen in the same fashion as how the initial 34 were selected from the original 109 phages. These 11 representative phages were used to construct a comparative pham map. Genome maps for each of the 11 phages were downloaded on Phamerator and annotated on Inkscape 1.0 (<https://inkscape.org/>).

Results

109 bacteriophages were designated as “small genome phages” and 34 representative phages were selected for analysis.

To determine the genome size cut-off as the upper limit of “small” phages, a size continuum graph was constructed to list all the phages from PhagesDB in order of increasing genome size (Figure 1A). A large gap in genome size was observed between the 109th and 110th smallest phage as of November 2020; there were no phages between *Arthrobacter* Cluster FD phage Anjali (the 109th phage with 19,679 bp) and *Propionibacterium* Cluster BU phage Attacne (the 110th phage with 28,876 bp). Since this was a large gap of 10,000 bp, we set 19,679 bp as the upper cutoff for “small” phages, resulting in 109 small phages that have genome sizes ranging from 14,270 bp to 19,679 bp (Figure 1B). Of the 109 phages, 48 phages were isolated on *Arthrobacter*, 49 on *Microbacterium*, 9 on *Gordonia*, 2 on *Streptomyces*, and 1 on *Rhodococcus* (Supplementary Table S1). Of note, despite more than half of the 3,222 sequenced phages on PhagesDB that have been isolated on *Mycobacterium*, none of the 109 small phages infect *Mycobacterium* (the smallest phage known to infect *Mycobacterium* is 38,341 bp), suggesting that host type may be correlated with phage genome size.

In order to better visualize and understand the genomic similarities and differences among the 109 phages, we selected 34 of those as representative phages. The 34 phages selected are highlighted in the list of 109 small phages (Supplementary Table S1) and compiled into a chart (Table 1) that summarizes general phenotypic characteristics of each phage, including cluster, host, morphotype, life cycle, and genome size. Our list of phages includes representatives from Clusters AN, AX, BO,

Genomic conservation in small phages

CW, DM, EE, FD, FE, and GE that were isolated on *Arthrobacter*, *Streptomyces*, *Gordonia*, *Microbacterium*, or *Rhodococcus* hosts. The majority of the selected phages display *Siphoviridae* morphology and are lytic, except for Cluster BO *Streptomyces* phages that have *Tectiviridae* morphology [23] and Clusters CW and DM *Gordonia* phages that have a temperate life cycle.

Nucleotide dot plot comparisons suggested minimal sequence conservation across the 34 representative phages.

One method of determining the extent of relatedness within a group of phages is through the construction of a nucleotide dot plot, which allows qualitative comparison of nucleotide identity throughout the entire genome. In a comparison of the 34 representative small phages using this methodology, most observed similarity occurred between phages of the same cluster (Figure 2), which often have sizable regions of nucleotide identity [2]. Even then, certain phages from the same cluster such as the *Arthrobacter* phages from Cluster AX, *Gordonia* phages from Cluster CW, and *Arthrobacter* phages from Cluster FE which overall exhibited lower nucleotide conservation with other phages from their respective clusters.

Amino acid dot plot and GCS comparisons exhibited greater levels of conservation, including significant similarity between *Arthrobacter* and *Gordonia* phages.

Nucleotide comparisons can identify close relationships, but may miss more distant relationships between phages due to accumulated silent mutations and synonymous codon usage. A previous study found that *Arthrobacter* singleton

Genomic conservation in small phages

BlueFeather, Cluster FE, and former Cluster FI phages shared low nucleotide similarity but considerable amino acid identity, resulting in the consolidation of these phages into Cluster FE. Contrary to the nucleotide dot plot analysis of small genome phages, dot plot analysis using whole genome concatenated amino acid sequences displayed multiple regions of amino acid identity not only between phages from the same cluster but also among phages from different clusters (Figure 3). For example, Cluster AN *Arthrobacter* phages exhibited similarity with Cluster AX *Arthrobacter* phages, and singleton *Rhodococcus* phage RRH1 was likewise similar to Clusters CW and DM *Gordonia* phages. These similarities were not apparent in the nucleotide similarity analysis.

To determine whether there were any phams that were shared between these small phages given the degree of amino acid similarity, we calculated gene content similarity among the 34 representative small phages. Consistent with the amino acid dot plot, significant GCS values (over 35%) were observed between phages from the same cluster as well as between Clusters CW and DM *Gordonia* temperate phages and Clusters AX and AN *Arthrobacter* lytic phages (Figure 4). Furthermore, singleton *Rhodococcus* phage RRH1 exhibited significant GCS values with Clusters AX and AN *Arthrobacter* and Clusters CW and DM *Gordonia* phages, further corroborating the amino acid identity seen in the dot plot (Figure 3). On the other hand, Cluster BO *Streptomyces* phages share the least amount of phams with other small phages — only displaying very low GCS values with Cluster GE *Microbacterium* phage MuffinTheCat. Although minimal nucleotide sequence similarity was seen between the small phages, our GCS analysis illustrates that the majority of small phages we sampled do share certain phams.

Genomic conservation in small phages

To better understand the subtleties of genetic relationships between these various small phage groups, a SplitsTree gene content network phylogeny was constructed for the 34 representative phages (Figure 5). *Arthrobacter* Cluster FE phages appear to have longer branch lengths from their most recent internal node than phages in the fellow *Arthrobacter* Clusters AN, AX, and FD, indicating that Cluster FE is one of the more diverse groups of *Arthrobacter* small phages, in correlation with its composition of phages that were originally in distinct clusters [24]. Likewise, the branches that include *Gordonia* Clusters CW and DM phages as well as singleton *Rhodococcus* RRH1 are highly diverse as well. In fact, based on the branches of this SplitsTree, the phages GRU3, GMA5, and Coeur (all subcluster CW2) and *Gordonia* phages McGonagall and Jeanie (all subcluster CW1) are more genetically similar to the *Rhodococcus* singleton phage RRH1 than to Schiebs (subcluster CW3). This is an instance in which phages from one cluster are more similar to non-cluster members than to phages within their cluster.

Pham maps suggested conservation of terminase and other genes among some small phages.

As previous studies suggested the conservation of subunits of the same terminase phams in atypically small *Gordonia* phages [8], we performed a Phamerator map analysis to see if terminase and other additional genes would be conserved in other small phages as well. Indeed, the *Gordonia* small phages SallySpecial (Cluster DM) and Schiebs (Cluster CW) displayed conservation of the terminase phams. We noticed this two-unit terminase conservation not only among these atypically small *Gordonia* phages, but also with singleton *Rhodococcus* phage RRH1. Furthermore,

Genomic conservation in small phages

these phages share significant GCS values, suggesting that despite their differing hosts, phage RRH1 is considerably similar to these *Gordonia* phages. Despite sharing no phams with any of the other phages in our Phamerator map analysis (Supplementary Figure S1), *Microbacterium* Cluster EE phage Lunatic shares its overall genome structure with the majority of other phages explored herein; all have a terminase protein, a portal protein, a capsid or protease fusion protein, and a tape measure protein, which can be expected given that these genes encode important functional components of the phage particle. Lunatic is not the only phage to exhibit similar genome architecture with small phages; other *Microbacterium* phages have also shared similar genome architecture to this group of atypically small phages [9].

Discussion

The focus of this study was to characterize and better understand the intricate relationships between phages with small genomes. This was prompted by the isolation of phages Whytu and BlueFeather, both *Arthrobacter* phages that displayed extraordinarily small genome sizes. Our analyses of these 34 representative small phages, including Whytu and BlueFeather, showed minimal nucleotide similarity but greater amino acid and gene content similarity. Strong amino acid similarity in the absence of nucleotide similarity suggests that small actinobacteriophages have diversified enough that only distant relationships remain, perhaps due to gradual purifying selection of essential genes over their evolutionary history [25]. Phamerator map analysis of 11 of the 34 selected small phages further supports this point, as we observed intracluster conservation of phams between small phages that infect different hosts such as *Arthrobacter*, *Gordonia*, and *Rhodococcus*. The layout of genes and overall genome architecture were similar on an intercluster level, even across phages that did not share any phams, such as the *Microbacterium* phage Lunatic. This synteny in genomic architecture across otherwise distantly related phages has been observed in larger phages [9,26–28].

In our comparison of genome sizes and the isolation hosts of small phages, we observed that none of the 109 small phages were isolated on *Mycobacterium*, and that the smallest phage known to infect *Mycobacterium* was 38,341 bp long. This led us to ask several questions, such as why none of the 109 small phages in our study were isolated on *Mycobacterium*, and whether larger phages could have progressively evolved from smaller phages through direct acquisition of genes or through alternative

Genomic conservation in small phages

mechanisms [29]. One study has suggested that some phams only present in one genome within a cluster may have been acquired by horizontal gene transfer [27], and another has suggested that larger phages may have progressively evolved from smaller phages through acquisition of genes via horizontal gene transfer [29]. Further research into an alternative hypothesis, such as whether reductive evolution in larger phages could have led to the loss of non-essential genes in small phages, would also provide more insight on the relationships between small and large phages.

Previous clustering methods have sorted phages with more than 50% nucleotide identity into the same cluster [27,28] and were later revised to utilize a 35% shared gene content parameter instead [8]. For instance, former Cluster FI phages and the singleton Phage BlueFeather have recently been combined with Cluster FE phages to form a single large cluster, Cluster FE. Therefore, it was expected that the branch lengths of the newly expanded Cluster FE in our SplitsTree analysis would be varied. On the other hand, branch lengths of the fellow *Arthrobacter* Cluster AN were extremely short and thus member phages were more similar. This variability in phage cluster diversity is observable in small phages across *Microbacterium*, *Gordonia*, and *Rhodococcus* hosts as well, and is an indicator of the continuum of actinobacteriophage diversity [8,9].

In a previous study, conservation of the large terminase subunits was observed among the atypically small *Arthrobacter*, *Gordonia*, and *Rhodococcus* phages studied [8]. Our Phamerator map analysis of 11 representative phages supported those findings, as the same terminase phams were present in *Gordonia* phages SallySpecial and Schiebs from Clusters DM and CW respectively. We furthermore noticed those same terminase phams present in the singleton *Rhodococcus* phage RRH1. Even more

Genomic conservation in small phages

interestingly, we observed the same portal protein pham in the three *Gordonia* and *Rhodococcus* phages mentioned above in addition to the *Arthrobacter* phages Atraxa (Cluster AX) and Decurro (Cluster AN). These phages not only share these phams but also have a similar genome architecture with each other on a whole; this may be evidence for horizontal gene transfer that allowed these similar genes to be present in phages from a wide array of clusters and hosts [30]. This is perhaps unsurprising, as terminase and portal proteins tend to be more conserved than other genes [31], but these results are nonetheless valuable for further understanding the conservation of critical genes among small phages. Given the multiple different forms of evidence presented in this study through GCS, genome architecture, and conservation of phams, it seems that the *Rhodococcus* phage RRH1 may be more closely related to *Gordonia* phages than previously thought. While genomic similarity does not necessarily imply overlapping host ranges [32], we suggest host range experiments be conducted on RRH1 in light of its similarity to the *Gordonia* phages analyzed herein. It is possible that, since the *Rhodococcus* phage RRH1 shares so much in common with the *Gordonia* phages, it may be able to infect *Gordonia* hosts as well. The reverse situation may also be promising: to explore if the small *Gordonia* phages listed above are also able to infect *Rhodococcus*.

Overall, this study has defined a specific group of phages as so-called “small phages”: those with genome sizes that are under 20,000 base pairs and thus at least 10,000 bp smaller than every other actinobacteriophage sequenced to date on PhagesDB. We explored several patterns of conservation in their genomes among the various phage hosts and clusters represented. In addition, we verified low nucleotide identity but substantial amino acid and gene content similarity between some of these

Genomic conservation in small phages

phages and also identified previously hidden relationships between the small *Rhodococcus* phage RRH1 and other small *Gordonia* and *Arthrobacter* phages, a testament to the complex dynamics that exist even among phages with such extraordinarily small genomes. This study only examined 109 small phages available on PhagesDB at the time, a mere handful of all the small phages that exist in the natural world. As more small phages continue to be collected and sequenced, further analysis of small phages will undoubtedly shed light on the concealed trends that define the nature of phage relationships.

Acknowledgements

We thank Rebecca A. Garlena and Daniel A. Russell at the Pittsburgh Bacteriophage Institute for genome sequence and assembly and Travis Mavrich, Welkin Pope, Debbie Jacobs-Sera, and Graham Hatfull with the HHMI Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program for programmatic support. This research was supported in part by the Department of Microbiology, Immunology, and Molecular Genetics and the Dean of Life Sciences Division at UCLA. The authors acknowledge the use of instruments at the Electron Imaging Center for NanoMachines supported by NIH (1S10RR23057 to ZHZ) and CNSI at UCLA.

Author Contributions

V.Y.T. and H.L. conceptualized the paper; V.Y.T., H.L., A.S.V., B.C.C., L.G.N., M.R., and S.N.F. performed experiments and drafted the paper; V.Y.T., H.L., A.K., A.C.F., and J.M.P. revised the paper; A.C.F. and J.M.P. supervised the research. All authors have reviewed and approved of this manuscript prior to submission.

This project was supported by the Microbiology, Immunology & Molecular Genetics Department and the Dean of Life Sciences Division at UCLA, with additional support for sequencing from the HHMI Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program.

Author Disclosure Statement

The authors declare that there is no conflict of interest regarding the publication of this article.

Figures

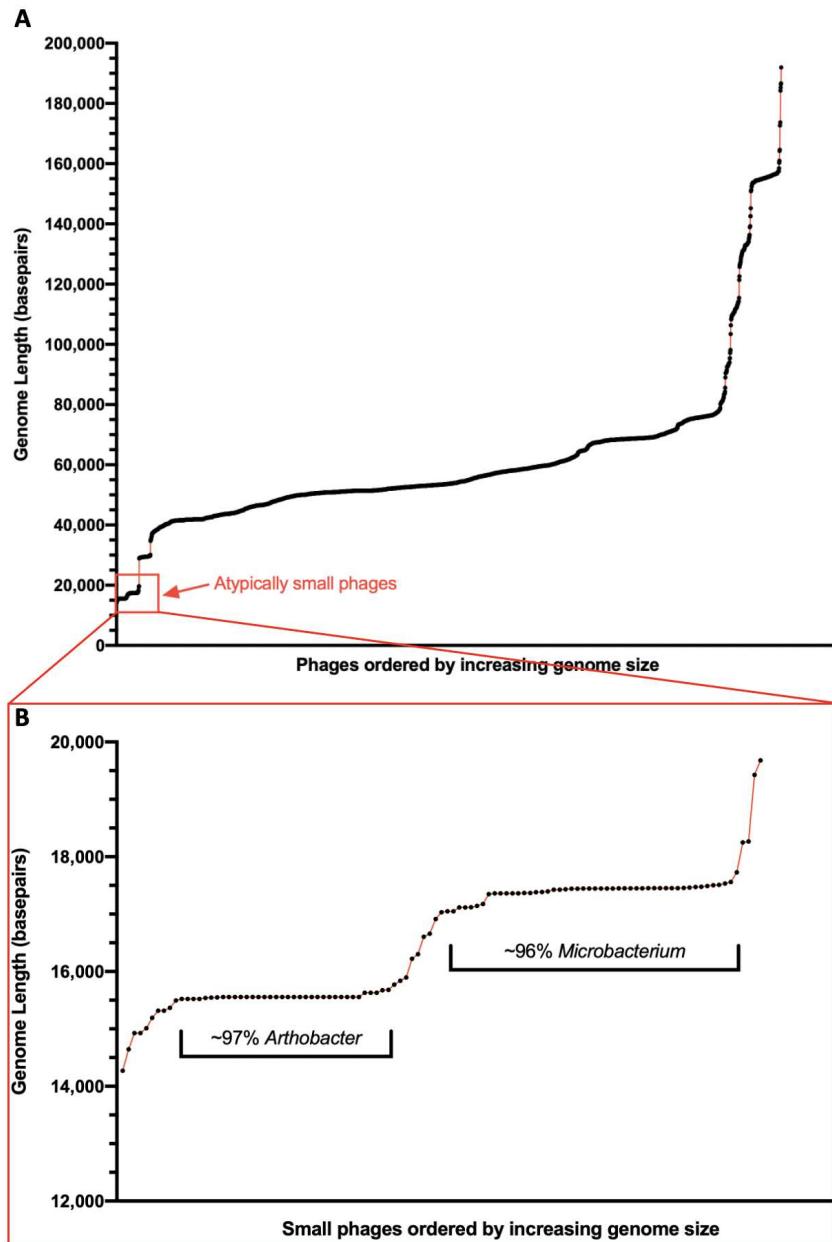


Figure 1. Atypically small phages primarily infect *Arthrobacter* and *Microbacterium*.

(A) All 3222 sequenced phages (excluding draft phages) listed on PhagesDB in November 2020 are plotted by increasing genome size.

(B) An enlarged graph of the atypically small phages region in panel A reveals 109 distinct phages, which primarily infect *Arthrobacter* and *Microbacterium*. Each black dot represents a distinct phage.

Genomic conservation in small phages

Table 1. 34 representative phages from each cluster of small genome phages show immense physical and genomic diversity.

Phage ID	Accession	Cluster	Host	Morphotype	Life Cycle	Genome Size (bp)
Seume	MF140426	AN	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	15319
Toulouse	KU160670	AN	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	15319
Decurro	KT355471	AN	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	15524
Moloch	KU160657	AN	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	15630
Stratus	KU160667	AN	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	15630
Swenson	MF140429	AN	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	15680
Atraxa	MH834597	AX	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	14927
Sputnik	MH834628	AX	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	14927
Adaia	MH834594	AX	<i>Arthrobacter</i>	nd	Lytic	15840
Forthebois	MK620900	BO	<i>Streptomyces</i>	<i>Tectiviridae</i>	nd	18251
WheeHeim	MK305890	BO	<i>Streptomyces</i>	<i>Tectiviridae</i>	nd	18266
Jeanie	KU998256	CW1	<i>Gordonia</i>	<i>Siphoviridae</i>	Temperate	17118
McGonagall	KU998255	CW1	<i>Gordonia</i>	nd	Temperate	17119
Coeur	MK801723	CW2	<i>Gordonia</i>	nd	Temperate	16223
GMA5	KR053198	CW2	<i>Gordonia</i>	nd	Temperate	17562
GRU3	KR053197	CW2	<i>Gordonia</i>	nd	Temperate	17727
Schiebs	MN234213	CW3	<i>Gordonia</i>	<i>Siphoviridae</i>	Temperate	14646
SallySpecial	MG812496	DM	<i>Gordonia</i>	<i>Siphoviridae</i>	Temperate	15896
Emperor	MH271296	DM	<i>Gordonia</i>	nd	Temperate	16604
EpicDab	MK660712	DM	<i>Gordonia</i>	nd	Temperate	16658
Lunatic	nd	EE	<i>Microbacterium</i>	nd	Lytic	16915
HarperAnne	MN586010	EE	<i>Microbacterium</i>	nd	Lytic	17116
Bullzi2019	MT498067	EE	<i>Microbacterium</i>	<i>Siphoviridae</i>	Lytic	17143
VitulaEligans	MH371124	EE	<i>Microbacterium</i>	<i>Siphoviridae</i>	Lytic	17534
Mendel	MK016500	FD	<i>Arthrobacter</i>	nd	Lytic	19428
Anjali	MK016490	FD	<i>Arthrobacter</i>	nd	Lytic	19679
Yavru	MT889364	FE	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	15193
Whytu	MT024870	FE	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	15369
Noely	MH834622	FE	<i>Arthrobacter</i>	nd	Lytic	15013
Corgi	MH834607	FE	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	15771
Idaho	MK757448	FE	<i>Arthrobacter</i>	nd	Lytic	15675
BlueFeather	MT024867	FE	<i>Arthrobacter</i>	<i>Siphoviridae</i>	Lytic	16302
MuffinTheCat	MT952848	GE	<i>Microbacterium</i>	nd	nd	15494
RRH1	NC_016651	Singleton	<i>Rhodococcus</i>	<i>Siphoviridae</i>	nd	14270

nd = not determined.

Phages are ordered based on appearance in the dot plots in Figures 2 and 3.

Genomic conservation in small phages

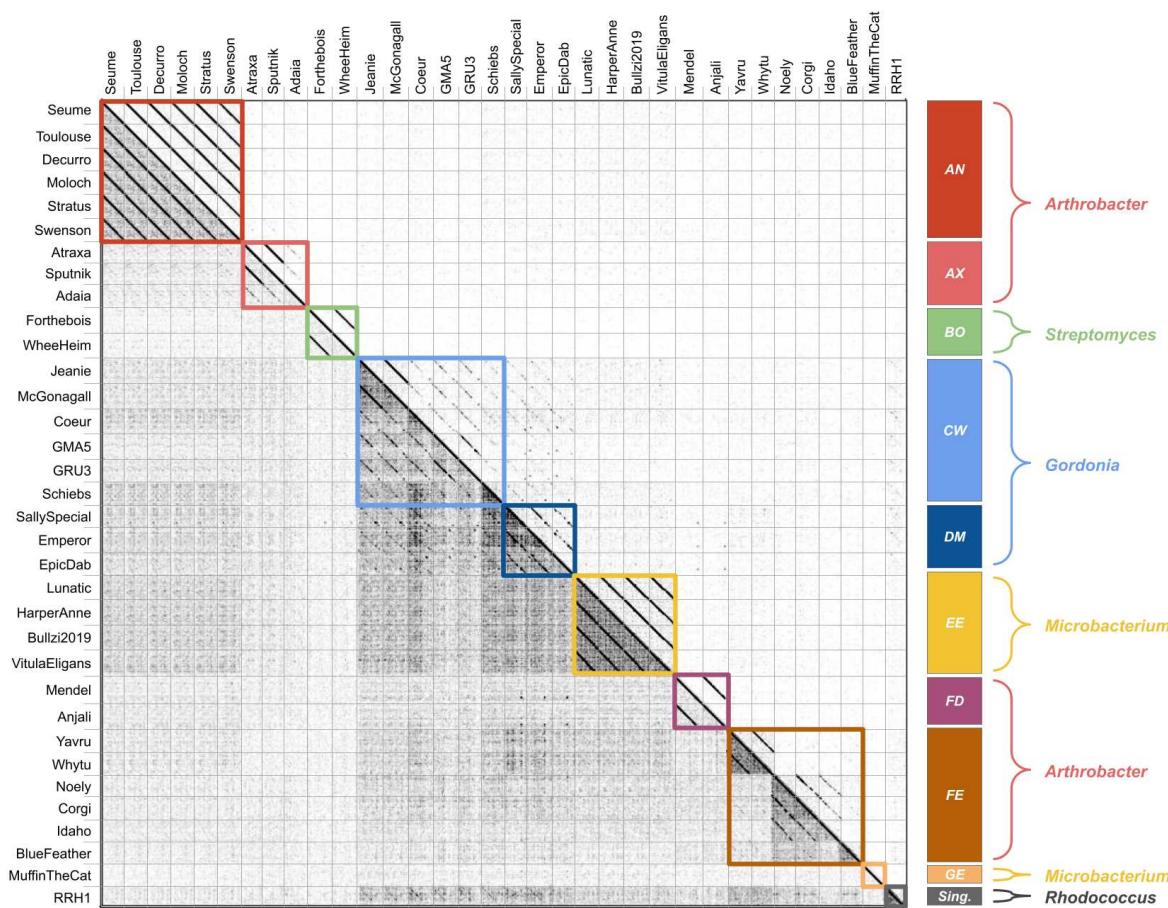


Figure 2. Nucleotide dot plot reveals minimal nucleotide sequence conservation among 34 representative small bacteriophages.

Using the software Gepard 1.40, a dot plot was generated to observe alignments among the whole nucleotide genomes of 34 representative small bacteriophages with a word size of 12 on the right side, and word size of 10 on the left side. Minimal alignments were observed between phages of different clusters. The *Arthrobacter* Clusters AX and FE, as well as the *Gordonia* Cluster CW, contained select phages that had strong alignments to some but not all other members of their individual clusters.

Genomic conservation in small phages

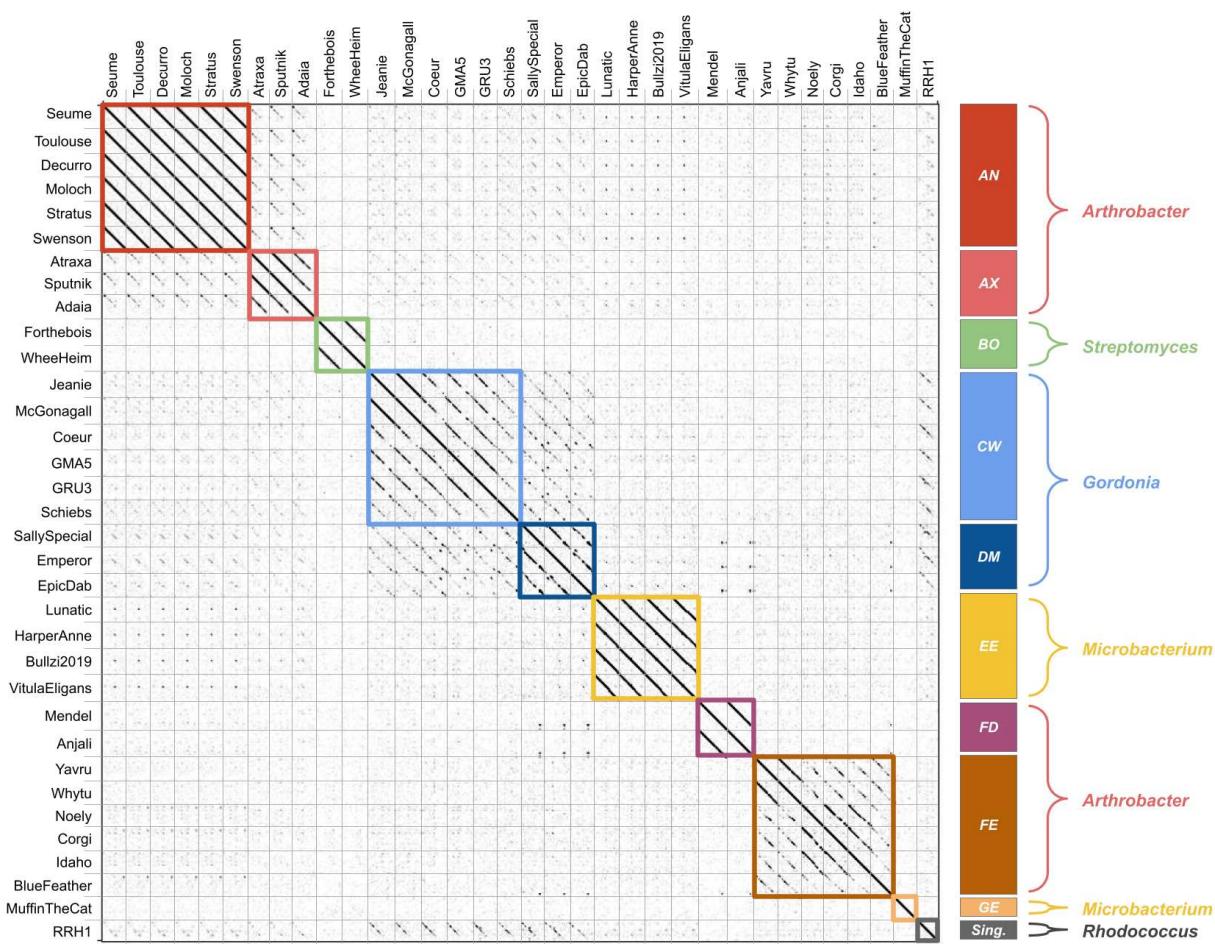


Figure 3. Amino acid dot plot exhibits substantial amino acid conservation among small phages.

Using the software Gepard 1.40, a dot plot was generated to observe alignments among the concatenated complete amino acid sequences of 34 representative small phages using a word size of 5. *Arthrobacter* Clusters AX and FE, as well as the *Gordonia* Cluster CW, contained phages that all had moderate to strong alignments to all other represented phages within their clusters. Furthermore, alignments were observed between phages belonging to different clusters. Most notably, the *Rhodococcus* phage RRH1 exhibits significant alignment to many *Gordonia* phages in Cluster CW and DM.

Genomic conservation in small phages

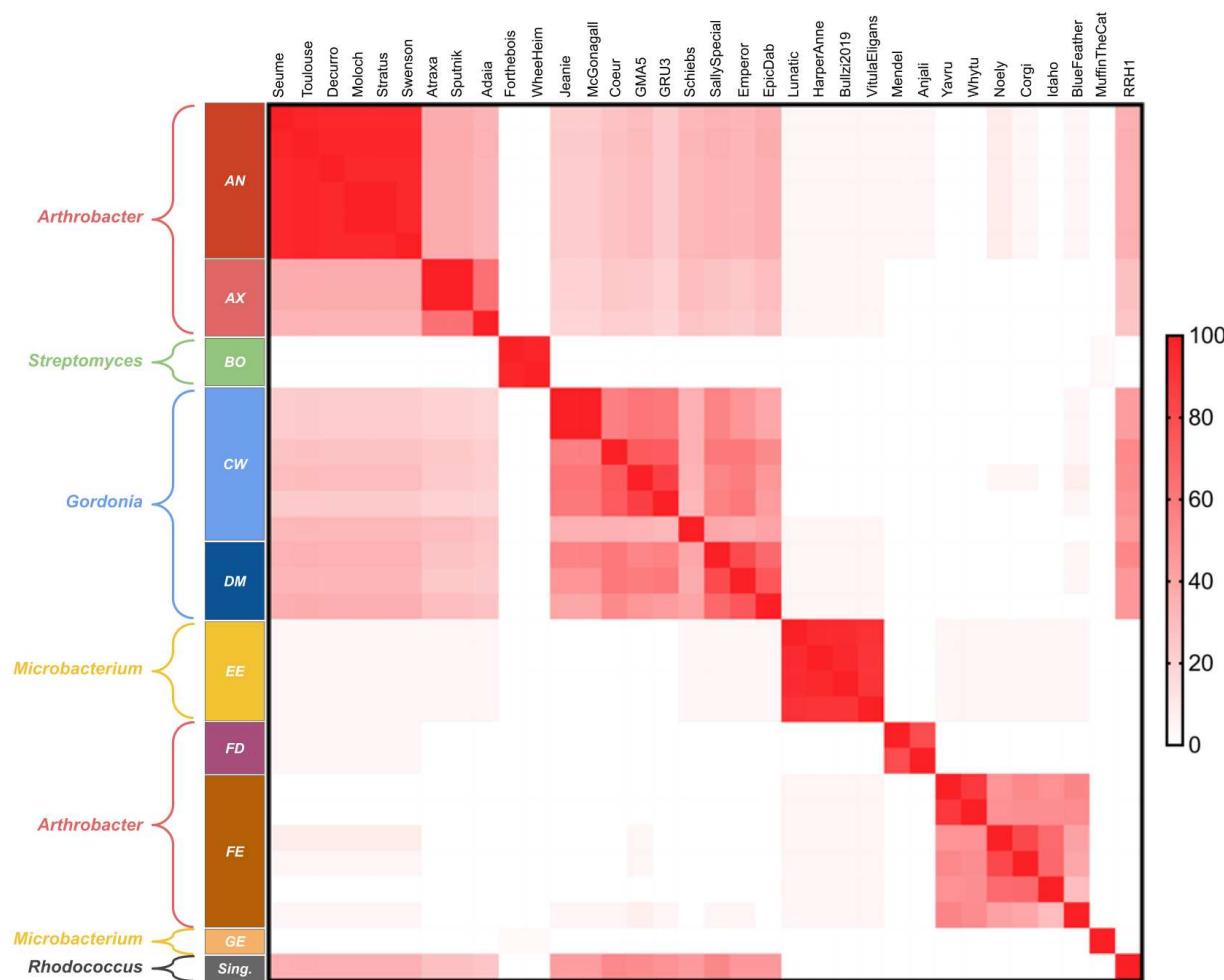


Figure 4. Shared gene content between 34 representative small bacteriophages.

Using Prism9, a heatmap was generated to visualize pairwise GCS comparisons among all of the 34 representative small phages. Clusters CW and DM *Gordonia* temperate phages show significant GCS values (over 35%) with Clusters AX and AN *Arthrobacter* lytic phages. *Rhodococcus* phage RRH1 displays significant GCS values with Clusters CW and DM *Gordonia* phages and Clusters AX and AN *Arthrobacter* phages.

Genomic conservation in small phages

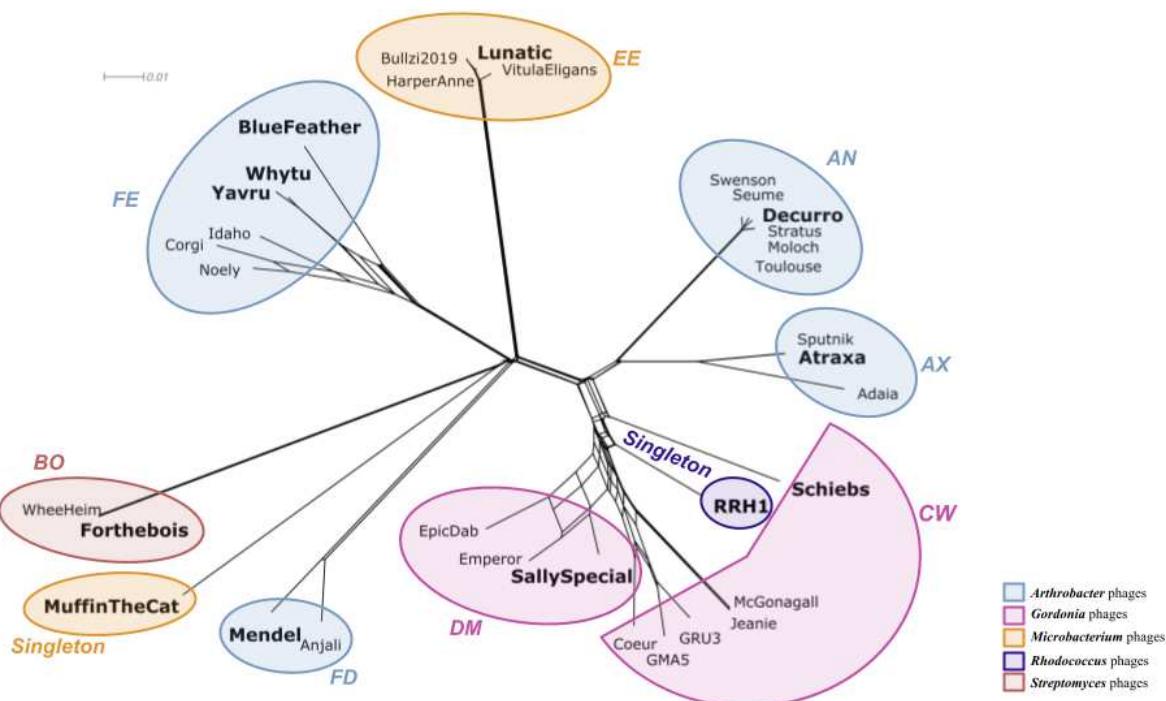


Figure 5. SplitsTree analysis unveils a continuum of diversity among small genome *Arthrobacter* and *Gordonia* phages.

The phages in *Arthrobacter* Cluster FE have longer branch lengths than phages in the fellow *Arthrobacter* Clusters AN, AX, and FD, indicating that Cluster FE is one of the more diverse groups of *Arthrobacter* phages. The *Gordonia* phages GRU3, GMA5, and Coeur are part of the subcluster CW2 and phages McGonagall and Jeanie are part of the subcluster CW1; yet they are more genetically similar to the singleton RRH1 than to Schiebs, which belongs to the subcluster CW3. The bolded phages are representative phages for each cluster and appear in the pham map (Supplementary Figure S1).

Supplementary Figures

Supplementary Table S1. 109 atypically small genome non-draft phages extracted from PhagesDB.

Accession	Phage ID	Year Found	Cluster (Subcluster)	Host Genus	Host Species	Host Strain	Morphotype	Life Cycle	Genome Size (bp)	GC %
NC_016651	RRH1	2012	Singleton	<i>Rhodococcus</i>	<i>rhodochrous</i>	<i>Rrho39 (DSMZ43241)</i>	<i>Siphoviridae</i>	nd	14270	68.4
MN234213	Schiebs	2016	CW (CW3)	<i>Gordonia</i>	<i>rubripertincta</i>	<i>NRRL B-16540</i>	<i>Siphoviridae</i>	Temperate	14646	70.6
MH834597	Atraxa	2017	AX	<i>Arthrobacter</i>	<i>sulfureus</i>	<i>NRRL B-14730</i>	<i>Siphoviridae</i>	Lytic	14927	58.0
MH834628	Sputnik	2017	AX	<i>Arthrobacter</i>	<i>sulfureus</i>	<i>NRRL B-14730</i>	<i>Siphoviridae</i>	Lytic	14927	58.0
MH834622	Noely	2017	FE	<i>Arthrobacter</i>	<i>globiformis</i>	<i>B-2979</i>	nd	Lytic	15013	68.3
MT889364	Yavru	2019	FE	<i>Arthrobacter</i>	<i>globiformis</i>	<i>B-2979</i>	<i>Siphoviridae</i>	Lytic	15193	64.3
MF140426	Seume	2013	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15319	60.3
KU160670	Toulouse	2013	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15319	60.3
MT024870	Whytu	2019	FE	<i>Arthrobacter</i>	<i>globiformis</i>	<i>B-2979</i>	<i>Siphoviridae</i>	Lytic	15369	64.8
MT952848	MuffinTheCat	2019	Singleton	<i>Microbacterium</i>	<i>testaceum</i>	<i>NRRL B-24232</i>	nd	nd	15494	55.1
MF140417	Link	2014	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15521	60.2
KT355471	Decurro	2014	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15524	60.2
KU160658	Muttlie	2014	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15524	60.2
KU160674	Yank	2014	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15524	60.2
MF140415	KylieMac	2015	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15540	59.8
MF140431	Taj14	2015	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15546	59.9
MN617842	Saphira	2015	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15549	60.0
MK737943	Arby	2017	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
MH576948	Azathoth	2017	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
MK494103	Blair	2015	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	59.6
MH576950	CGermain	2017	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
KY434670	Chestnut	2015	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
MH576951	Copper	2017	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
KX443695	Courtney3	2014	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
MH576952	Dewayne	2017	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
MF140409	Elkhorn	2015	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	59.6
MH779506	Guntur	2017	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
MH576955	Hunnie	2017	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
MH576957	Inspire2	2017	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
KT355473	Jessica	2014	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
MK820640	Laila	2015	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	59.6
MF140419	Lore	2016	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	59.6
MN813695	LouisXIV	2017	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
KU160655	Maggie	2013	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
MF140420	Mariposa	2015	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
KX576642	Massimo	2015	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
KX610765	Prospero	2015	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1
MH576961	Ronnie	2017	AN	<i>Arthrobacter</i>	sp.	<i>ATCC 21022</i>	<i>Siphoviridae</i>	Lytic	15556	60.1

Genomic conservation in small phages

MK494111	StewieGriff	2015	AN	<i>Arthrobacter</i>	sp.	ATCC 21022	<i>Siphoviridae</i>	Lytic	15556	59.6
MF140433	TinoCrisci	2016	AN	<i>Arthrobacter</i>	sp.	ATCC 21022	<i>nd</i>	Lytic	15556	60.1
KT783672	TymAbreu	2013	AN	<i>Arthrobacter</i>	sp.	ATCC 21022	<i>Siphoviridae</i>	Lytic	15556	60.1
KU160657	Moloch	2013	AN	<i>Arthrobacter</i>	sp.	ATCC 21022	<i>Siphoviridae</i>	Lytic	15630	60.0
KT355475	Sandman	2013	AN	<i>Arthrobacter</i>	sp.	ATCC 21022	<i>Siphoviridae</i>	Lytic	15630	60.0
KU160667	Stratus	2014	AN	<i>Arthrobacter</i>	sp.	ATCC 21022	<i>Siphoviridae</i>	Lytic	15630	60.0
MK757448	Idaho	2017	FE	<i>Arthrobacter</i>	sp.	ATCC 21022	<i>nd</i>	Lytic	15675	63.6
MF140429	Swenson	2015	AN	<i>Arthrobacter</i>	sp.	ATCC 21022	<i>Siphoviridae</i>	Lytic	15680	59.9
MH834607	Corgi	2017	FE	<i>Arthrobacter</i>	<i>globiformis</i>	B-2979	<i>Siphoviridae</i>	Lytic	15771	67.6
MH834594	Adaia	2017	AX	<i>Arthrobacter</i>	<i>sulfureus</i>	NRRL B-14730	<i>nd</i>	Lytic	15840	56.1
MG812496	SallySpecial	2016	DM	<i>Gordonia</i>	<i>terrae</i>	3612	<i>Siphoviridae</i>	Temperate	15896	70.1
MK801723	Coeur	2016	CW (CW2)	<i>Gordonia</i>	<i>terrae</i>	3612	<i>nd</i>	Temperate	16223	67.9
MT024867	BlueFeather	2019	FE	<i>Arthrobacter</i>	<i>globiformis</i>	B-2979	<i>Siphoviridae</i>	Lytic	16302	64.3
MH271296	Emperor	2016	DM	<i>Gordonia</i>	<i>terrae</i>	3612	<i>nd</i>	Temperate	16604	70.1
MK660712	EpicDab	2018	DM	<i>Gordonia</i>	<i>neofelifaecis</i>	NRRL 59395	<i>nd</i>	Temperate	16658	69.1
<i>nd</i>	Lunatic	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	16915	68.8
MH779512	Miaurora	2017	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17032	69.0
MT310864	Nobel	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17049	69.0
MT498047	TinyTruffula	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17050	69.0
MN586010	HarperAnne	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17116	68.8
KU998256	Jeanie	2015	CW (CW1)	<i>Gordonia</i>	<i>neofelifaecis</i>	NRRL 59395	<i>Siphoviridae</i>	Temperate	17118	68.6
KU998255	McGonagall	2015	CW (CW1)	<i>Gordonia</i>	<i>neofelifaecis</i>	NRRL 59395	<i>nd</i>	Temperate	17119	68.6
MT498067	Bullzi2019	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17143	68.8
MW055909	Namsahir	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17176	68.9
MH399783	Noelani	2017	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17349	68.2
MH744416	Dongwon	2017	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17362	68.5
MH651181	Minima	2017	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17362	68.5
MH045561	PaoPu	2016	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17362	68.5
MT639642	YertiPhresh	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17362	68.5
MW055907	JoBros	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17363	68.5
MN284892	Bradley2	2017	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17367	68.5
MT498051	Jahseh	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17368	68.5
MT639647	JRok	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17382	68.7
MN585990	Azizam	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17385	68.7
MK878899	Hulk	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17395	68.7
MN369745	Kaijohn	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17426	68.8
MK524502	TimoTea	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17427	68.7
MT316458	McShie	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17434	68.9
MT521990	Bri160	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17442	68.6
MT776807	Majesty	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17444	68.7
MH045557	BurtonThePup	2016	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17445	68.8
MN010759	Ciel	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17445	68.8
MT657332	NeumannU	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17445	68.7
MN428064	Owens	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>nd</i>	Lytic	17445	68.8
MN586008	Scrunchy	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17445	68.8
MN010755	Slentz	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	NRRL B-24224	<i>Siphoviridae</i>	Lytic	17445	68.7

Genomic conservation in small phages

MN234191	Naby	2018	EE	<i>Microbacterium</i>	<i>paraoxydans</i>	<i>NWU1</i>	<i>nd</i>	Lytic	17448	68.2
MN428062	TeddyBoy	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>nd</i>	Lytic	17449	68.8
MT498040	Livingwater	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17450	68.6
MH371111	Quaker	2017	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17450	68.6
MH045556	BonaeVitae	2015	EE	<i>Microbacterium</i>	<i>paraoxydans</i>	<i>NWU1</i>	<i>Siphoviridae</i>	Lytic	17451	68.2
MT316462	Danno	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17452	68.7
MH576963	Scamander	2017	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>nd</i>	Lytic	17452	68.7
MK894434	LaviMo	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17453	68.7
MK620897	Rhysand	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>nd</i>	Lytic	17453	68.7
MN329679	Vanisius	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17453	68.7
MT316459	Otwor	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>nd</i>	Lytic	17454	68.7
MH371118	KayPaulus	2017	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17455	68.5
MT498038	Luxx	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17463	68.8
MT952850	Sippinontea	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>nd</i>	Lytic	17473	68.6
MT498046	Hernandez44	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17477	68.8
MH825700	Efeko	2017	EE	<i>Microbacterium</i>	<i>paraoxydans</i>	<i>NRRL B-14843</i>	<i>Siphoviridae</i>	Lytic	17491	68.6
MK801730	Belthelas	2018	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17502	68.6
MT498065	PoRanda	2019	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17510	69.1
MH371124	VitulaEligans	2017	EE	<i>Microbacterium</i>	<i>foliorum</i>	<i>NRRL B-24224</i>	<i>Siphoviridae</i>	Lytic	17534	68.8
KR053198	GMA5	2015	CW (CW2)	<i>Gordonia</i>	<i>malaquae</i>	<i>BEN700</i>	<i>nd</i>	Temperate	17562	66.4
KR053197	GRU3	2015	CW (CW2)	<i>Gordonia</i>	<i>rubripertincta</i>	<i>Grub38</i>	<i>nd</i>	Temperate	17727	66.5
MK620900	Forthebois	2018	BO	<i>Streptomyces</i>	<i>scabiei</i>	<i>RL-34</i>	<i>Tectiviridae</i>	<i>nd</i>	18251	53.6
MK305890	WheeHeim	2017	BO	<i>Streptomyces</i>	<i>scabiei</i>	<i>RL-34</i>	<i>Tectiviridae</i>	<i>nd</i>	18266	54.6
MK016500	Mendel	2017	FD	<i>Arthrobacter</i>	<i>globiformis</i>	<i>B-2979</i>	<i>nd</i>	Lytic	19428	59.8
MK016490	Anjali	2017	FD	<i>Arthrobacter</i>	<i>globiformis</i>	<i>B-2979</i>	<i>nd</i>	Lytic	19679	59.4

nd = not determined. The 34 representative phages used in the study's analyses are bolded. Phages are ordered by increasing genome length.

Genomic conservation in small phages



Supplementary Figure S1. Pham maps reveal conservation of terminase and other phams among most small genome phages. 11 Phamerator genome maps, of at least one representative phage from each cluster within the group of 34 small genome phages, are displayed as of November 2020. Shared phams are colored; unshared phams are grayed out; orphans are uncolored. Shading indicates nucleotide similarity. Terminase conservation was observed in atypically small *Gordonia* phages: the group of phages that include RRH1 (Singleton), Decurro (AN), Schiebs (CW3), Atraxa (AX) and SallySpecial (DM) all exhibit intercluster conservation of the same terminase pham. All phages on this figure also share similar genome architectures; intercluster conservation of other genes besides terminase are also evident, such as the portal protein and tape measure protein.

References

1. Keen EC. A century of phage research: Bacteriophages and the shaping of modern biology. *BioEssays News Rev Mol Cell Dev Biol.* 2015;37: 6–9. doi:10.1002/bies.201400152
2. Grose JH, Casjens SR. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. *Virology.* 2014;0: 421–443. doi:10.1016/j.virol.2014.08.024
3. Hatfull GF. Dark Matter of the Biosphere: the Amazing World of Bacteriophage Diversity. *J Virol.* 2015;89: 8107–8110. doi:10.1128/JVI.01340-15
4. Russell DA, Hatfull GF. PhagesDB: the actinobacteriophage database. Wren J, editor. *Bioinformatics.* 2017;33: 784–786. doi:10.1093/bioinformatics/btw711
5. Petrovski S, Dyson ZA, Seviour RJ, Tillett D. Small but Sufficient: the *Rhodococcus* Phage RRH1 Has the Smallest Known Siphoviridae Genome at 14.2 Kilobases. *J Virol.* 2012;86: 358–363. doi:10.1128/JVI.05460-11
6. Hatfull GF. Bacteriophage genomics. *Curr Opin Microbiol.* 2008;11: 447–453. doi:10.1016/j.mib.2008.09.004
7. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, et al. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *eLife.* 2015;4: e06416. doi:10.7554/eLife.06416
8. Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, Montgomery MT, et al. Bacteriophages of *Gordonia* spp. Display a Spectrum of Diversity and Genetic Relationships. Losick R, editor. *mBio.* 2017;8: e01069-17, /mbio/8/4/e01069-17.atom. doi:10.1128/mBio.01069-17
9. Jacobs-Sera D, Abad LA, Alvey RM, Anders KR, Aull HG, Bhalla SS, et al. Genomic diversity of bacteriophages infecting *Microbacterium* spp. *PLoS ONE.* 2020;15. doi:10.1371/journal.pone.0234636
10. Hatfull GF. Complete Genome Sequences of 61 Mycobacteriophages. *Genome Announc.* 2016;4. doi:10.1128/genomeA.00389-16
11. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, et al. A Broadly Implementable Research Course in Phage Discovery and Genomics for First-Year Undergraduate Students. *mBio.* 2014;5. doi:10.1128/mBio.01051-13
12. Russell DA. Sequencing, Assembling, and Finishing Complete Bacteriophage Genomes. *Methods Mol Biol Clifton NJ.* 2018;1681: 109–125. doi:10.1007/978-1-4939-7343-9_9
13. Pope WH, Jacobs-Sera D. Annotation of Bacteriophage Genome Sequences Using DNA Master: An Overview. In: Clokie MRJ, Kropinski AM, Lavigne R, editors. *Bacteriophages: Methods and Protocols, Volume 3.* New York, NY: Springer; 2018. pp. 217–229. doi:10.1007/978-1-4939-7343-9_16
14. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 1999;27: 4636–4641. doi:10.1093/nar/27.23.4636
15. Besemer J, Borodovsky M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.* 2005;33: W451–W454. doi:10.1093/nar/gki487

Genomic conservation in small phages

16. SEA-PHAGES/starterator. SEA-PHAGES Program; 2020. Available: <https://github.com/SEA-PHAGES/starterator>
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2
18. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 2011;39: D225–D229. doi:10.1093/nar/gkq1189
19. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 2015;43: D222–D226. doi:10.1093/nar/gku1221
20. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005;33: W244–W248. doi:10.1093/nar/gki408
21. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics.* 2007;23: 1026–1028. doi:10.1093/bioinformatics/btm039
22. Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. *Mol Biol Evol.* 2006;23: 254–267. doi:10.1093/molbev/msj030
23. Caruso SM, deCarvalho TN, Huynh A, Morcos G, Kuo N, Parsa S, et al. A Novel Genus of Actinobacterial Tectiviridae. *Viruses.* 2019;11: 1134. doi:10.3390/v11121134
24. Demo S, Kapinos A, Bernardino A, Guardino K, Hobbs B, Hoh K, et al. BlueFeather, the singleton that wasn't: Shared gene content analysis supports expansion of *Arthrobacter* phage cluster FE. *Microbiology*; 2020 Aug. doi:10.1101/2020.08.15.252577
25. Ha AD, Denver DR. Comparative Genomic Analysis of 130 Bacteriophages Infecting Bacteria in the Genus *Pseudomonas*. *Front Microbiol.* 2018;9. doi:10.3389/fmicb.2018.01456
26. Hatfull GF, Hendrix RW. Bacteriophages and their Genomes. *Curr Opin Virol.* 2011;1: 298–303. doi:10.1016/j.coviro.2011.06.009
27. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko C-C, et al. Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J Mol Biol.* 2010;397: 119–143. doi:10.1016/j.jmb.2010.01.011
28. Klyczek KK, Bonilla JA, Jacobs-Sera D, Adair TL, Afram P, Allen KG, et al. Tales of diversity: Genomic and morphological characteristics of forty-six Arthrobacter phages. *PLOS ONE.* 2017;12: e0180517. doi:10.1371/journal.pone.0180517
29. Yuan Y, Gao M. Jumbo Bacteriophages: An Overview. *Front Microbiol.* 2017;8. doi:10.3389/fmicb.2017.00403
30. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, et al. Origins of Highly Mosaic Mycobacteriophage Genomes. *Cell.* 2003;113: 171–182. doi:10.1016/S0092-8674(03)00233-2
31. Casjens SR, Thuman-Commike PA. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. *Virology.* 2011;411: 393–415. doi:10.1016/j.virol.2010.12.046
32. Hatfull GF. Mycobacteriophages. *Microbiol Spectr.* 2018;6: 10.1128/microbiolspec.GPP3-0026-2018.

Genomic conservation in small phages

doi:10.1128/microbiolspec.GPP3-0026-2018