Novel Cluster AZ *Arthrobacter* phages Powerpuff, Lego, and YesChef exhibit close functional relationships with *Microbacterium* phages

Andrew Kapinos, Pauline Aghamalian, Erika Capehart, Anya Alag, Heather Angel, Eddie Briseno, Emily Farag, Hilory Foster, Abbas Hakim, Daisy Hernandez-Casas, Calvin Huang, Derek Lam, Maya Mendez, Ashley Min, Nikki Nguyen, Emily Ortiz, Lizbeth Shelly Saldivar, Jack Arthur Shannon, Rachel Smith, Mihika V. Sridhar, An Ta, Malavika C. Theophilus, Ryan Ngo, Canela Torres, Krisanavane Reddi, Amanda C. Freise, Jordan Moberg Parker*

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

Keywords: bacteriophage, *Arthrobacter*, *Microbacterium*, clustering, lysogeny, virology, microbiology, comparative genomics

*Corresponding Author:
Jordan Moberg Parker, jmobergparker@ucla.edu

Brief Running Title
Cluster AZ Comparative Analysis

Abstract

Bacteriophages exhibit a vast spectrum of relatedness and there is increasing evidence of close genomic relationships independent of host genus. The variability in phage similarity at the nucleotide, amino acid, and gene content levels confounds attempts at quantifying phage relatedness, especially as more novel phages are isolated. This study describes three highly similar novel *Arthrobacter globiformis* phages–Powerpuff, Lego, and YesChef–which were assigned to Cluster AZ using a nucleotide-based clustering parameter. Phages in Cluster AZ and *Microbacterium* Cluster EH, as well as the former *Microbacterium* singleton Zeta1847, exhibited low nucleotide similarity but gene content similarity in excess of the recently adopted *Microbacterium* clustering parameter, which resulted in the reassignment of Zeta1847 to Cluster EH. Additionally, Clusters AZ and EH phages encode a shared integrase indicative of a lysogenic life cycle; in the first experimental verification of a Cluster AZ phage's life cycle, we show that phage Powerpuff is a true temperate phage and forms stable lysogens that exhibit immunity to superinfection by related phages, despite lacking identifiable repressors typically required for lysogenic maintenance and superinfection immunity. The ability of phage Powerpuff to undergo and maintain lysogeny suggests that other closely related phages encoding an integrase but lacking an identified repressor may be temperate as well. Our findings further highlight the importance of using multiple metrics to capture phage relatedness, provide additional evidence of significant shared phage genomic content spanning multiple actinobacterial

host genera, and demonstrate the continued need for verification and characterization of life cycles in newly isolated phages.

Introduction

Bacteriophages comprise the most abundant group of biological entities on the planet, with an estimated 10^{31} phage particles in existence (1). Despite the growing body of evidence suggesting the immense role phages play in ecological regulation through interactions with their bacterial hosts (2–4), the phage population as a whole remains relatively understudied with only 3,758 actinobacteriophage genomes published to PhagesDB as of August 2021 (5).

Actinobacteriophages display immense genomic and biological diversity (6). Past studies have observed that phages infecting the same bacterial host and exhibiting the same viral life cycle tend to share the highest amount of nucleotide similarity, with a more conserved evolutionary history (7); however, substantial levels of genomic diversity have been identified even amongst phages known to infect a common host (6). Studies of phage relatedness are further complicated by the mosaic nature of phage genomes, due to widespread exchanges of modules of genetic material (8). Given that host barriers to genetic exchange are more readily violable than previously thought, this can result in phages of unique bacterial hosts sharing considerable gene content (9). A recent study of a large collection of *Microbacterium* phages described significant shared gene content amongst a group of phages infecting *Microbacterium*, *Streptomyces*, *Rhodococcus*, *Gordonia*, and *Arthrobacter* spp. (10). It was also found that sequenced *Microbacterium* phages exhibited shared gene content or genome architecture with *Arthrobacter* phages. This evidence suggested that phages infecting *Microbacterium* and *Arthrobacter* spp. may exhibit proximal phylogenetic relationships.

Few studies have specifically explored the phages that infect *Arthrobacter*, a genus of bacteria that is primarily soil-dwelling and engaged in the biochemical processing of natural compounds (11–13). Klyczek *et al.* described a collection of *Arthrobacter* phages, all isolated on *Arth*robacter sp. ATCC 21022, which shared no nucleotide sequence similarity with phages infecting other actinobacterial hosts (14). These *Arthrobacter* phages were considered to be primarily lytic, similar to sequenced *Microbacterium* phages (10), and unlike the *Mycobacterium* and *Gordonia* phages which are more likely to be temperate (9,15). This predominance of lytic *Arthrobacter* phages continues to be evidenced today–of the 331 sequenced *Arthrobacter* phages on PhagesDB as of August 2021, only 61 are predicted to be temperate, comprising Clusters AS, AY, AZ, FA, FF, and FG (5). Importantly, many predictions of *Arthrobacter* phage life cycles have depended on bioinformatic evidence, such as the presence of a known integrase, and have yet to be verified experimentally. Additional analyses of potential genomic relationships of *Arthrobacter* phages, including investigations of amino acid identity and Gene Content Similarity (GCS) (9), have so far been limited in scope. The isolation of novel *Arthrobacter* phages allows for more thorough genomic comparisons to phages infecting both *Arthrobacter* and other actinobacterial hosts and provides the opportunity for experimental verification of phage life cycles.

This study describes the relationships of novel *Arthrobacter globiformis* phages Powerpuff, Lego, and YesChef to phages infecting *Arthrobacter* and non-*Arthrobacter* hosts. These phages were determined to be members of the actinobacteriophage Cluster AZ using a nucleotide-based clustering parameter (14,16). We discovered that,

while Cluster AZ phages shared minimal similarity with *Microbacterium* and *Streptomyces* phages at the nucleotide level, sequence similarities at the amino acid level were more well-conserved. Additionally, gene-level analyses showed that phages in Cluster AZ shared genome architecture with *Microbacterium* phages in Cluster EH, as well as GCS in excess of the recently adopted 35% clustering threshold for *Microbacterium* phages (10). The *Microbacterium* phage Zeta1847, which was sequenced in 2016 and previously designated as a singleton phage, was also found to share over 35% GCS with all Cluster EH phages and was thus reassigned to Cluster EH.

In the first experimental verification of a Cluster AZ phage's life cycle, we found that the novel phage Powerpuff formed stable lysogens which exhibited immunity to superinfection by Cluster AZ phages Powerpuff, Lego, and YesChef. Thus, phage Powerpuff is likely a true temperate phage encoding a repressor that has yet to be identified. This experimental verification suggests that Clusters AZ and EH phages also encoding integrases but lacking identifiable repressors are perhaps temperate as well. In sum, we present the first comparative genomic study of phages belonging to actinobacteriophage Cluster AZ and describe, for the first time, significant conserved genomic content and likely shared biological functions between *Arthrobacter* phages of Cluster AZ and phages of other actinobacterial hosts.

Methods

Phage isolation, purification, and amplification

Three soil samples were collected from within Los Angeles County, CA, USA: 34.443624 N, 118.609545 W (Powerpuff), 34.016253 N, 118.501056 W (Lego), and 34.052707 N, 118.44657 W (YesChef). Direct isolation of phage YesChef was performed at 30ºC using PYCa broth (Yeast Extract 1 g/L, Peptone 15 g/L, 4.5mM $CaCl_2$, Dextrose 0.1%), while enriched isolations of phages Powerpuff and Lego were performed at 25ºC and 30ºC, respectively, using 10X PYCa broth (Yeast Extract 10 g/L, Peptone 150 g/L, 45mM $CaCl_2$, Dextrose 10%) and *Arthrobacter globiformis* B-2979. Filter-sterilized samples were spot tested using *A. globiformis* B-2979 and PYCa media using the double agar overlay method. Samples containing putative phage were purified and amplified as described previously (17).

Transmission electron microscopy (TEM)

Each high titer lysate was aliquoted onto a carbon-coated grid and stained using 1% (w/v) uranyl acetate. Each carbon grid was imaged using a FEI T12 TEM Instrument (Thermo Fisher Scientific, MA, USA) at magnifications between 30,000X and 42,000X. Phage capsid and tail measurements were determined using ImageJ (18).

DNA extraction and sequencing

DNA was extracted from high titer lysates using the Wizard® Clean-Up Kit (cat. # A7280, Promega, WI, USA). Sequencing libraries were constructed with the NEBNext® UltraTM II DNA Library Prep kit (New England Biolabs, MA, USA), and shotgun sequenced by Illumina-MiSeq at the Pittsburgh Bacteriophage Institute. Genome

assembly and finishing were conducted as previously described (19).

Genome annotation

The Phage Evidence Collection and Annotation Network (PECAAN) was used to document evidence during manual annotation of phage genomes (https://discover.kbrinsgd.org/). Genes were preliminarily auto-annotated using DNA Master (http://cobamide2.bio.pitt.edu). GeneMark (20) and Glimmer (21) were used to assess coding potential. Phamerator was used to assign genes to phamilies (phams) on the basis of amino acid similarity and to examine synteny with related phages (22). Conserved start sites were identified using Starterator (23). PhagesDB BLASTp (5), NCBI BLASTp (24), the NCBI Conserved Domain Database (25), and HHpred (26) were used for gene function calls. Membrane protein topology programs TmHmm (27) and TOPCONS (28) were used to identify putative transmembrane domains within draft genes.

Comparative genomic analyses

Upon the completion of manual annotation, the final version of each phage genome was downloaded from Phamerator and used to create a linear genome map using Inkscape 1.0 (https://inkscape.org/). NCBI Nucleotide BLAST (BLASTn) was optimized for highly similar sequences (megablast) and used to identify similar phage genomes. Gepard 1.40 was used to generate dotplots using word sizes of 15 and 5 for nucleotide and amino acid inputs, respectively (29). OrthoANIu and coverage values were calculated using the command-line OrthoANIu tool provided by EZBioCloud (30)

and visualized as a heat map using Prism 8.0.0 (Graphpad Software, San Diego, California, USA).

Pham data for phages of interest from the Actino_Draft database (version 382) were input into SplitsTree 4.16.1 to produce a network phylogeny using default parameters (31). Gene Content Similarity between phages of interest was calculated using the PhagesDB Explore Gene Content tool (5) and visualized as a heatmap using Graphpad Prism 8.0.0. Specific information regarding pham presence and function in each phage of interest was collected using PhagesDB and Phamerator.

Host range assay

Undiluted high titer lysates of phages Powerpuff, Lego, and YesChef were spotted onto *Microbacterium foliorum* NRRL B-24224 and a control lawn of *A. globiformis* B-2979 using PYCa media. All plates were incubated at 25°C for 48 hours then examined for lytic activity.

Preparation of stable lysogens and immunity assays

Powerpuff high titer lysate was serially diluted and spotted onto *A. globiformis* B-2979, then incubated at 30ºC for 96 hours. All subsequent immunity assay plates were incubated at 30ºC for 48 hours. Bacterial mesas from spot dilutions 10^0 through 10^{-3} were streak purified three times on PYCa media to remove exogenous phage particles. Experimental plates were prepared by streaking putative lysogens onto a prepared lawn of *A. globiformis* B-2979, while control plates were prepared in the absence of host cells.

To verify the presence of stable lysogens, liquid cultures of streak purified putative lysogens were incubated at 30ºC for 48 hours and then pelleted. Ten-fold serial dilutions of supernatants were spot tested on *A. globiformis* B-2979 to confirm phage release and calculate titer. Immunity assays of *Arthrobacter* Cluster AZ phages Powerpuff, Lego, and YesChef, Cluster FE phage BlueFeather, Cluster AU phage Giantsbane, and Cluster AO phage Abba were performed using ten-fold serial dilutions of phage lysate on wild-type (WT) *A. globiformis* B-2979 and *A. globiformis* B-2979 lysogens of Powerpuff.

Results

Phages Powerpuff, YesChef, and Lego are highly similar *Siphoviridae* members of Cluster AZ

Phages Powerpuff, Lego, and YesChef all exhibited 1-3 mm turbid bullseye plaques after 24 hours of incubation at their respective isolation temperatures (Fig 1). Transmission electron microscopy of the three phages revealed similar particle dimensions, with an average head diameter of 56.9 nm and an average tail length of 124.5 nm (Table 1). All phages exhibited long, flexible tails, indicative of *Siphoviridae* (32).

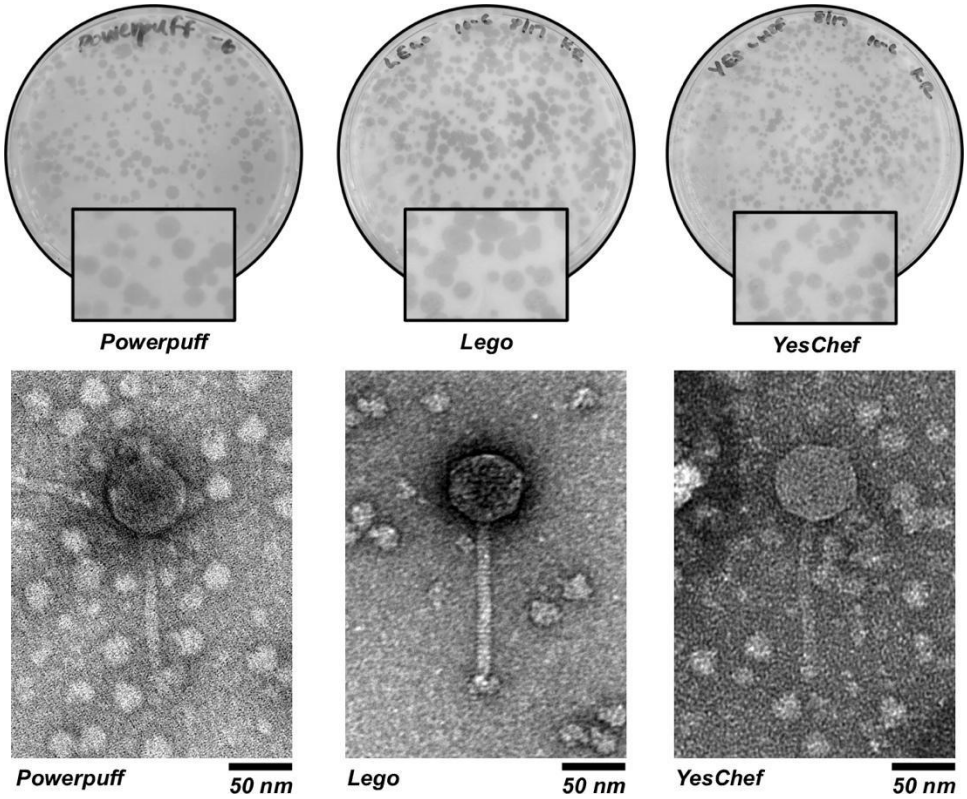

Fig 1: Phages Powerpuff, Lego, and YesChef exhibit turbid bullseye plaque morphologies and are *Siphoviridae.*
Purified phage lysates were plated using the double agar overlay method for plaque visualization. Each phage exhibited 1-3 mm turbid bullseye plaques, which were suggestive of a lysogenic life cycle. The presence of long, flexible tails suggested their classification as *Siphoviridae*.

Table 1: Phages Powerpuff, Lego, and YesChef share similar genomic and physical characteristics.

Phage	Accession	Genome length (bp)	%GC Content	No. of genes	Head diam. (nm)	Tail length (nm)
Powerpuff	MN703413	44651	67.6%	71	56.7 ± 13.9	126.7 ± 17.0
Lego	MT024869	43446	67.5%	69	57.6 ± 2.1	120.4 ± 10.7
YesChef	MT024871	43510	67.7%	69	56.5 ± 5.2	126.3 ± 1.9

Genome sequencing and assembly determined that all three phages exhibited 11 base 3' sticky overhangs (CGAAGGGGCAT), with similar genome length, percent GC content, and number of genes (Table 1). Powerpuff, Lego, and YesChef were assigned to Cluster AZ using a nucleotide-based parameter (14,16). Powerpuff, Lego,

and YesChef were also found to be nearly genomically identical according to a preliminary BLASTn search, which revealed pairwise BLASTn coverages of at least 99%, with E-values of 0 and identities of at least 98.63% (Table S1). These nucleotide similarities translated into high similarity in gene content and genome architecture, with few differences between these phages at the gene level (Fig S1). Only Cluster AZ phages Tbone and Kaylissa shared between 96.26-96.86% BLASTn identity and 92-98% coverage with Powerpuff, Lego, and YesChef, with all pairwise comparisons having E-values of 0. The remaining 7 phages analyzed from Cluster AZ shared between 80.14-87.79% BLASTn identity and 8-81% coverage with the novel phages. While Powerpuff, Lego, and YesChef were isolated from within Los Angeles County, phages Tbone and Kaylissa were isolated from Louisiana and New York state, respectively, based on records from PhagesDB.org (5). This provides another interesting case of phages which are extremely similar genomically, despite being isolated from locations which are geographically distant (33).

Cluster AZ *Arthrobacter* phages are diverse and share nucleotide similarity with Cluster EH *Microbacterium* phages

Nucleotide comparisons of phages Powerpuff, Lego, and YesChef to the most genomically similar actinobacteriophages both within and outside of Cluster AZ were performed. It was expected that the most similar set of phages to Powerpuff, Lego, and YesChef at the nucleotide level would also be members of *Arthrobacter*-infecting Cluster AZ, which was confirmed using PhagesDB BLASTn (Table S2). The most similar phages to Powerpuff, Lego, and YesChef outside of Cluster AZ included phages in

Microbacterium Clusters EH and EB, *Arthrobacter* Cluster AK, and *Streptomyces* Cluster BJ.

Nucleotide dotplot comparisons using a word length of 15 revealed that Cluster AZ phages Liebe and Maureen exhibited strong alignments to one another but weak alignments when compared to the remainder of their cluster (Fig 2). As expected from the BLASTn results, Cluster AZ shared some degree of nucleotide similarity with phages belonging to *Microbacterium* Cluster EH, and displayed significantly less similarity with Cluster EB phages. There were no nucleotide alignments observed between Cluster AZ and phages in Clusters AK or BJ, despite being the next most similar set of phages to Powerpuff per BLASTn.

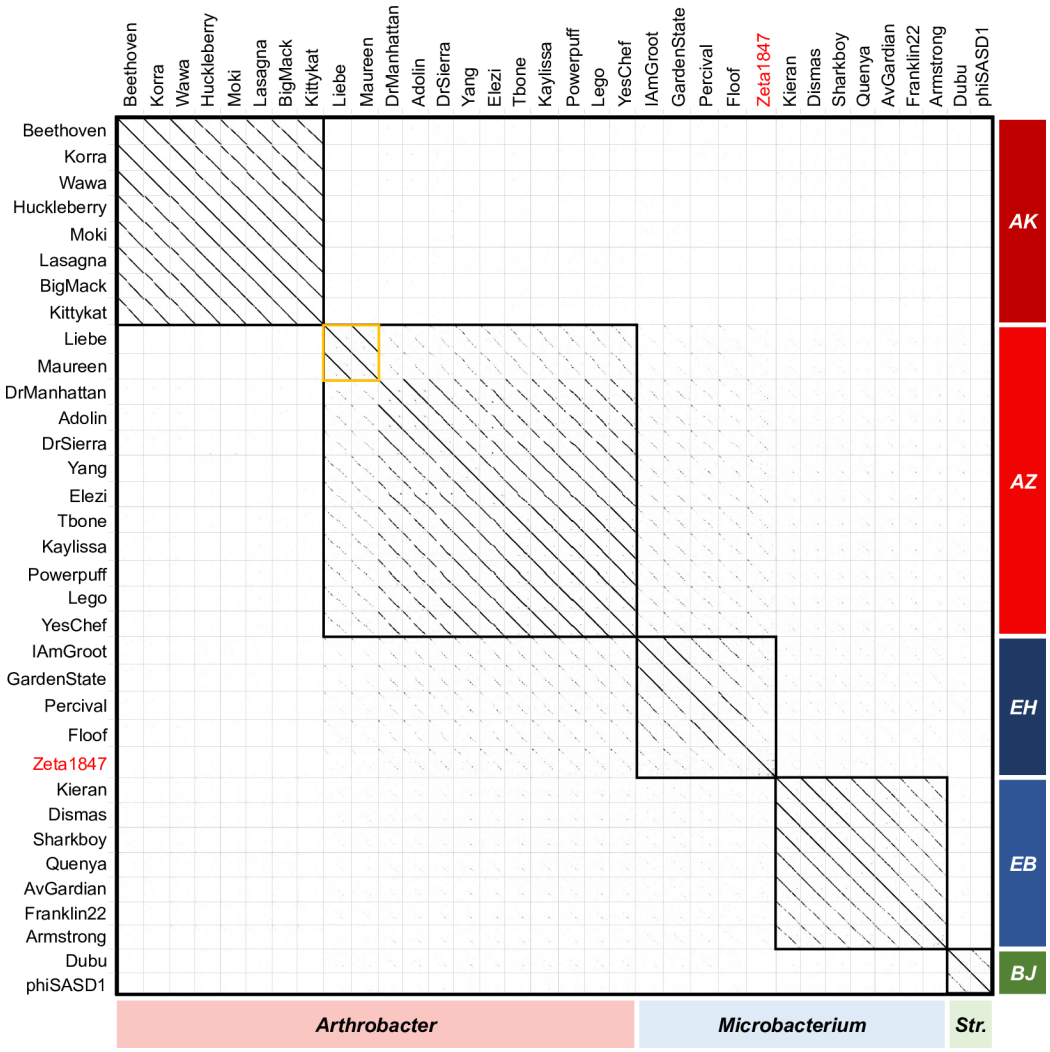

Fig 2: Nucleotide dotplots reveal two groups of highly similar phages within Cluster AZ, with some similarity to *Microbacterium* phages sampled.
Whole genome nucleotide sequences were analyzed using Gepard dotplot software and a word size of 15. Cells boxed in black represent phage clusters and the former singleton Zeta1847 is indicated in red. "*Str.*" indicates phages infecting *Streptomyces.* There were few alignments between phages of different clusters. Within Cluster AZ, phages Liebe and Maureen exhibited strong alignments to each other but weak alignments to the remainder of their cluster (indicated in yellow).

Similarities observed at the nucleotide level were computationally confirmed using OrthoANIu (Fig 3), which quantifies the similarity in orthologous nucleotide sequences between genomes (34). Within phage clusters, OrthoANIu tended to be high, with intracluster OrthoANIu values at or above 70%, well in excess of the 50%

identity threshold required for clustering under nucleotide-based parameters (16). It is notable that high OrthoANIu values existed for many pairwise comparisons between phages of unique clusters, such as those between Clusters AK, AZ, and EB phages. The vast majority of such comparisons exhibited coverage values below 5%.

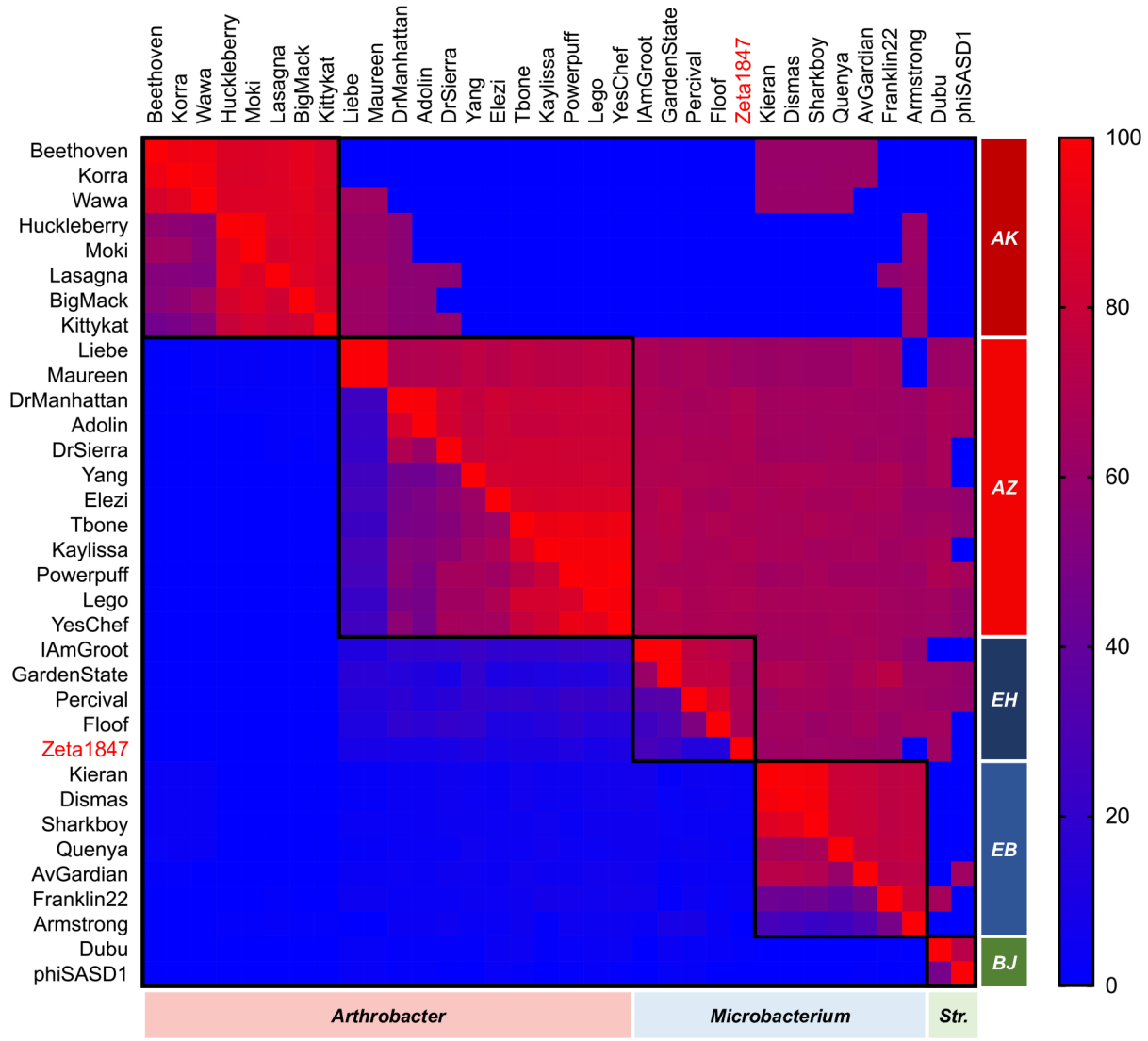

Fig 3: OrthoANIu indicates widespread shared genomic features between Cluster AZ and *Microbacterium* phages.
Average nucleotide identities between orthologous regions of each genome (OrthoANIu) and respective coverages were calculated using a command-line OrthoANIu tool, then visualized as a heat map using Prism 8.0.0. Upper right values represent pairwise OrthoANIu and bottom left values represent average coverage. Cells boxed in black represent phage clusters and the former singleton Zeta1847 is indicated in red. "*Str.*" indicates phages infecting *Streptomyces.* OrthoANIu values supported

findings of the nucleotide dotplot and indicated widespread presence of small but well-conserved genomic features.

As expected from weak alignments on the nucleotide dotplot, comparisons to Cluster AZ phages Liebe or Maureen accounted for the lowest OrthoANIu and coverage values both within Cluster AZ and between Clusters AZ and EH (Fig 3). The most similar Clusters AZ and EH phages exhibited stronger OrthoANIu and coverage values than the most dissimilar phages within Cluster AZ. This was confirmed using PhagesDB BLASTn, in which the score and E-value of the weakest comparisons between two Cluster AZ phages (Liebe/Maureen v. Adolin; 313 bits score, E-value $3e^{-82}$) were weaker than the strongest comparison between a Cluster AZ and Cluster EH phage (Yang v. IAmGroot/GardenState; 389 bits score, E-value $1e^{-105}$). This suggests that, at the nucleotide level, Cluster AZ phages Liebe and Maureen are about as similar to the remainder of Cluster AZ (including Powerpuff, Lego, and YesChef) as these phages are to *Microbacterium* phages in the next most closely related cluster, highlighting the diversity of the Cluster AZ phages.

Amino acid sequences are similar between *Arthrobacter, Microbacterium, and Streptomyces* phages

Codon degeneracy allows for phage sequences to be shared at the amino acid level but not at the nucleotide level (9,35), which may limit the apparent similarity of phage genomes when comparing nucleotide sequences alone. Amino acid dotplot comparisons revealed similarity between *Arthrobacter* phage Clusters AZ and AK, despite minimal nucleotide identity (Fig 4). Within Cluster AZ, Liebe and Maureen exhibited strong alignments to the remainder of their cluster, in contrast with the

nucleotide similarity results. Similarities between Cluster AZ and *Microbacterium* phage Clusters EH and EB, as well as *Streptomyces* phages belonging to Cluster BJ, were also strengthened when compared to nucleotide dotplot comparisons. This increase in alignment strength when comparing phages at the amino acid level is indicative of synonymous substitutions in the nucleotide code, perhaps suggesting a distant evolutionary relationship for alignments which are strengthened or apparent only at the amino acid level (36).

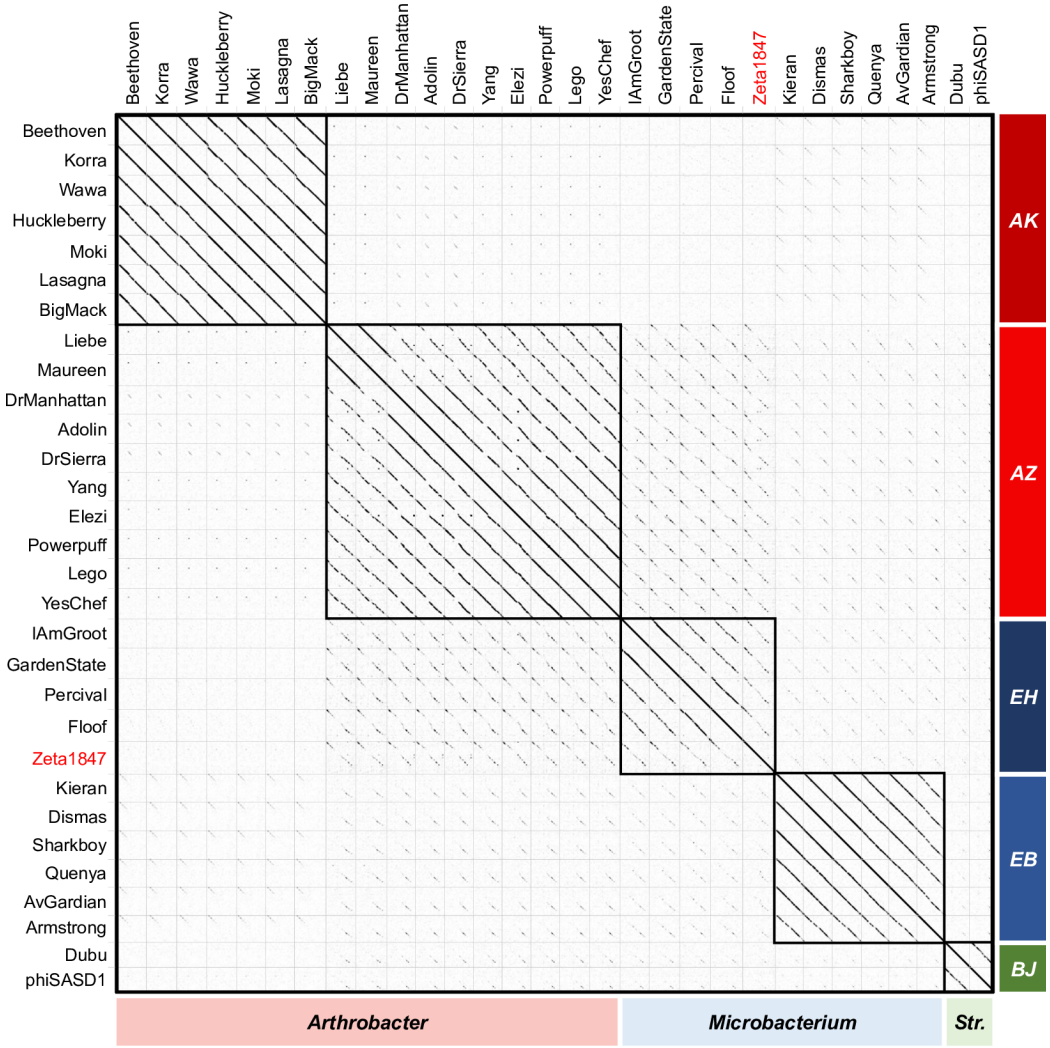

Fig 4: Amino acid dotplots reveal a history of synonymous substitutions. Whole genome amino acid sequences were analyzed using Gepard dotplot software and a word size of 5. Cells boxed in black represent phage clusters and the former

singleton Zeta1847 is indicated in red. "*Str.*" indicates phages infecting *Streptomyces.* Increased alignment strength at the amino acid level indicated a history of synonymous substitutions and suggested distant relationships.

Clusters AZ and EH share genome architecture and gene content in excess of *Microbacterium* clustering parameters

The observation of phages which share substantial portions of their gene content despite lacking significant nucleotide similarity and/or span-length coverage (9,35) has prompted the adjustment of clustering parameters for new phage clusters from a nucleotide-based parameter (16) to an updated threshold of at least 35% shared gene content (9,10). Thus, while nucleotide and amino acid comparisons serve as important preliminary metrics for determining similarity between phages, analyses of shared gene content may serve as more functionally relevant metrics for phage comparison.

GCS values were calculated for each genome pair included in the nucleotide and amino acid comparisons performed above. Interestingly, the putative singleton phage Zeta1847 displayed GCS values with Cluster EH phages in excess of the *Microbacterium* phage clustering parameter, sharing between 37.5% and 40% GCS with Cluster EH phages. This finding resulted in the assignment of Zeta1847 into Cluster EH by phagesdb.org.

With a mean GCS value of 32.69%, there were multiple pairwise comparisons between phages in *Arthrobacter* Cluster AZ and *Microbacterium* Cluster EH which met or exceeded the 35% GCS clustering parameter, indicated by a white outline on the GCS heatmap (Fig 5). Cluster AZ phages DrManhattan and Adolin shared between 35.2% and 37.3% GCS with Cluster EH phages IAmGroot, GardenState, and Percival.

Cluster AZ phages DrSierra and Yang also shared 35.8% and 35.3% GCS with phage Percival, respectively, while phages Liebe and Maureen shared 36.4% GCS with Cluster EH phage Floof. These comparisons appear to be as substantial as other recently identified relationships between *Arthrobacter* and *Microbacterium* phages, which also shared up to 40% GCS (10).

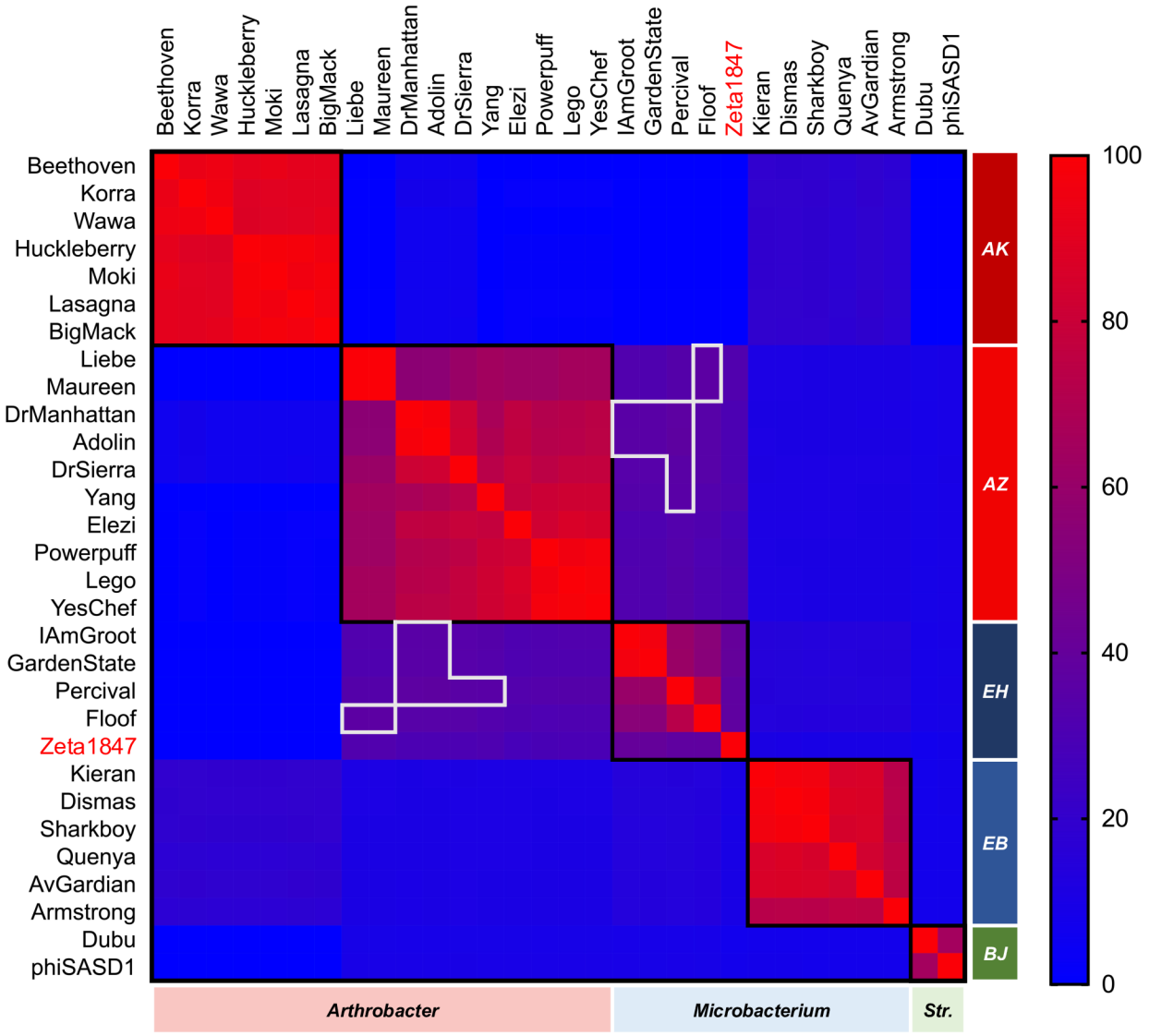

Fig 5: Cluster AZ and EH phages share gene content in excess of *Microbacterium* clustering parameters.
GCS values were recorded using the PhagesDB Explore Gene Content tool and visualized as a heat map using Prism 8.0.0. Cells boxed in white represent pairwise GCS values in excess of gene content clustering parameters (≥35%) between phages

belonging to different clusters. Cells boxed in black represent phage clusters and the former singleton Zeta1847 is indicated in red. "*Str.*" indicates phages infecting *Streptomyces.* Some phages in Clusters AZ and EH shared over 35% GCS, in excess of the *Microbacterium* clustering parameter. The former singleton Zeta1847 shared over 35% GCS with Cluster EH phages, which resulted in the clustering of this phage with Cluster EH.

Phages in Clusters AZ and EH also shared similar genome architecture (Fig 6). The right arm of the genome displayed less synteny between these phages; however, genes found in the same gene phamilies (phams) tended to be arranged in the same order. While Cluster AZ representative phage Powerpuff encoded an endolysin in the right arm of the genome, the Cluster EH phages encoded endolysins in the left arm. Within Cluster AZ, only phages Elezi, Liebe, and Maureen also encoded endolysins in the left arm as the Cluster EH phages do. The relative proximity of the relationships between Clusters AZ and EH was further evidenced by a SplitsTree network phylogeny of the shared gene content of these phages, in which Clusters AZ and EH formed a large branch separate from the remainder of the phages (Fig 7). The *Arthrobacter* Clusters AK and AZ were segregated from each other on the tree. This supports the notion that there exists great diversity even amongst phages infecting the same host (6), and that phages are perhaps just as likely to share gene-level similarities with phages infecting other hosts when compared to phages infecting the same host but belonging to different clusters.

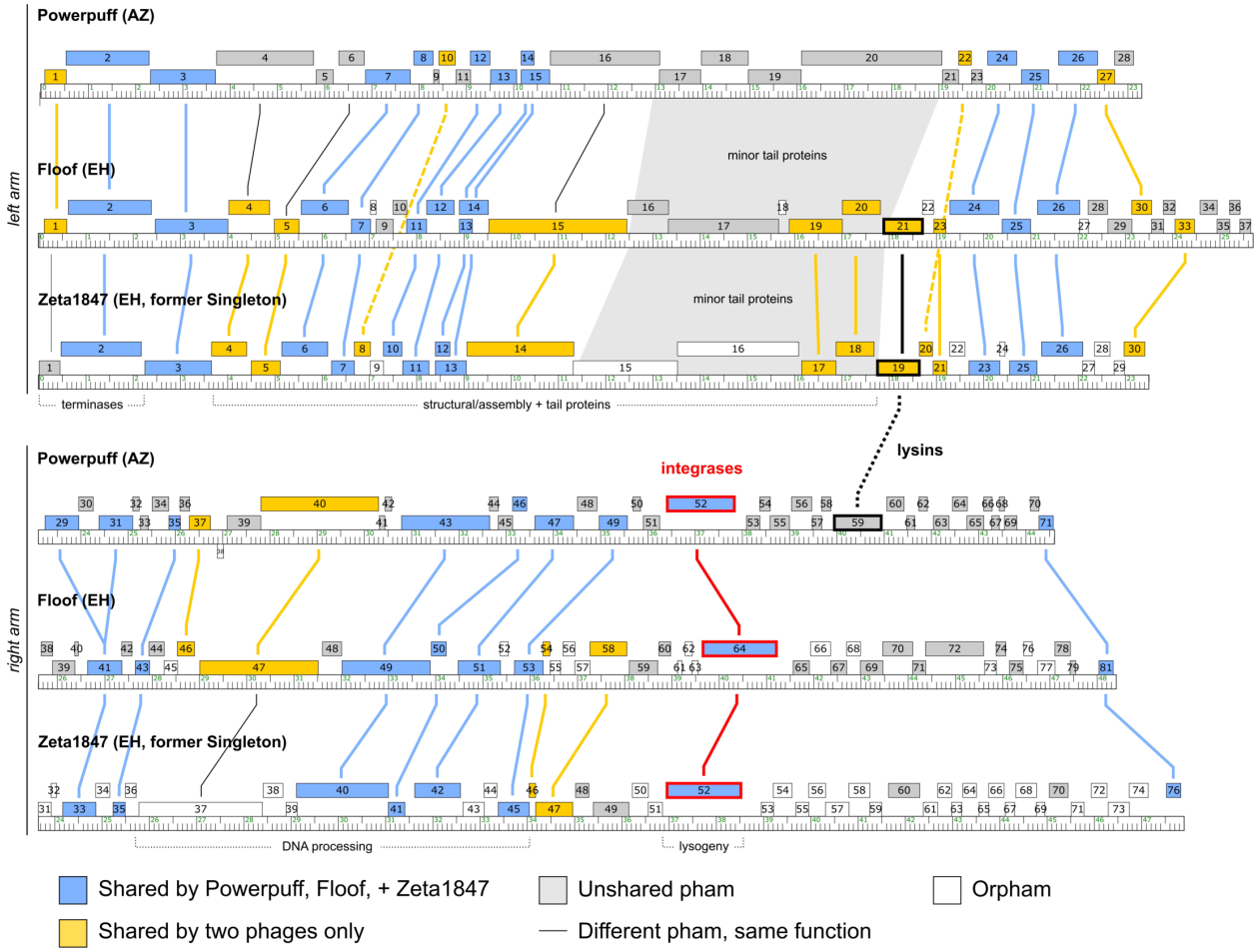

Fig 6: Clusters AZ and EH phages share similar genome architectures. Genome maps were downloaded from Phamerator and formatted using Inkscape 1.0. Genes in different phams with conserved functions are indicated by thin black lines and shaded regions. Integrases are highlighted in red, while lysins are highlighted in black. The left arm (top panel) of each genome was highly similar, with a less conserved right arm (bottom panel). Genes belonging to the same phams exhibited a conserved order. In Powerpuff, endolysin was found in the right arm rather than the left arm. Only Cluster AZ phages Elezi, Liebe, and Maureen encode endolysins in the left arm as do the Cluster EH phages.

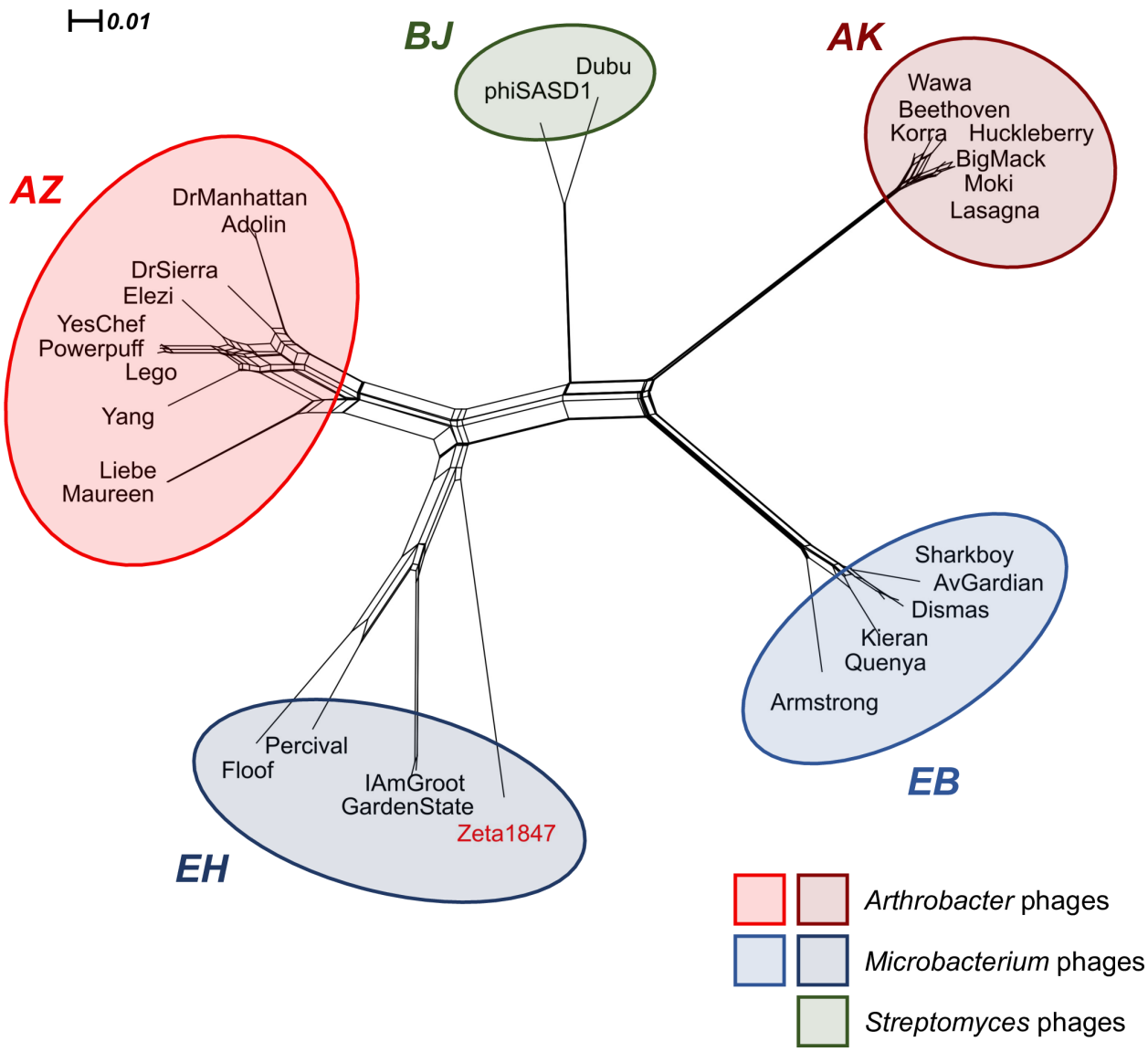

Fig 7: Cluster AZ is more similar to *Microbacterium* and *Streptomyces* phages than to any other *Arthrobacter* phages.
Pham information was obtained from the Actino_Draft database (version 382) and input into SplitsTree 4.16.1 to produce a network phylogeny using default parameters. Phage clusters are colored by host as indicated in the legend and the former singleton Zeta1847 is indicated in red. Each cluster formed a distinct branch on the tree. The most distant group of phages from Cluster AZ comprised Cluster AK, indicating low gene similarity despite being the most closely related *Arthrobacter* phages according to BLASTn.

Clusters AZ and EH phages encode for integrase and are likely temperate

An examination of the genes shared between Clusters AZ and EH revealed common functional biological features. Many DNA processing genes appeared to be shared between Clusters AZ and EH, including the genes encoding both terminase subunits, holliday junction resolvase, DNA polymerase I, DNA primase/helicase, and SprT-like protease. Many structural and virion assembly genes were also shared, including those encoding the portal protein, major capsid protein, head-to-tail adaptor, head-to-tail stopper, tail terminator, major tail protein, and tail assembly chaperone (Fig S1). These genes all encode for vital proteins involved in the phage life cycle and imply common biological features between these phages (9). Given this implication, we tested the lytic activity of phages Powerpuff, Lego, and YesChef on *Microbacterium foliorum*, the isolation host of Cluster EH phages Percival and Floof. Despite sharing considerable gene content with phages infecting *M. foliorum*, Powerpuff, Lego, and YesChef were unable to infect this host.

All Cluster AZ phages and all Cluster EH phages except Percival shared a pham encoding a serine integrase (pham 29778 as of 11/19/20). Cluster EH phage Percival also encoded a serine integrase assigned to a different pham (14036 as of 11/19/20). BLASTp alignment of these two phams (Floof_64 v. Percival_59) revealed 24.89% sequence identity over 87% query coverage, with an E-value of $2e^{-07}$, suggesting a distant relationship. While predicted to be temperate based on the presence of a serine integrase, none of these phages encoded a known repressor typically required for lysogenic maintenance and conferral of superinfection immunity (10). However, a

lysogenic life cycle was supported by each phage's turbid plaque morphology throughout isolation (37). Using patch and liquid release assays, we determined that phage Powerpuff forms stable lysogens and appears to be a true temperate phage (Fig 8). Furthermore, immunity assays showed that Cluster AZ phages Powerpuff, Lego, and YesChef were unable to infect Powerpuff lysogens, while the unrelated *Arthrobacter* Cluster FE phage BlueFeather, Cluster AU phage Giantsbane, and Cluster AO phage Abba retained their infectivity with reduced efficiency (Table S3). We thus expect that repressors likely remain unidentified in Clusters AZ and EH phages (38).

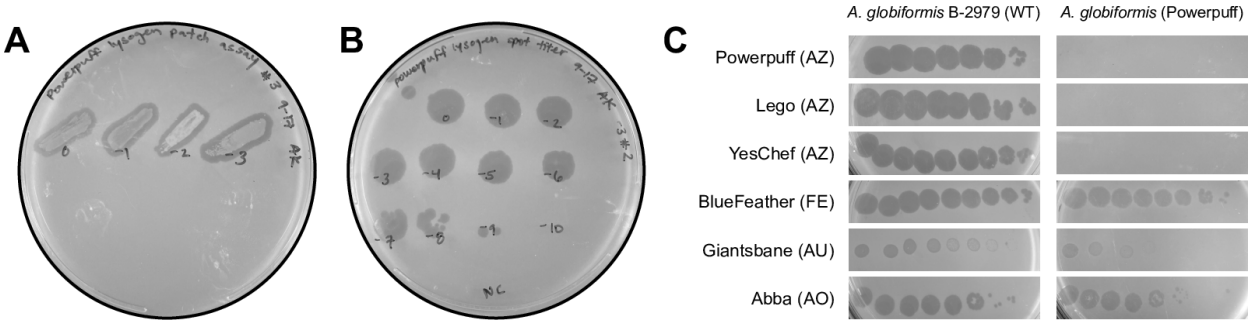

Fig 8: Phage Powerpuff forms stable lysogens that are immune to superinfection by related phages.
(A) Purified putative lysogens were streaked on a prepared lawn of *A. globiformis* B-2979. Zones of clearing were indicative of phage release. (B) Putative lysogens were grown in liquid culture then pelleted, after which supernatant was spot tested on *A. globiformis* B-2979 for titer. Plaque formations at high dilutions indicated spontaneous liquid release of phage particles and verified presence of stable lysogens. (C) 1:10 dilutions of *Arthrobacter* phage lysates were plated on WT *A. globiformis* B-2979 (left) and Powerpuff lysogens (right). Closely related phages were unable to lyse the Powerpuff lysogen, while unrelated phages retained infectivity with reduced efficiency.

Discussion

The goal of this study was to describe the novel *A. globiformis* phages Powerpuff, Lego, and YesChef, as well as to characterize their relationships to phages infecting a variety of actinobacterial hosts. Previous large scale studies of *Arthobacter* phages revealed minimal similarity to phages infecting non-*Arthrobacter* hosts (14).

However, more recent studies of *Microbacterium* phages have indicated similarities in both genome architecture and gene content between *Microbacterium* and *Arthrobacter* phages (10). These findings confirm that such relationships do exist in the environment and suggest that as we isolate new phages we will continue to identify cross-host relationships involving *Arthrobacter* phages.

As the number of sequenced actinobacteriophages increases, so does our understanding of the relationships among them. It was previously thought that phages infecting a common bacterial host would be most likely to exhibit increased genomic similarity (7). More recent studies have provided evidence of vast genomic diversity amongst phages infecting a common host (6), as well as instances in which phages infecting unique hosts display substantial genomic similarities (9). At the nucleotide level, the most similar phages to Powerpuff, Lego, and YesChef outside of Cluster AZ belonged to *Microbacterium* Clusters EH and EB, *Arthrobacter* Cluster AK, and *Streptomyces* Cluster BJ. We observed high OrthoANIu values for many pairwise comparisons between phages of unique clusters. The vast majority of such comparisons (excluding those between Clusters AZ and EH) exhibited coverage values below 5%. This indicates that while there is perhaps a widespread prevalence of shared, well-conserved genomic features amongst many of the phages included in this study, such features comprise only a small portion of each genome and are unlikely to represent a significant phylogenetic relationship. These results support previous findings which stated that *Arthobacter* phages are unlikely to share significant sequence similarity with actinobacteriophages infecting other host genera (14).

In general, phages Liebe and Maureen accounted for the least similarity in nucleotide comparisons both within Cluster AZ and between Clusters AZ and EH. These phages were as different from the other phages within their cluster as the Cluster AZ phages are collectively different from Cluster EH phages. The nucleotide dissimilarity of phages Liebe and Maureen from the remainder of Cluster AZ, at a level which is approximately equivalent to the similarities between Clusters AZ and EH, provides interesting evidence to further illustrate the diversity of Cluster AZ and the complexity of these phage relationships.

While nucleotide similarities between phage clusters were found to be minimal, amino acid comparisons yielded stronger alignments between almost all of the genomes analyzed. The increase in alignment strength at an equivalent amino acid word length indicated a history of synonymous substitutions and a distant evolutionary relationship amongst these phages, particularly those in Clusters AZ and EH (36).

Analyses of shared gene content further supported nucleotide and amino acid comparisons. GCS values between *Arthrobacter* phages in Cluster AZ and *Microbacterium* phages in Cluster EH either approached or exceeded the clustering parameter that has been applied to *Microbacterium* phages (10) and provided additional evidence for the close relationship between these phages. Though these values exceeded the new gene-content-based clustering parameter, we do not suggest that phages in Clusters AZ and EH should be clustered together. It is important to note that clustering parameters depend upon the available dataset of sequenced phages and do not reflect fundamental separation points between groups of phages (9). As more novel

phages are isolated it is expected that previously discrete clusters may become less well-separated, even amongst phages infecting unique actinobacterial hosts.

Previous application of the 35% gene-content-based clustering parameter to *Microbacterium* phages placed phage Zeta1847 as a singleton outside of any established clusters, given GCS values of only ~20% with the Cluster EH phages (10). Our updated GCS analysis revealed that the shared gene content of phage Zeta1847 with Cluster EH phages was in excess of the *Microbacterium* clustering parameter, indicating that Zeta1847 is less genomically isolated from clustered *Microbacterium* phages than was previously thought. This finding resulted in the placement of Zeta1847 into Cluster EH. As more phage genes are sequenced, pham assignments may change and reveal previously unidentified relationships between both novel and previously isolated phages. In this case, the close relationship between Zeta1847 and the rest of Cluster EH is evidenced functionally as well, given that these phages are the only isolated *Microbacterium* phages which are known to encode an integrase and which may be able to undergo lysogeny (10).

Phages in Clusters AZ and EH also shared a conserved genome architecture, with a high degree of synteny in the left arm of the genome and a similar order of conserved phams in the right arm. While some Cluster AZ phages encoded endolysins in the right arm of the genome, others encoded endolysins in the left arm as do the Cluster EH phages, evidencing variability in the similarity of genome architecture both within and outside of each cluster. Many genes in the right arm of these genomes were either orphams (no known gene homologs) or had no known function. It is possible that

as additional gene functions are assigned, further functional similarities and synteny will be observed between Clusters AZ and EH. A SplitsTree network phylogeny of shared gene content supported the proximity of the relationships between Clusters AZ and EH, while also providing additional evidence describing the diversity which exists amongst *Arthrobacter* phages (6,14).

Previous research has stated that some genes are thought to “travel together” when being exchanged amongst genomes, including tail genes and DNA replication genes (7). The functional significance of the genes which were shared between Clusters AZ and EH, including a large number of vital DNA processing, structural, and virion assembly genes, suggested common and conserved biological features and behaviors. Despite sharing considerable gene content with Cluster EH phages infecting *M. foliorum*, phages Powerpuff, Lego, and YesChef were unable to infect this host. This is not entirely unexpected, however, given that shared gene content does not necessitate an expanded host range. For instance, while the majority of Cluster A phages are known to infect mycobacterial hosts, phages belonging to the closely related Subcluster A15 are known to only infect *Gordonia* (39).

Surprisingly, phage Powerpuff was able to form stable lysogens that were immune to superinfection by phages Powerpuff, Lego, and YesChef, despite lacking an identifiable repressor gene typically required for maintenance of lysogeny and conferral of superinfection immunity. The other Cluster AZ phages, as well as most Cluster EH phages, also lacked repressor genes but encoded an integrase in the same pham as Powerpuff’s serine integrase, with the exception of Cluster EH phage Percival. Phages

are thought to be limited by the kinetics of DNA packaging (40), which offers limited "genomic real estate" and makes the long-term conservation of unused or non-functional genes, or their replacement with functional homologs, seemingly unlikely. Thus, if the Cluster EH phages were indeed lytic, it would be unexpected for phage Percival to encode a functionally homologous integrase gene in a different pham than the rest of these phages. Moreover, Cluster EH phages exhibit a genome architecture that has been previously described as distinct from known temperate phages (10). However, we find that these Cluster EH phages do in fact share a similar genome architecture with the Cluster AZ phages, including phage Powerpuff which is a true temperate phage. This lends secondary evidence suggesting that all Cluster AZ phages, as well as the Cluster EH phages, could be true temperate phages with a yet unidentified repressor. If true, this would make the Cluster EH phages the first identified *Microbacterium* phages which are able to undergo lysogeny (10). Further experiments investigating the ability of Cluster EH phages to form stable lysogens would be necessary to confirm their life cycles.

In sum, this research describes another case in which phages infecting different hosts share considerable genomic and biological similarities. As the first comparative study of phages in Cluster AZ, these findings describe, for the first time, significant conserved genomic content between *Arthrobacter* phages of Cluster AZ and phages of other actinobacterial hosts–particularly those belonging to Cluster EH, which infect *Microbacterium*. We also present the first experimental verification of lysogeny for a Cluster AZ phage, and suggest that other closely related phages lacking known

repressors, such as those in Cluster EH, may in fact be temperate. While the phage "puzzle" certainly remains incomplete, our findings serve to further illustrate the complexity of phage taxonomy and contribute to our understanding of actinobacteriophages and the characteristics which define them.

Acknowledgements

We would like to thank the Microbiology, Immunology, and Molecular Genetics Department, and the Dean of Life Sciences Division at UCLA for programmatic support. We thank Esther Choe, Byron Corado-Perez, Alexa Omholt, and Sophia Wang for their assistance in phage isolation, genome annotation, preliminary analysis, and original manuscript draft preparation. We also 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. We would also like to thank Daniel A. Russell and Graham Hatfull for clustering assistance and manuscript feedback. The authors acknowledge the use of instruments at the Electron Imaging Center for NanoMachines supported by NIH (1S10RR23057 to ZHZ) and CNSI at UCLA. The authors also acknowledge the Agricultural Research Service (ARS) Culture Collection for providing *Microbacterium foliorum* NRRL B-24224.

References

1. Keen EC. A century of phage research: Bacteriophages and the shaping of modern biology: Cause to reflect. BioEssays. 2015 Jan;37(1):6–9.

2. Clokie MRJ, Millard AD, Letarov AV, Heaphy S. Phages in nature. Bacteriophage. 2011 Jan;1(1):31–45.
3. Grose JH, Casjens SR. Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family Enterobacteriaceae. Virology. 2014 Nov;468–470:421–43.
4. Pratama AA, van Elsas JD. The 'Neglected' Soil Virome – Potential Role and Impact. Trends Microbiol. 2018 Aug;26(8):649–62.
5. Russell DA, Hatfull GF. PhagesDB: the actinobacteriophage database. Wren J, editor. Bioinformatics. 2017 Mar 1;33(5):784–6.
6. Hatfull GF. Dark Matter of the Biosphere: the Amazing World of Bacteriophage Diversity. Goodrum F, editor. J Virol. 2015 Aug 15;89(16):8107–10.
7. Hatfull GF, Hendrix RW. Bacteriophages and their genomes. Curr Opin Virol. 2011 Oct;1(4):298–303.
8. Hatfull GF. Bacteriophage genomics. Curr Opin Microbiol. 2008 Oct;11(5):447–53.
9. 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 Sep 6;8(4):e01069-17, /mbio/8/4/e01069-17.atom.
10. Jacobs-Sera D, Abad LA, Alvey RM, Anders KR, Aull HG, Bhalla SS, et al. Genomic diversity of bacteriophages infecting Microbacterium spp. Cloeckaert A, editor. PLOS ONE. 2020 Jun 18;15(6):e0234636.
11. Camargo FAO, Bento FM, Okeke BC, Frankenberger WT. Hexavalent Chromium Reduction by an Actinomycete, Arthrobacter crystallopoietes ES 32. Biol Trace Elem Res. 2004;97(2):183–94.
12. O'Loughlin EJ, Sims GK, Traina SJ. Biodegradation of 2-methyl, 2-ethyl, and 2-hydroxypyridine by an Arthrobacter sp. isolated from subsurface sediment. Biodegradation. 1999;10(2):93–104.
13. Westerberg K, Elväng AM, Stackebrandt E, Jansson JK. Arthrobacter chlorophenolicus sp. nov., a new species capable of degrading high concentrations of 4-chlorophenol. Int J Syst Evol Microbiol. 2000 Nov 1;50(6):2083–92.
14. 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. Schuch R, editor. PLOS ONE. 2017 Jul 17;12(7):e0180517.
15. 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. Kolter R, editor. eLife. 2015 Apr 28;4:e06416.
16. 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 Mar;397(1):119–43.
17. 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. Complete Genome Sequences of 61 Mycobacteriophages. Genome Announc. 2016 Aug

25;4(4):e00389-16, /ga/4/4/e00389-16.atom.
18. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012 Jul;9(7):671–5.
19. Russell DA. Sequencing, Assembling, and Finishing Complete Bacteriophage Genomes. In: Clokie MRJ, Kropinski AM, Lavigne R, editors. Bacteriophages [Internet]. New York, NY: Springer New York; 2018 [cited 2020 Sep 10]. p. 109–25. (Methods in Molecular Biology; vol. 1681). Available from: http://link.springer.com/10.1007/978-1-4939-7343-9_9
20. Besemer J, Borodovsky M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 2005 Jul 1;33(Web Server):W451–4.
21. Delcher A. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 1999 Dec 1;27(23):4636–41.
22. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics. 2011 Dec;12(1):395.
23. SEA-PHAGES/Starterator [Internet]. SEA-PHAGES Program; 2016. Available from: https://github.com/SEA-PHAGES/starterator
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct;215(3):403–10.
25. 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 Jan 28;43(D1):D222–6.
26. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005 Jul 1;33(Web Server):W244–8.
27. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Edited by F. Cohen. J Mol Biol. 2001 Jan;305(3):567–80.
28. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res. 2015 Jul 1;43(W1):W401–7.
29. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics. 2007 Apr 15;23(8):1026–8.
30. Yoon S-H, Ha S, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017 Oct;110(10):1281–6.
31. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics. 1998 Feb 1;14(1):68–73.
32. Yuan Y, Gao M. Jumbo Bacteriophages: An Overview. Front Microbiol [Internet]. 2017 Mar 14 [cited 2020 Sep 10];8. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2017.00403/full
33. Kalatzis P, Rørbo N, Castillo D, Mauritzen J, Jørgensen J, Kokkari C, et al. Stumbling across the Same Phage: Comparative Genomics of Widespread Temperate Phages Infecting the Fish Pathogen Vibrio anguillarum. Viruses. 2017

May 20;9(5):122.
34. Lee I, Ouk Kim Y, Park S-C, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016 Feb 1;66(2):1100–3.
35. 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. Dykeman EC, editor. PLOS ONE. 2021 Mar 12;16(3):e0248418.
36. Koonin EV, Galperin MY. Sequence - evolution - function: computational approaches in comparative genomics. Boston: Kluwer Academic; 2003. 461 p.
37. Kropinski AM, Mazzocco A, Waddell TE, Lingohr E, Johnson RP. Enumeration of Bacteriophages by Double Agar Overlay Plaque Assay. In: Clokie MRJ, Kropinski AM, editors. Bacteriophages [Internet]. Totowa, NJ: Humana Press; 2009 [cited 2020 Sep 10]. p. 69–76. (Walker JM, editor. Methods in Molecular Biology; vol. 501). Available from: http://link.springer.com/10.1007/978-1-60327-164-6_7
38. Dedrick RM, Jacobs-Sera D, Bustamante CAG, Garlena RA, Mavrich TN, Pope WH, et al. Prophage-mediated defence against viral attack and viral counter-defence. Nat Microbiol. 2017 Mar;2(3):16251.
39. Hatfull GF. Mycobacteriophages. Microbiol Spectr [Internet]. 2018 Oct 5 [cited 2020 Sep 10];6(5). Available from: http://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.GPP3-0026-2018
40. Hendrix RW. Bacteriophages: Evolution of the Majority. Theor Popul Biol. 2002 Jun;61(4):471–80.

Supporting Information

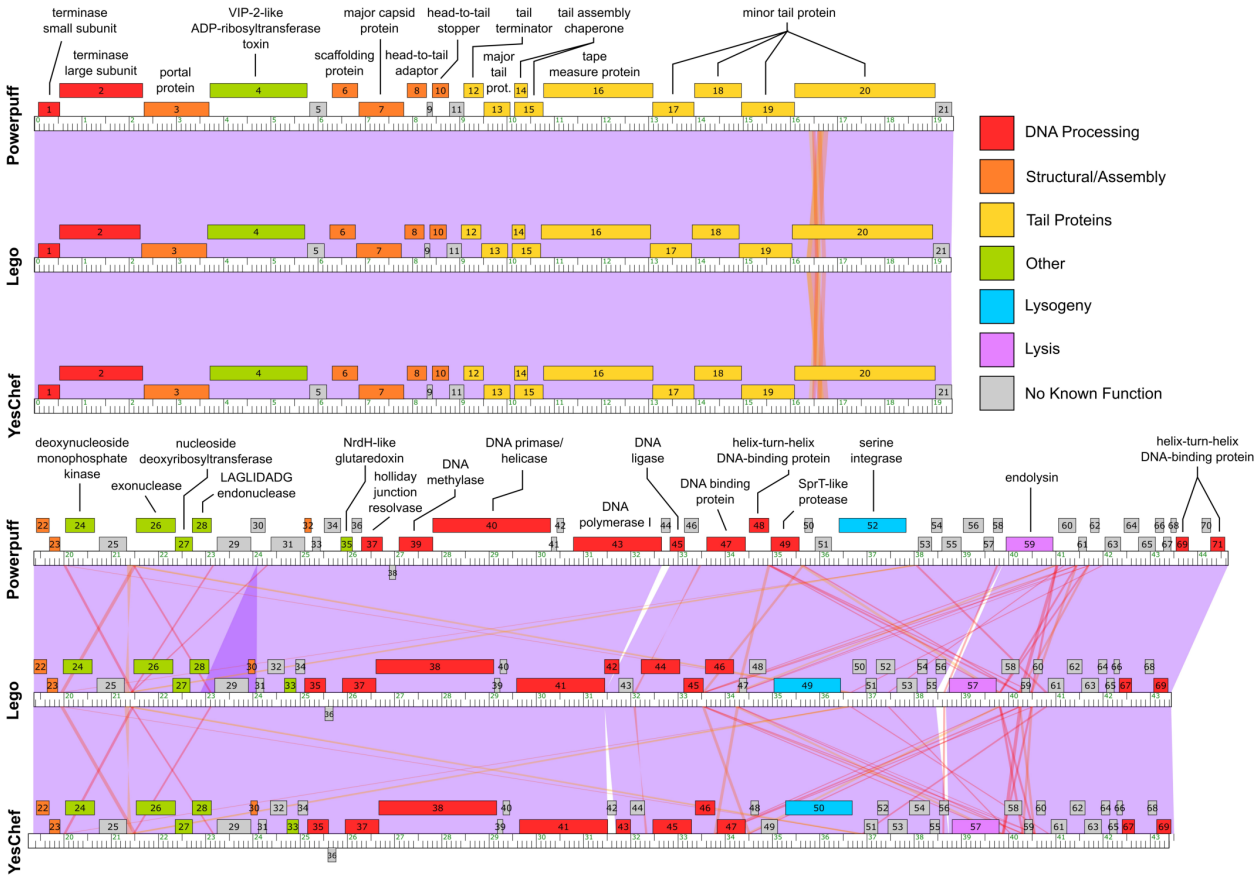

Fig S1: Powerpuff, Lego, and YesChef have highly similar genomes. Genomes were downloaded from Phamerator and formatted using Inkscape 1.0. Genes were sorted by general function or type, as indicated in the legend above. Powerpuff, Lego, and YesChef have highly similar genomes, with pairwise BLASTn scores of over 98.63% identity with at least 99% coverage and E-values of 0. Each genome was found to be between 43,446 and 44,651 bp in length, encoding between 69 and 71 genes. Notable dissimilarities included a gene duplication in phage Powerpuff (Powerpuff_29 and Powerpuff_31), which twice encoded a gene of unknown function that was present only once in phages Lego and YesChef. Phage Lego was found to encode a gene of unknown function (Lego_56) not found in Powerpuff or YesChef, located directly upstream of the gene encoding an endolysin. Phages Powerpuff and YesChef also encoded a gene of unknown function (Powerpuff_44 and YesChef_42) not found in phage Lego. This gene was positioned within a cassette of DNA processing genes in these phages.

Table S1: NCBI BLASTn results for Powerpuff, Lego, YesChef, Tbone, and Kaylissa.

Note: E-values for all comparisons were 0.0

	Query Phage	Subject Phage				
		Powerpuff	**Lego**	**YesChef**	**Tbone**	**Kaylissa**
Coverage (%)	**Powerpuff**	100	99	100	94	98
	Lego	99	100	99	93	98
	YesChef	100	99	100	93	98
	Tbone	92	92	92	100	94
	Kaylissa	97	97	97	94	100
Identity (%)	**Powerpuff**	100	99.10	100	96.26	96.86
	Lego		100	98.63	96.52	97.23
	YesChef			100	96.26	96.86
	Tbone				100	95.26
	Kaylissa					100

Table S2: PhagesDB BLASTn scores for Powerpuff query.

Host genera	Cluster	Phage	Score (bits)	E-value
Arthrobacter	AZ	Powerpuff	8.85e+04	0
Arthrobacter	AZ	YesChef	4.78e+04	0
Arthrobacter	AZ	Lego	4.29e+04	0
Arthrobacter	AZ	Tbone	1.53e+04	0
Arthrobacter	AZ	Kaylissa	1.46e+04	0
Arthrobacter	AZ	Elezi	3416	0
Arthrobacter	AZ	DrSierra	2397	0
Arthrobacter	AZ	Yang	2089	0
Arthrobacter	AZ	DrManhattan	1475	0
Arthrobacter	AZ	Adolin	1475	0
Arthrobacter	AZ	Maureen	765	0
Arthrobacter	AZ	Liebe	765	0
Microbacterium	EH	Percival	244	3e-61
Microbacterium	EH	Zeta1847	240	4e-60
Microbacterium	EH	GardenState	232	1e-57
Microbacterium	EH	IAmGroot	230	4e-57
Microbacterium	EH	Floof	226	6e-56
Microbacterium	EB	Sharkboy	131	3e-27
Arthrobacter	AK	Kittykat	129	1e-26
Microbacterium	EB	Franklin22	125	2e-25
Arthrobacter	AK	BigMack	125	2e-25
Microbacterium	EB	Dismas	123	7e-25
Arthrobacter	AK	Moki	117	4e-23
Arthrobacter	AK	Huckleberry	117	4e-23
Microbacterium	EB	Kieran	107	4e-20
Microbacterium	EB	AvGardian	107	4e-20
Microbacterium	EB	Quenya	103	6e-19
Streptomyces	BJ	Dubu	101	3e-18
Microbacterium	EB	Armstrong	100	1e-17
Arthrobacter	AK	Wawa	94	6e-16
Arthrobacter	AK	Korra	94	6e-16
Arthrobacter	AK	Beethoven	94	6e-16
Arthrobacter	AK	Lasagna	78	4e-11
Streptomyces	BL	phiSASD1	56	1e-04

Table S3: Efficiencies of Plating (EOP) for phages spot titered on Powerpuff lysogens, relative to WT *A. globiformis* B-2979.

Phage	Cluster	EOP
Powerpuff	AZ	0
Lego	AZ	0
YesChef	AZ	0
BlueFeather	FE	8.3×10^{-2}
Giantsbane	AU	6.25×10^{-2}
Abba	AO	3.67×10^{-2}