Virology 538 (2019) 45–52

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

A cornucopia of *Shigella* phages from the Cornhusker State

Sarah M. Doore[a,b,3], Jason R. Schrad[a,3], Hailee R. Perrett[a,1,3], Kevin P. Schrad[c], William F. Dean[a,2], Kristin N. Parent[a,b,*]

[a] Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, 48824, USA
[b] BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI, 48824, USA
[c] Department of Science, Lincoln Southwest High School, Lincoln, NE, 68512, USA

ARTICLE INFO

Keywords:
Bacteriophage hunting
Shigella bacteriophage
Autographivirinae
Ounavirinae
Mooglevirus
High school research
Bacteriophage diversity
Virus ecology
Genomics
Phylogenetics

ABSTRACT

Bacteriophages are abundant in the environment, yet the vast majority have not been discovered or described. Many characterized bacteriophages infect a small subset of *Enterobacteriaceae* hosts. Despite its similarity to *Escherichia coli*, the pathogenic *Shigella flexneri* has relatively few known phages, which exhibit significant differences from many *E. coli* phages. This suggests that isolating additional *Shigella* phages is necessary to further explore these differences. To address questions of novelty and prevalence, high school students isolated bacteriophages on non-pathogenic strains of enteric bacteria. Results indicate that *Shigella* phages are abundant in the environment and continue to differ significantly from *E. coli* phages. Our findings suggest that *Shigella*-infecting members of the *Ounavirinae* subfamily continue to be over-represented and show surprisingly low diversity within and between sampling sites. Additionally, a podophage with distinct genomic and structural properties suggests that continued isolation on non-model species of bacteria is necessary to truly understand bacteriophage diversity.

1. Introduction

Bacteriophages, or viruses that infect bacteria, are estimated to be the most abundant biological entities on the planet (Hendrix et al., 1999; Suttle, 2005). With at least 10^{31} viral particles in the biosphere, there is great diversity among these viruses in terms of host organism, genomic content, morphology, and infection cycles (Pope et al., 2015; Suttle, 2005). Given that only a fraction of bacteria can be cultured, our knowledge of bacteriophages—or phages—is limited. In addition, only a subset of culturable bacteria have been used to isolate a majority of the currently described phages (Hatfull and Hendrix, 2011).

One genera of bacteria that is understudied in terms of its phages is *Shigella*, which contains four species: *S. boydii*, *S. dysenteriae*, *S. flexneri*, and *S. sonnei*. These bacteria are all causative agents of shigellosis, a type of bacillary dysentery that affects an estimated 164.7 million people each year (Kotloff et al., 1999). Despite the global prevalence and burden of *Shigella* bacteria, very few *Shigella*-infecting phages had been described until recent years. In 2016, concurrent with a dysentery outbreak in the region, 16 new *Shigella* phages were isolated from environmental samples collected in central Michigan (Doore et al., 2018). The overwhelming majority of these new *Shigella* phages were most similar to the *Citrobacter* phage Moogle within the *Ounavirinae* subfamily of myoviruses. Members of this subfamily have uncommon genome sizes of 85.0–95.0 kbp and similarly uncommon T = 9 capsid geometry. In the previous study, relatively few phage isolates were siphoviruses (2 out of 18), and no podoviruses were found. In addition, all phages were isolated from surface water samples, which led us to ask: 1) how common are *Shigella* phages in the environment, 2) in what sample type(s) are they most prevalent, 3) is T = 9 geometry common in *Shigella* phages, and 4) is the observed increase in 2016 unique to one area, or are *Shigella* phages found in other geographical regions as well?

In order to address these questions, we targeted the distant midwestern state of Nebraska at a site approximately 1200 km from the original Michigan isolation zone. On average, Nebraska is drier and warmer than Michigan (Arguez et al., 2012), leading to the possibility of different phage environments. Nebraska and Michigan also have different primary water sources, providing another potential avenue for the discovery of new phages. Most of the water in Michigan stems from

* Corresponding author. Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, 48824, USA.
E-mail address: kparent@msu.edu (K.N. Parent).
[1] Present address: Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.
[2] Present address: Department of Cell, Developmental, and Integrative Biology, University of Alabama at Birmingham, Birmingham, AL 35205, USA.
[3] These authors contributed equally to the work.

https://doi.org/10.1016/j.virol.2019.09.007
Received 12 June 2019; Received in revised form 5 September 2019; Accepted 16 September 2019
Available online 18 September 2019
0042-6822/ © 2019 Elsevier Inc. All rights reserved.

the Great Lakes, particularly Lake Michigan and Lake Superior. Nebraska, on the other hand, gets most of its water from rivers starting in the Rocky Mountains. Finally, the location where we worked with students in Lincoln, Nebraska lies on a unique salt marsh (Harvey et al., 2007), which is an environment that has not been extensively studied for phages. In Nebraska, we challenged high school students to collect environmental samples, then screen them for the presence of bacteriophages. As was done in Michigan, the students tested their samples for phages that could infect non-pathogenic strains of the enteric bacteria *S. flexneri* (serotype Y), *E. coli* (K1/K-12), and *S. enterica* sv. Typhimurium. The overall goal of this work was to determine the relative frequency of *S. flexneri* phage isolation compared to total isolation of enteric bacteria-infecting phages and to involve students in real-world scientific processes.

In this study, we characterized 12 individual phages isolated from a total of 50 samples, including 10 Moogle-like myoviruses and two podoviruses. Similar to Michigan, the Moogleviruses continued to be overrepresented in terms of isolation frequency. These viruses exhibit relatively low genetic diversity both within and between geographic regions, with the exception of the gene encoding a putative capsid decoration protein. These decoration proteins, which exhibit low structural conservation, have been described in phages such as lambda (Lander et al., 2008), T5 (Vernhes et al., 2017), and L (Gilcrease et al., 2005). Despite structural differences, these decoration proteins appear to stabilize phage capsids. Though typically annotated as a tail protein, we have determined that this gene is widespread throughout the *Ounavirinae* subfamily of myoviruses. Finally, the podoviruses proved to be both elusive and unique, sharing only 10% nucleotide identity to any other previously-described bacteriophage. These isolates share some morphologic properties with T7 but may otherwise form an independent species of podoviruses.

2. Materials and methods

2.1. Bacterial strains

The following strains of bacteria used for host range tests and phage amplification have been previously described in (Doore et al., 2018): *S. flexneri* serotypes Y (PE577), 2a (CFS100), and 5a (SD100); *E. coli* K1/K-12 (EV36), B (REL606), and C (C122); and *S. enterica* sv. Typhimurium (DB7136). Shannon Manning at the Shiga Toxin-Producing *Escherichia coli* (STEC) Center at Michigan State University kindly provided *S. boydii* serotype 13 (ATCC 12032), *S. dysenteriae* serotype 1 (strain CDC-3823/69), and *S. sonnei* (strain 16372).

2.2. Phage isolation and purification

High school students were introduced to the concept of bacteriophage and where enteric-infecting phages might be found in the environment by a short, lecture-style class period. Students were given a 50 ml conical tube and instructed to collect samples from either the school campus or from home after thinking about where they would likely find phages. These samples, reported in Supplementary Table 1, included pond or swamp water (n = 27), water or soil from athletic fields (n = 9), creek water (n = 3), soil from backyards (n = 2), liquid from dogs' water bowls (n = 2), soil from a deer footprint (n = 1), or unmarked locations (n = 6). Luria broth (LB) was added to solid media to a final volume of 50 ml and thoroughly mixed, then particulates were allowed to settle. Supernatant from the settled solid material andliquid samples were filtered through a 0.45 μm filter to remove bacteria and any particulates. 250 μl aliquots of filtrate were added to LB plates with 0.3% top agar overlays seeded with test strains of bacteria: PE577 (*S. flexneri)*, EV36 (*E. coli*), or DB7136 (*S. enterica* sv. Typhimurium). Plates were incubated overnight at 37 °C and screened for plaque formation. Original samples and plates containing plaques were both sent to Michigan State University for subsequent purification and amplification in liquid culture, as described in (Doore et al., 2018).

2.3. Negative staining and transmission electron microscopy

Negative stain transmission electron microscopy (TEM) was used to determine the gross particle morphology of the plaque purified and amplified phage isolates. Small aliquots (3–5 μL) of phage were applied to continuous carbon grids that had been glow discharged for 35 s using a Pelco EasiGlow glow discharger. The phage were adsorbed to the carbon for approximately 30 s, after which excess sample was washed from the grid. The remaining particles were stained with 1% uranyl acetate. The stained samples were imaged using a Talos Arctica operating at 200 keV with 2.5 s exposures collected on a Ceta CCD detector at 45,000× nominal magnification (2.24 Å/pixel).

2.4. Consolidation of bacteriophage isolates

Isolates were grouped to reduce the resources necessary to fully characterize all phages; consequently, only one representative of each group was used for all future experiments. This consolidation was done by three methods—host range assays, protein profiles, and restriction fragment length polymorphism analysis—to ensure groups were formed with high confidence. Initially, host range assays were performed by spotting purified phage onto LB agar plates seeded with test strains of bacteria. Quantitative host range assessment was then carried out by plaque assays as previously described (Doore et al., 2018) (Table 1).

Table 1
Host range of newly-isolated bacteriophages. Spot tests and/or quantitative plaque assays were performed for all phage-host combinations. Efficiency of plating frequencies at or above 10^{-8} are reported as numbers, whereas frequencies below 10^{-8} are indicated with a dashed line.

	Host									
Phage	*S. boydii*	*S. dysenteriae*	*S. flexneri* Y	*S. flexneri* 2a	*S. flexneri* 5a	*S. sonnei*	*E. coli* B	*E. coli* C	*E. coli* K1/K12	*S. enterica* sv. Typhimurium
10	–	–	1.0	–	–	–	–	–	–	–
11	–	–	1.0	0.1	1.0	–	–	–	–	–
17	–	–	1.0	0.1	1.0	–	–	–	–	–
KPS64	–	–	0.8	10^{-2}	1.0	–	–	–	–	–
24	–	–	1.0	0.1	1.0	–	–	–	–	–
HRP29	–	–	1.0	–	–	–	–	–	–	–
32	–	–	0.3	1.0	0.1	–	–	–	–	–
44	–	–	1.0	0.1	0.8	–	–	–	–	–
CHB7	–	–	10^{-5}	10^{-5}	10^{-5}	$\sim 10^{-8}$	10^{-5}	1.0	1.0	–
48	–	–	0.7	0.1	1.0	–	–	–	–	–
49	–	–	1.0	0.1	0.2	–	–	–	–	–
Silverhawkium	–	–	0.8	0.1	1.0	–	–	–	–	–

To examine protein profiles, samples were TCA precipitated following standard procedure, ensuring that the samples were normalized to approximately 1.0×10^{10} plaque forming units (pfu) prior to precipitation. After resuspension in equal volumes of gel loading buffer, samples were loaded on 15% acrylamide gels alongside a Bio-Rad Precision Plus protein standard. Electrophoresis proceeded at 200 V for approximately 1 h, then gels were stained with Coommassie Blue. To examine restriction enzyme patterns, 0.4 μg of phage genomic DNA was digested with EcoRI and/or HindIII for 2 h at 37 °C. Fragments were examined by agarose gel electrophoresis alongside an Invitrogen 1 kb Plus DNA standard. For this, 1.2% agarose gels containing 0.5 μg/ml ethidium bromide were electrophoresed at 100 V for 2 h. Bacteriophages with identical restriction fragment length polymorphism patterns *and* identical protein profiles were grouped together.

2.5. Genome sequencing and annotation

Four representative phage isolates were chosen for whole-genome sequencing through MicrobesNG at the University of Birmingham, UK. Genomic DNA was extracted from 1×10^{10}- 1×10^{11} phage particles using phenol-chloroform as described in (Dover et al., 2016). Purified genomic DNA was quantified in triplicate with the Quantit dsDNA HS assay in an Ependorff AF2200 plate reader, then libraries were prepared using Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: 2 ng of DNA instead of 1 were used as input, and PCR elongation time was increased to 1 min from 30 s. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system. Pooled libraries were quantified using the Kapa Biosystems Library Quantification Kit for Illumina on a Roche light cycler 96 qPCR machine. Libraries were sequenced on the Illumina HiSeq using a 250 bp paired end protocol.

Raw reads were trimmed and assembled into contigs using the A5 pipeline (Tritt et al., 2012), version 08-25-2016. Open reading frames were identified using GeneMark.hmm S (Besemer et al., 2001) and annotated using blast2GO (Conesa et al., 2005; Gotz et al., 2008). Since these genomes are circularly permuted, they were aligned with the genome of the most closely-related reference sequence—Moogle for the Sf13-likes and KP34 for HRP29—to determine "ends" of the genome. tRNAs were identified using tRNAscan-SE (Lowe and Chan, 2016). When comparing genomes or coding regions across phages, average nucleotide identity (ANI) was calculated as percent coverage multiplied by percent identity. Protein-based comparisons were visualized using BLAST Image Ring Generator (BRIG) version 0.95 (Alikhan et al., 2011). For HRP29, proteins found in the mature virion were confirmed via mass spectrometry on excised and trypsin-digested bands from a 10% SDS-PAGE gel. Mass spectrometry experiments were performed by the Michigan State University Proteomics Core.

2.6. Phylogenetic analysis

Primers to amplify genes encoding the major capsid protein, the putative Mooglevirus decoration protein, and the large terminase were designed according to observed variation across the genus. The following primers were used: capsid, 5′-CAG TNT CGG TGC AGA GGA TTC-3′ and 5′-GCG AAA GGT ATT GAT TTC GTC CC-3'; decoration protein, 5′-GGT GAN GTN TCA CCN CAA CAC AAA GG-3′ and 5′-CCT TTA CGT GAN AGT TNC ATT ATC AAC CTC C-3'; large terminase, 5′-GGA AGG CTT TCT TGT GTT TCT GTG-3′ and 5′-GGT GTA GAA NCT AAA CAG GAA TCT G-3'. PCR products were Sanger sequenced at the Research Technology Support Facility – Genomics Core at Michigan State University. Whole genomes or Sanger-sequenced regions were aligned using Clustal Omega (Sievers et al., 2011). From this alignment, phylogenetic trees were generated using MrBayes version 3.2.6 under a mixed model with haploid genome and gamma variation settings (Ronquist and Huelsenbeck, 2003). Two parallel replicates were simultaneously run until convergence with a standard deviation below 0.05. Trees were viewed using FigTree version 1.4.3 (Rambaut, 2012).

3. Results

3.1. A majority of newly-isolated bacteriophages from Nebraska infected Shigella

A total of 50 samples were collected by students from two classrooms. All sample locations are presented in Supplementary Table 1, including "reported" and "interpreted" locations, as certain labels were incomplete or rubbed away between collection and analysis. From the material collected, 12 samples (marked with * and in bold font in Supplementary Table 1) contained detectable phages. All 12 phages were subsequently purified and characterized more thoroughly, including a test of host range on non-pathogenic strains of *S. flexneri*, *E. coli*, and *S. enterica* sv. Typhimurium. Efficiency of plating data shown in Table 1 indicates that 11 of the 12 infected *S. flexneri* exclusively. The remaining phage—sample 47, officially named CHB7—primarily infected *E. coli*, with the ability to infect *S. flexneri* at a much lower frequency. No isolates were able to form plaques on *S. enterica* sv. Typhimurium. These results are similar to the host range of bacteriophages isolated in Michigan in 2016, with the vast majority of phages being specific to *S. flexneri* (Doore et al., 2018). As in 2016, a majority of phages were also able to infect all three strains of *S. flexneri* used for testing, representing serotypes Y, 2a, and 5a. Only one Nebraska isolate was restricted to *S. flexneri* serotype Y.

The 12 samples that tested positive for phages in Nebraska were shipped to Michigan State University for subsequent analysis. Upon arrival at Michigan State University, all bacteriophages were tested against a broader panel of potential host strains. This panel included pathogenic strains of *S. boydii*, *S. dysenteriae*, and *S. sonnei* in addition to *S. flexneri*. Unlike the 2016 Michigan bacteriophages, isolates from Nebraska were not able to infect any other *Shigella* strains.

3.2. Most phages were similar to Moogleviruses in morphology and genomic content

Once purified, bacteriophages were subjected to a battery of tests to classify them into groups. Although 12 isolates were purified, some of these were collected within close proximity and may therefore have belonged to the same species. To consolidate phages into groups, we first performed a preliminary characterization of host range (described above), imaging by transmission electron microscopy, restriction digestion of genomic DNA, and analysis of protein composition by SDS-PAGE. Based on these data, we then identified a single isolate from each group to serve as a representative. This was done to reduce the time and materials required for a thorough analysis of every isolate, since we likely had redundancy in our sampling.

After host range had been determined, negative stain transmission electron microscopy revealed that ten of the new isolates were myoviruses, having long, contractile tails. These also had icosahedral capsids of about 80 nm and short tail fibers, consistent with the "Moogle-like" appearance of previous *Shigella* phage isolates (Doore et al., 2018). The two remaining isolates, "10" and "29," were short-tailed podoviruses. These data are shown in Fig. 1.

Next, genomic DNA was extracted and digested with the restriction enzymes EcoRI and/or HindIII. Restriction fragment length polymorphisms (RFLPs) were subsequently analyzed to determine which phages produced identical patterns (Fig. 2A and B). Finally, purified phage particles were analyzed by SDS-PAGE to determine protein composition (Fig. 2C and D). Phages with identical morphology, RFLP, and protein composition were then grouped together.

With these data combined, all phages were classified into four groups, represented by the following isolates: HRP29 (Sample 29;

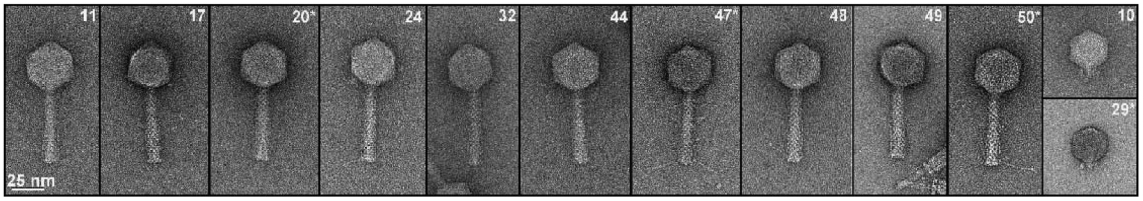

Fig. 1. Negative stain micrographs of all bacteriophages isolated. Representatives chosen for in-depth characterization are marked with *. Sample numbers indicate the phage identifier prior to naming, where 20 = KPS64, 29 = HRP29, 47 = CHB7, and 50 = Silverhawkium.

named by Hailee R. Perrett), CHB7 (Sample 47; named by Charley H. Bittle), KPS64 (Sample 20; named by Kevin P. Schrad), and Silverhawkium (Sample 50; named collectively by the highschool class). To analyze the genomic content of each group, whole genomes from these representative phages were sequenced in full. Basic characteristics are listed in Table 2. A majority of the newly-isolated phages, represented by myoviruses KPS64 or Silverhawkium, shared the greatest average nucleotide identity (ANI) with Michigan isolates *Shigella* phage Sf17 and *Citrobacter* phage Moogle isolated in Texas (GenBank accessions MF327004 and NC_027293, respectively). The *Escherichia* phage CHB7 was more similar to the *Escherichia* phages SUSP1 and SUSP2 isolated from the east coast (accessions KT454805 and KT454806) than to either Sf17 or Moogle, further distinguishing it from the other isolates. The two HRP29-like podophages were not clearly related to any phage in available databases based on ANI. These phage appeared to be members of the *Autographivirinae* subfamily based on morphology, and specifically the Drulisvirus genus based on synteny.

Given the low similarity to other phages, we performed mass spectrometry on purified HRP29 phage particles to determine which of these proteins are contained or associated within the mature virions. The results from this experiment, reported in Table 3, indicate that many of the annotated hypothetical proteins may be structural proteins. They are considered hypothetical proteins as BLAST searches (both using nucleotide and amino acid sequences) did not reveal any obvious homology to proteins of known function. Further studies will need to be conducted to determine the role of these proteins in HRP29 biology.

3.3. Overall inter- and intrastate diversity is low for Sf13-and Moogle-type viruses

The three myoviruses CHB7, KPS64, and Sivlerhawkium exhibited slightly different host ranges, RFLP patterns, and protein profiles (Table 1, Fig. 2), but their genomic content was strikingly similar. Since CHB7 showed the greatest difference in host range, we suspected this phage would be least similar to other Moogleviruses or to FelixO1, the type member of the *Ounavirinae*; however, a genome map, shown in Fig. 3A, suggests that CHB7 appears to be similar to previously-isolated Sf13 phages and to Moogle phages. Based on coding sequences, the three phages from Nebraska also aligned extremely well with Sf13, Sf14, Moogle, and FelixO1 (Fig. 3B).

To determine additional similarities between the other Sf13-and Moogleviruses isolated, for which whole genomes were not available, three structural genes were chosen for further analysis: the putative decoration protein, the major capsid protein, and the large terminase. In Sf13- and Moogle-type phages, the decoration protein refers to an open reading frame of approximately 370 amino acids which contains

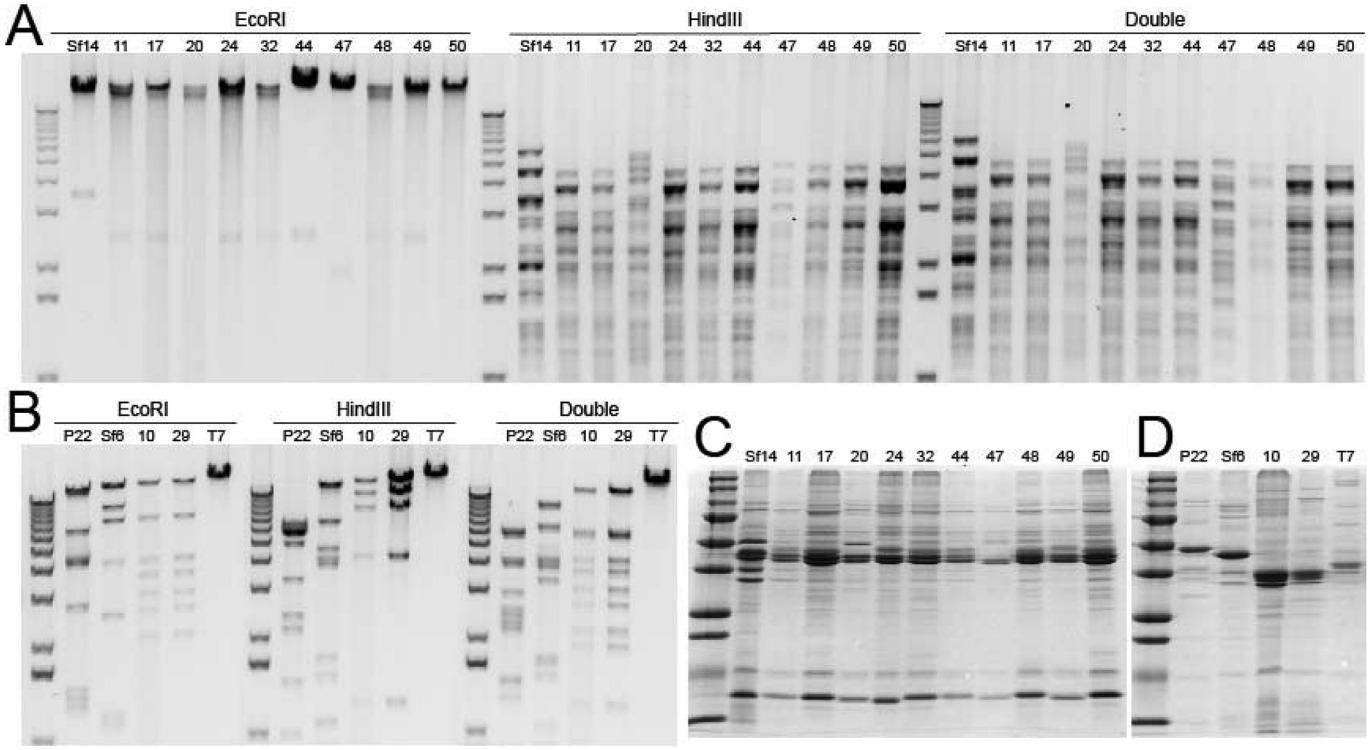

Fig. 2. Genome and protein profiles of phage isolates. For genome profiles, restriction fragment length polymorphisms are shown of A) *Ounavirinae* or B) podovirus phage genomic DNA digested with the indicated restriction enzyme(s). The left-most lane of each gel is a 1 kb ladder run as a standard. For protein profiles, SDS-PAGE are shown of C) *Ounavirinae* or D) podovirus particles. Sample numbers indicate the phage identifier prior to naming.

Table 2
Genomic and phylogenetic characteristics of newly-isolated, representative phages.

Phage	No. Isolates	Genome Length (kbp)	No. ORFs	No. tRNAs	% GC	Morphology	Taxonomic Subfamily (Genus)	% ANI to similar or reference genomes	GenBank Accession
HRP29	2	43.6	53	0	54.1	Podo	*Autographi-virinae*	10.0 to phiKDA1 3.2 to KP34	MK562503
CHB7	1	88.0	134	26	40.0	Myo	*Ounavirinae* (Suspvirus)	89.0 to SUSP2 85.9 to SUSP1	MK562504
Silver-hawkium	7	88.8	135	26	39.1	Myo	*Ounavirinae* (Mooglevirus)	85.0 to Sf17 80.6 to Moogle	MK562505
KPS64	1	90.2	137	26	39.0	Myo	*Ounavirinae* (Mooglevirus)	95.0 to Sf17 84.3 to Moogle	MK562502

Table 3
Proteins detected by mass spectrometry of purified HRP29 particles.

MW (kDa)	Protein	gp	Accession
139	Internal virion protein	43	QBP32943.1
94	Internal virion protein	42	QPB32942.1
85	Tail protein	40	QBP32940.1
57	Portal protein	35	QBP32935.1
54	Tailspike protein	52	QBP32952.1
37	Capsid protein	37	QBP32937.1
29	Scaffolding protein	36	QBP32936.1
29	Tail adapter protein	44	QBP32944.1
22	Internal virion protein	41	QBP32941.1
21	Tail tube protein	39	QBP32939.1
20	Hypothetical protein	3	QBP32903.1
14	Spanin	49	QBP32949.1
13	Hypothetical protein	47	QBP32947.1
10	Hypothetical protein	38	QBP32938.1
6	Hypothetical protein	48	QBP32948.1

three immunoglobulin-like domains. These types of proteins provide capsid stability in phages like phage L (Gilcrease et al., 2005; Tang et al., 2006) and lambda (Sternberg and Weisberg, 1977; Wendt and Feiss, 2004) in addition to hyperthermophilic phages (Bayfield et al., 2019; Stone et al., 2018). The presence of putative decoration proteins in the *Ounavirinae* may play a similar role. Previous work illustrated that for *Shigella* phage Sf14, the decoration proteins exhibit preferential binding as observed through cryo-EM structure analysis (Doore et al., 2018). These proteins are attached to the center of hexamers at the quasi 3-fold axes of symmetry, surrounding each 5-fold vertex, but do not associate with hexamers at the true 3-fold axes. Though frequently annotated as a putative tail protein, similar proteins can be found across the *Ounavirinae* subfamily including FelixO1, SUSP1, and Mooglevirus. We used SDS-PAGE and mass spectrometry to determine whether the putative decoration protein was indeed found in the mature virions. Results indicated that these Mooglevirus proteins (the Sf14 gp20 homolog) were found in mature particles and at high abundance.

When the capsid and terminase genes were compared across the subfamily, both showed relatively low diversity. Although the phages generally formed clades according to geographic location, these genes were still very similar between locations. By contrast, the putative decoration protein exhibited greater divergence both within and between the groups from each location (Fig. 3C–E). Whether this holds true for phages spanning more significant distances is yet to be determined. In addition, since this analysis was performed only with structural genes, other portions of the genome may exhibit a different pattern.

3.4. HRP29 is a new species of Drulisvirus

In contrast to the Mooglevirus, the podovirus HRP29 appears to have diverged significantly from previously-described bacteriophages. With only 10% ANI to any known phage, this isolate presented a taxonomic challenge. In addition, the HRP29 genome did not strictly adhere to properties of other well-characterized podoviruses. The order of structural genes is nearly identical to that of bacteriophage T7 (Fig. 4A), and the tail morphology based on negative staining (Fig. 1) resembles T7 more closely than P22. However, HRP29 encodes a

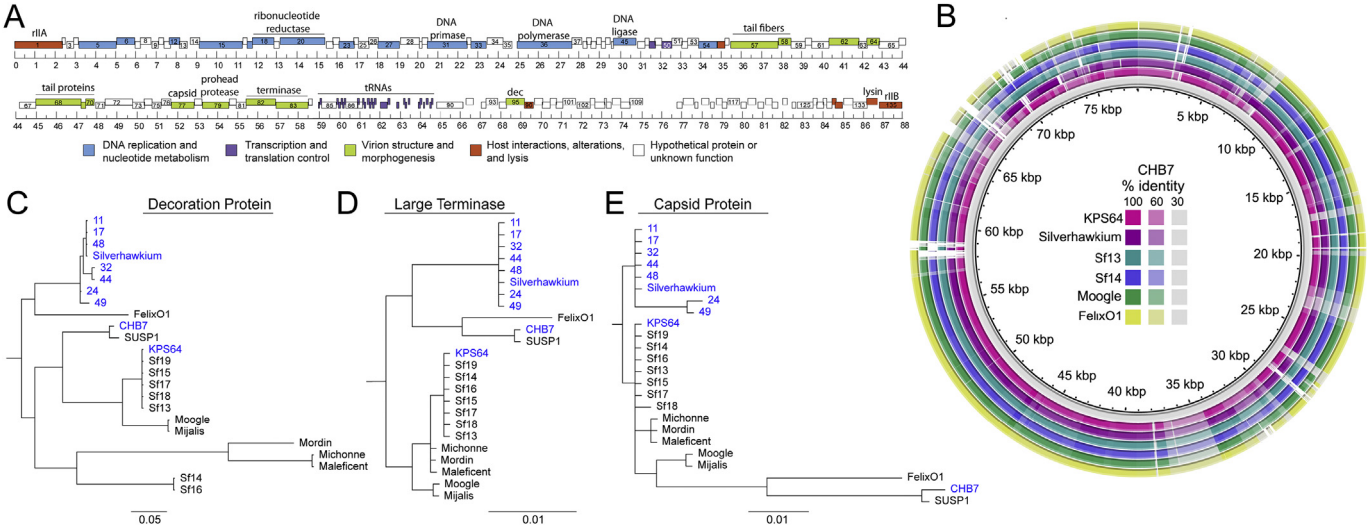

Fig. 3. Genomic and phylogenetic analyses of Mooglevirus isolates. A) Genome map of CHB7, with scale bar in kbp. B) Protein alignments across members of the *Ounavirinae*, with CHB7 as the reference sequence in gray. C-E) Phylogenetic trees of decoration protein, large terminase, and capsid proteins from phages within the *Ounavirinae* subfamily. Phages described in this paper are indicated by blue text; other phages infect *Salmonella* (FelixO1), *Citrobacter* (isolates beginning with the letter M), *Shigella* (isolates indicated by Sf) or *Escherichia* (SUSP1). Scale bar represents amino acid changes per site.

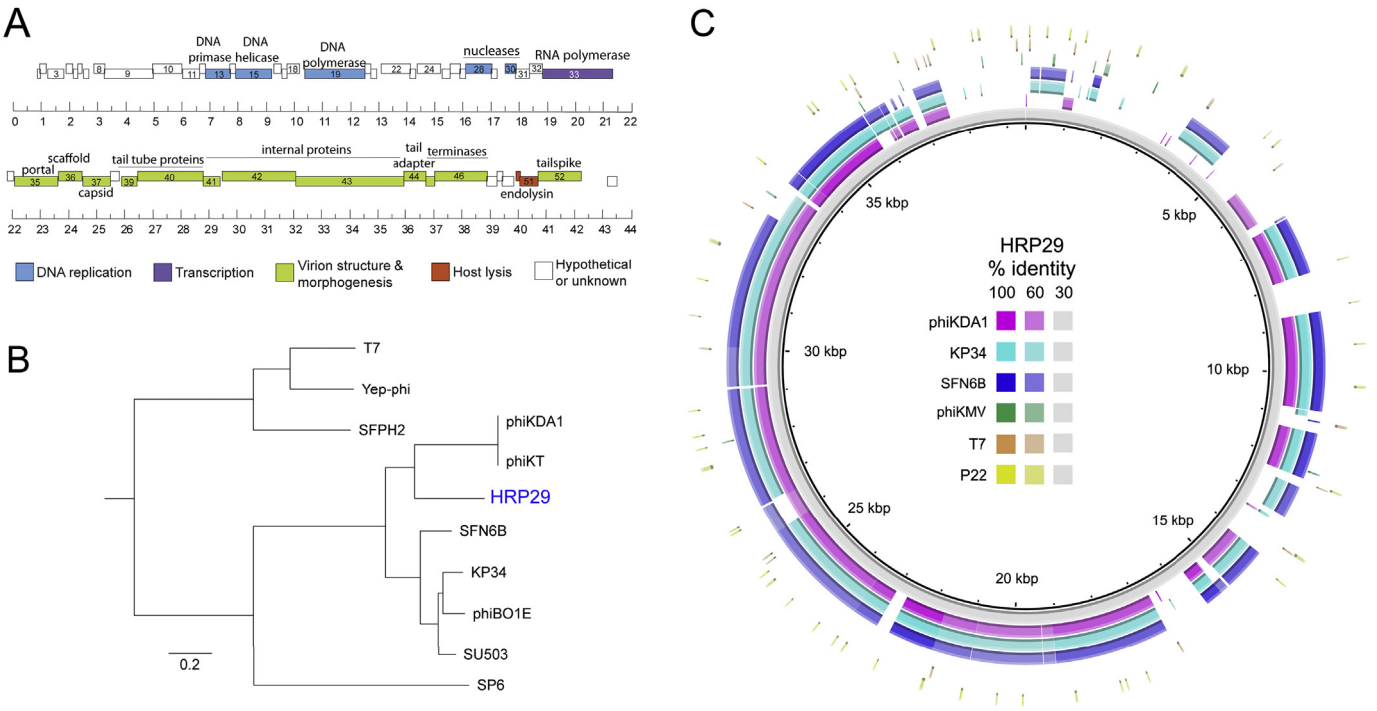

Fig. 4. Genomic and phylogenetic analyses of *Shigella* phage HRP29: A) Genome map of HRP29, with scale bar in kbp and B) phylogeny with other related podoviruses based on whole genome sequence. Scale bar represents nucleic acid substitutions per site. Other phages infect a variety of *Enterobacteriaceae*, including *Escherichia coli* (T7, phiKT), *Yersinia pestis* (Yep-phi), *Shigella flexneri* (SFPH2, SFN6B), *Enterobacter cloaceae* (phiKDA1), *Salmonella enterica* (SP6), or *Klebsiella pneumoniae* (KP34, phiBO1E, SU503). C) Protein-based alignments of HRP29 compared to representative podoviruses, illustrating similarity but not complete identity to Drulisviruses.

tailspike protein that resembles that of Sf6, which is a member of the P22-like viruses. These features suggest a hybrid tail with T7-and P22-like elements. The remainder of the genome is unlike either T7 or P22, instead resembling Drulisvirus KP34 and the unclassified (but putative Drulisvirus) phiKDA1.

To compare the genomic content of HRP29 more thoroughly, whole genomes of T7, P22, KP34, and related phages were used to generate a phylogenetic tree. As shown in Fig. 4B, HRP29 does clearly belong in the *Autographivirinae* subfamily, though it is only distantly related to previously-characterized isolates. Even within the group of its closest relatives—phiKDA1 and phiKT—HRP29 still exhibits significant differences in genomic content. Finally, alignments of protein sequences indicate that HRP29 may tentatively be classified as a Drulisvirus, as it shares the greatest overall similarity to KP34 and phiKDA1 (Fig. 4C). However, many regions share no significant similarity to these two phages. Among these is the region encoding the tailspike, resembling the tailspikes of phages Sf6 and SFN6B, which are both P22-like viruses. Though evolutionarily distant to each other (Fig. 4B), Sf6 and SFN6B also both infect *S. flexneri*. The tailspike is one of the first proteins to contact the bacterial surface, which may explain the conservation of this protein across the three phages. The isolation of additional HRP29-like phages and/or phiKDA1-like phages may further resolve this evolutionary relationship.

4. Discussion

4.1. *Shigella* phages are likely ubiquitous in the environment regardless of outbreak status

Previously, we isolated *Shigella* phages from the environment following an outbreak of shigellosis in the area. Here, we isolated numerous *Shigella* phages from the environment in a different geographic regionand at a time no outbreak had occurred. The prevalence of *Shigella* phages was higher than expected at this second site during three separate sampling sessions across two years (data not shown). The relative overabundance of *Shigella*-infecting phages in Nebraska may suggest that these phages are generally ubiquitous in the environment. Whether *Shigella* is the native host of these phages is still unknown. Since these phages infect all tested serotypes of *S. flexneri*, they may not rely on O-antigen. Lipopolysaccharide of *S. flexneri* is comprised of the inner core, outer core, and O-antigen. While the O-antigen varies between serotypes, the outer core varies between species (Knirel et al., 2011). Conversely, the inner core is conserved across all *Shigella*. Given these properties, we hypothesize that these O-antigen-independent phages are instead specific towards the inner core, explaining their specificity to *S. flexneri* but not their specificity to a given serotype.

The persistence of *S. flexneri* in aquatic environments has been documented in some countries and is analyzed in greater detail elsewhere (Connor et al., 2015; Faruque et al., 2002). Connor and others also discuss the finding that neither *S. boydii* nor *S. dysenteriae* are dominant endemic species in any country, while *S. sonnei* has only recently become endemic in industrialized countries. By contrast, *S. flexneri* has a much longer history of persistence worldwide and was the dominant species in the United States until recently. Our finding that *S. flexneri*-infecting phages are ubiquitous may suggest that the endemic phage population in some locations is still in flux and does not yet reflect this shift in bacterial species dominance.

Unlike the samples collected in Michigan, which originated from various locations along the Red Cedar River, the aqueous samples from Nebraska were largely collected from a pair of static ponds or from drainage rivulets running through and around athletic fields. Even though these two aqueous environments provide fairly different living conditions, it appears that the bacteriophages that inhabit these environments—or at least the *Shigella* phages—generally exhibit low structural and genetic diversity. Our sampling did not include a thorough analysis of the bacteria in these environments beyond screening on colorimetric dye indicator plates. We positively identified both *Salmonella* and *Escherichia* species in the samples, but so far we have not

found *Shigella*. The extent of diversity in terms of potential host strains remains unknown and will need to be further investigated.

4.2. Phages isolated on S. flexneri are overwhelmingly Moogle-like

Prior to the isolation of phages described in (Doore et al., 2018), relatively few bacteriophages could be described as having T = 9 capsids or 85.0–95.0 kbp genomes (Choi et al., 2008; Grose et al., 2014). This may be representative of bacteriophages in general or it may be an artifact of which phages are typically isolated. Relatively few *Shigella* phages had also been isolated prior to 2016. Given that a majority—56%—of *Shigella* phages isolated in two locations appear to be members of this rare group, we favor the latter hypothesis. Interestingly, one of the *E. coli* phages isolated also belongs to this group, so the characteristic does not seem to be confined to *Shigella* phages. Additional isolation of bacteriophages that infect various bacteria, and from various geographic locations, may be informative. Furthermore, our isolation conditions are derived from optimized conditions used in the purification of both podo- and siphoviruses. The strikingly high percentage of Moogle-like *Shigella* in our samples' phages may reflect that this is indeed a common morphology of *Shigella* phages, although additional experimentation and perhaps metagenomic analysis will be required for this conclusion.

4.3. The podovirus HRP29 is a new species of bacteriophage

A majority of the bacteriophages from Nebraska were myoviruses, with only two identical podoviruses being found. These phages, represented by HRP29, morphologically resemble phages of the subfamily *Autographivirinae* but are not genomically similar to any virus in the clade. Only one other T7-like *Shigella* phage has been reported (Yang et al., 2018). This virus, SFPH2, shares > 80% ANI to multiple *Citrobacter* and *Escherichia* phages in the family. Conversely, HRP29 shares a low ANI of ≤10% to a few *Klebsiella* and other miscellaneous phages within the family. SFPH2 and HRP29 also share no detectable similarity between their genomes. Thus, HRP29 is likely a novel virus that is only distantly related to T7 but still falls within the *Autographivirinae* subfamily of phages. Phylogenetic analyses and an examination of synteny suggests that this virus belongs to the Drulisvirus genus, yet it still exhibits significant differences in terms of genomic content.

Prior to our 2016 and 2018 hunts for *Shigella* phages, information for only 35 *Shigella* phages had been deposited into public databases. Our contributions increased this number to 54. Although a majority of *Shigella* phages appear to be Moogle-like members of the *Ounavirinae*, highly novel phages such as HRP29 indicate that we have not saturated the known representative members of *Shigella* phages and have certainly not reached a point where all phages are known. Similar to observations made from SEA-PHAGES data (Pope et al., 2015) and from data of *Enterobacteriaceae*-infecting phages (Grose and Casjens, 2014), additional isolation of environmental phages will be necessary to understand the enormous diversity of bacterial viruses.

Statement of author contributions

SMD, JRS, HRP, KPS, and KNP conceived of and designed the protocols and experiments. SMD, JRS, HRP, WFD, and KPS collected the data. SMD, JRS, and HRP performed the analyses. SMD, JRS, HRP, and KNP wrote the paper.

Acknowledgements

The authors would like to thank Mr. Charley Bittle and all the students at Lincoln Southwest. We also thank Mr. John Dover and Ms. Tori Brown for assistance with experiments and Dr. Sundharraman Subramanian for intellectual contributions. This work was supported by the National Institutes of Health GM110185, the National Science Foundation CAREER Award 1750125, the AAAS Marion Milligan Mason Award for Women in the Chemical Sciences, the NVIDIA GPU Grant Program, and the JK Billman Jr, MD Endowed Research Professorship to KNP. This material is based in part upon work supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454 to SMD. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2019.09.007.

References

Alikhan, N.F., Petty, N.K., Ben Zakour, N.L., Beatson, S.A., 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12, 402.
Arguez, A., Durre, I., Applequist, S., Vose, R.S., Squires, M.F., Yin, X.G., Heim, R.R., Owen, T.W., 2012. NOAA'S 1981-2010 U.S. Climate normals an overview. Bull. Am. Meteorol. Soc. 93, 1687–1697.
Bayfield, O.W., Klimuk, E., Winkler, D.C., Hesketh, E.L., Chechik, M., Cheng, N., Dykeman, E.C., Minakhin, L., Ranson, N.A., Severinov, K., Steven, A.C., Antson, A.A., 2019. Cryo-EM structure and in vitro DNA packaging of a thermophilic virus with supersized T=7 capsids. Proc. Natl. Acad. Sci. U. S. A. 116, 3556–3561.
Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607–2618.
Choi, K.H., McPartland, J., Kaganman, I., Bowman, V.D., Rothman-Denes, L.B., Rossmann, M.G., 2008. Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. J. Mol. Biol. 378, 726–736.
Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.
Connor, T.R., Barker, C.R., Baker, K.S., Weill, F.X., Talukder, K.A., Smith, A.M., Baker, S., Gouali, M., Pham Thanh, D., Jahan Azmi, I., Dias da Silveira, W., Semmler, T., Wieler, L.H., Jenkins, C., Cravioto, A., Faruque, S.M., Parkhill, J., Wook Kim, D., Keddy, K.H., Thomson, N.R., 2015. Species-wide whole genome sequencing reveals historical global spread and recent local persistence in Shigella flexneri. Elife 4, e07335.
Doore, S.M., Schrad, J.R., Dean, W.F., Dover, J.A., Parent, K.N., 2018. Shigella phages isolated during a dysentery outbreak reveal uncommon structures and broad species diversity. J. Virol. 92.
Dover, J.A., Burmeister, A.R., Molineux, I.J., Parent, K.N., 2016. Evolved populations of Shigella flexneri phage Sf6 acquire large deletions, altered genomic architecture, and faster life cycles. Genome Biol. Evol. 8, 2827–2840.
Faruque, S.M., Khan, R., Kamruzzaman, M., Yamasaki, S., Ahmad, Q.S., Azim, T., Nair, G.B., Takeda, Y., Sack, D.A., 2002. Isolation of Shigella dysenteriae type 1 and S. flexneri strains from surface waters in Bangladesh: comparative molecular analysis of environmental Shigella isolates versus clinical strains. Appl. Environ. Microbiol. 68, 3908–3913.
Gilcrease, E.B., Winn-Stapley, D.A., Hewitt, F.C., Joss, L., Casjens, S.R., 2005. Nucleotide sequence of the head assembly gene cluster of bacteriophage L and decoration protein characterization. J. Bacteriol. 187, 2050–2057.
Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420–3435.
Grose, J.H., Belnap, D.M., Jensen, J.D., Mathis, A.D., Prince, J.T., Merrill, B.D., Burnett, S.H., Breakwell, D.P., 2014. The genomes, proteomes, and structures of three novel phages that infect the Bacillus cereus group and carry putative virulence factors. J. Virol. 88, 11846–11860.
Grose, J.H., Casjens, S.R., 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. Virology 468–470, 421–443.
Harvey, F.E., Ayers, J.F., Gosselin, D.C., 2007. Ground water dependence of endangered ecosystems: Nebraska's eastern saline wetlands. Gr. Water 45, 736–752.
Hatfull, G.F., Hendrix, R.W., 2011. Bacteriophages and their genomes. Curr. Opin. Virol. 1, 298–303.
Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., Hatfull, G.F., 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. U. S. A. 96, 2192–2197.
Knirel, Y.A., Kondakova, A.N., Vinogradov, E., Lindner, B., Perepelov, A.V., Shashkov, A.S., 2011. Lipopolysaccharide core structures and their correlation with genetic groupings of Shigella strains. A novel core variant in Shigella boydii type 16. Glycobiology 21, 1362–1372.
Kotloff, K.L., Winickoff, J.P., Ivanoff, B., Clemens, J.D., Swerdlow, D.L., Sansonetti, P.J., Adak, G.K., Levine, M.M., 1999. Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. Bull. World Health Organ. 77, 651–666.

Lander, G.C., Evilevitch, A., Jeembaeva, M., Potter, C.S., Carragher, B., Johnson, J.E., 2008. Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location, and mechanism of attachment determined by cryo-EM. Structure 16, 1399–1406.
Lowe, T.M., Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44, W54–W57.
Pope, W.H., Bowman, C.A., Russell, D.A., Jacobs-Sera, D., Asai, D.J., Cresawn, S.G., Jacobs, W.R., Hendrix, R.W., Lawrence, J.G., Hatfull, G.F., Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science; Phage Hunters Integrating Research and EducationMycobacterial Genetics Course, 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Elife 4, e06416.
Rambaut, A., 2012. FigTree v1.4: Molecular Evolution, Phylogenetics and Epidemiology.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.
Sternberg, N., Weisberg, R., 1977. Packaging of coliphage lambda DNA. II. The role of the gene D protein. J. Mol. Biol. 117, 733–759.
Stone, N.P., Hilbert, B.J., Hidalgo, D., Halloran, K.T., Lee, J., Sontheimer, E.J., Kelch, B.A., 2018. A hyperthermophilic phage decoration protein suggests common evolutionary origin with herpesvirus triplex proteins and an anti-CRISPR protein. Structure 26, 936–947 e933.
Suttle, C.A., 2005. Viruses in the sea. Nature 437, 356–361.
Tang, L., Gilcrease, E.B., Casjens, S.R., Johnson, J.E., 2006. Highly discriminatory binding of capsid-cementing proteins in bacteriophage L. Structure 14, 837–845.
Tritt, A., Eisen, J.A., Facciotti, M.T., Darling, A.E., 2012. An integrated pipeline for de novo assembly of microbial genomes. PLoS One 7, e42304.
Vernhes, E., Renouard, M., Gilquin, B., Cuniasse, P., Durand, D., England, P., Hoos, S., Huet, A., Conway, J.F., Glukhov, A., Ksenzenko, V., Jacquet, E., Nhiri, N., Zinn-Justin, S., Boulanger, P., 2017. High affinity anchoring of the decoration protein pb10 onto the bacteriophage T5 capsid. Sci. Rep. 7, 41662.
Wendt, J.L., Feiss, M., 2004. A fragile lattice: replacing bacteriophage lambda's head stability gene D with the shp gene of phage 21 generates the Mg2 + -dependent virus, lambda shp. Virology 326, 41–46.
Yang, C., Wang, H., Ma, H., Bao, R., Liu, H., Yang, L., Liang, B., Jia, L., Xie, J., Xiang, Y., Dong, N., Qiu, S., Song, H., 2018. Characterization and genomic analysis of SFPH2, a novel T7virus infecting Shigella. Front. Microbiol. 9, 3027.